Comparative and evolutionary pharmacogenetics of *ABCB1*: complex signatures of positive selection on coding and regulatory regions

Haijian Wang[a], Keyue Ding[d], Yang Zhang[a], Li Jin[a,b], Iftikhar J. Kullo[d] and Fuchu He[a,c]

Background **As a major mediator in the complex interplay between humans and the xenobiotic environment, the *ABCBI* transporter gene is an obvious candidate for comparative and evolutionary pharmacogenetic studies. It has been recently reported that common variants in its coding region, which are variously associated with drug response and disease susceptibility, may have conferred differential selective sweep in various populations. Fully profiling the alletic architecture and explicitly interrogating the natural selection at *ABCBI* are needed to understand its evolutionary population genetics.**

Methods and results **Using a comprehensive single nucleotide polymorphism variants in coding and regulatory regions, as well as comparable genotype data from the Environmental Genome Project, we systematically characterized the extent and length of linkage disequilibrium throughout the *ABCBI* locus in three major ethnic populations (African, European, and Chinese). We observed pronounced signals of recent positive selection on the derived alleles of three common single nucleotide polymorphisms coding regions: e12/1236T, e21/2677T, and e26/3435T in the Chinese, as well as on extended haplotype homozygosity were also observed for two potentially functional common variants in the 5′f/-4489G (rs17149810) in the Chinese and 5′f/-693T (rs3213619) in the Africans, respectively, which may have shaped the phylogenetically inferred star-like haplotype structure of the 5′flanking region.**

Conclusion **Our finding reveal complex signatures of natural selection on both coding and regulatory regions of the human ABCBI gene, point to potential functional relevance of its regulatory variants, and suggest that evolutionary dynamics and transcriptional regulation may underline the phenotypic variation in xenobiotic disposition and varying predisposition to complex in which xenobiotics play a role. *Pharmacogenetics and Genomics* 17:667–678**

Pharmacogenetics and Genomics 2007, **17**:667–678

Keywords: ***ABCB1*****, haplotype, positive selection, regulatory variation**

[a]MOE Key Laboratory of Contemporary Anthropology, Center for Evolutionary Biology, State Key Laboratory of Genetic Engineering, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, [b]CAS-MPG Partner Institute of Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, [c]Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Chinese National Human Genome at Beijing, Beijing, P.R. China and [d]Division of Cardiovascular Diseases, Mayo Clinic and Foundation, Rochester, Minnesota, USA

Correspondence to Dr Fuchu He, Institutes of Biomedical Sciences, Fudan University, 138 Yixueyuan Rd., Shanghai 200032, P.R. China
Tel: +86 21 54237586; fax: +86 21 54237586; e-mail: hefc@fudan.edu.cn;
or
Dr Iftikhar J. Kullo, Division of Cardiovascular Diseases, Mayo Clinic, 200 First Street Southwest, Rochester, MN 55905, USA
Tel: +1 507 284 9049; fax: +1 507 266 1702;
e-mail: kullo.iftikhar@mayo.edu

*The first two authors contributed equally to this work.

Received 16 December 2006 **Accepted** 12 March 2007

Introduction

The multidrug resistance 1, MDR1 (P-glyco-protein, P-gp), is one of the well-characterized members of a family of transmembrane transporters (ABCB1, the ATP-binding cassette, subfamily B, member 1), whose function is the energy-dependent export of substances from inside of cells and from membranes to outside [1]. P-gp was initially reported to be overexpressed in tumor cells and accounts for much of the cross-resistance to structurally and mechanistically unrelated drugs, a multifactorial trait known as multidrug resistance [2]. The range of substrates for P-gp, its tissue expression profile and physiological role, however, has since expanded. Expressed in many normal tissues, such as intestinal epithelial cells [3], biliary hepatocytes [4], proximal tubules of kidney [4], blood–brain barrier [5], and placenta [6], P-gp is now considered to be one of the major determinants of the ADME (absorption, distribution, metabolism, and excretion) – Tox (toxicological) profiles, of a large range of hydrophobic exogenous substrates, including nearly every category of clinically important drugs. The degree of expression and functionality of the transporter, which is largely determined by variations in its coding gene, *ABCB1* (MIM 171050), may not only influence xenobiotic disposition, affect outcome of therapy for several diseases such as acute myeloid leukemia [7], epilepsy [8], and HIV-1 infection [9], but also influence interindividual susceptibility to diseases

1744-6872

such as inflammatory bowel disease and renal epithelial tumors [10–12].

The identification of sequence variants of *ABCB1* and functional characterization of their biological significance and clinical relevance have stimulated interest in the pharmacogenetics of drug transporters [3]. The primary focus has been on common single nucleotide polymorphisms (SNPs) in the coding regions, such as 1236 C > T (rs1128503) in exon 12 (synonymous), 2677 G > T/A (rs2032582) in exon 21 (Ala893Ser/Thr), and 3435 C > T (rs1045642) in exon 26 (synonymous). Only a few of these SNPs, including the well-defined synonymous 3435 C > T, serve as predictors of P-gp biochemical phenotype, drug response, clinical condition and outcome [13,14]. The causal relevance of this marker, however, is yet to be corroborantly established because of conflicting results from population-based association studies [13]. Additionally, several studies investigating its functional *cis*-regulatory polymorphisms, identified a couple of SNPs and phased haplotypes in the 5′ upstream region, and further characterized their effect on transcriptional activation [15–17]. These studies suggested that different haplotypes in the *ABCB1* 5′ flanking region have different effects on the promoter activity in various cell lines, indicating that nucleotide and/or haplotype variants not only in the coding but also regulatory region of *ABCB1* may be important for interindividual differences of P-gp expression and activity.

Parallel to exploring phenotypic effects and clinical relevance of *ABCB1* genetic polymorphisms, investigating molecular adaptation to environmental xenobiotics is an intriguing complementary strategy for pharmacogenetic studies. The aforementioned three common SNPs in coding region have been reported to occur in different ethnic populations with markedly varying frequencies. The level of linkage disequilibrium (LD) in the intragenic region of *ABCB1* is strong but varies among different populations [18,19]. Genomic evidence for positive selection at the exon26/3435T alleles in a Chinese population, as well as at the alternative 3435C in an African population, has been recently reported, suggesting that independent mutational events may have occurred on the *ABCB1* gene to confer positive selection in the non-African and African populations, respectively [20]. It should be noted that this population genetics study was based on only 12 SNPs with no information about ancestral or derived states and with limited coverage of the ~200 kb genomic sequence of this large gene, and did not encompass polymorphisms in the regulatory region. Therefore, the suggested different selective forces acting at the *ABCB1* locus in different populations, with respect to geographically and temporally fluctuating xenobiotic environments in our demographic history, are not yet fully explained.

Our aim in this study was to better delineate sequence variation and haplotype architecture of the *ABCB1* gene, and thereby to gain more explicit and logical evolutionary clues to the underlying process of natural selection on this locus. We resequenced ~11 kb of the locus in three major ethnic populations (African, European, and Chinese) and obtained a full spectrum of variants in both its coding and regulatory regions. In addition, we used the genotype data of *ABCB1* gene from a publicly available database – the Environmental Genome Project (EGP), *http://egp.gs.washington.edu/*. On the basis of these two independent data sets, we observed signatures of positive selection on both coding and regulatory regions of *ABCB1* gene, which point to potential functional relevance of its regulatory variants and provide insight into evolutionary dynamics and transcriptional regulation of the key component of the xenobiotic disposition system.

Materials and Methods

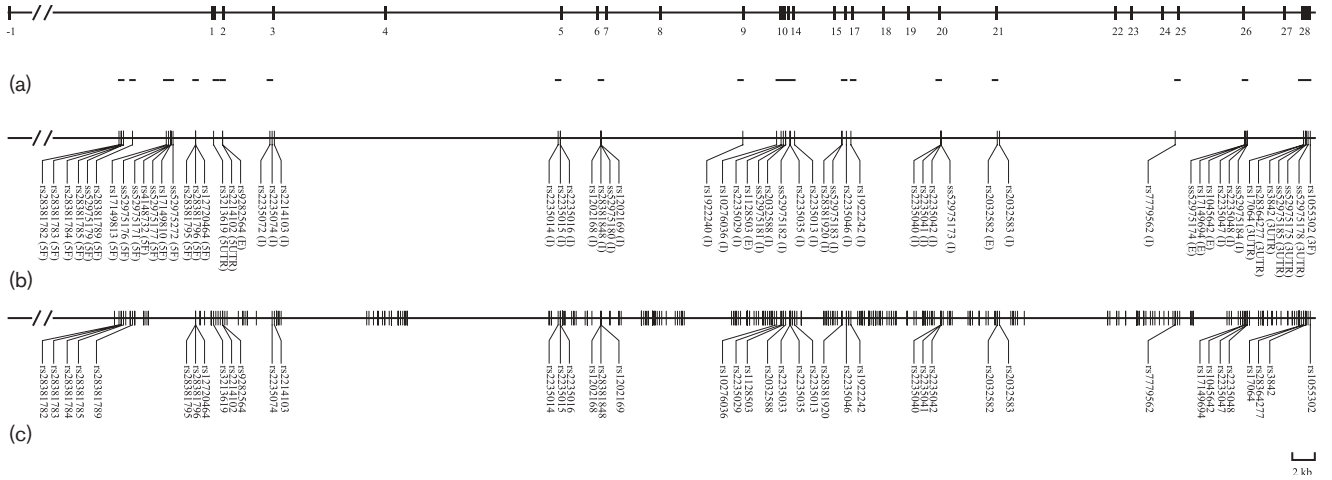
DNA samples

We sampled 74 unrelated human individuals from Africans (Yorubans, 23 individuals), Europeans (24 individuals), and Asians (27 Chinese individuals). The sample size in the settings of this SNP discovery with resequencing enabled to provide 95% detection rate in a subpopulaton for SNPs with a minor allele frequency (MAF) of ≥ 0.05 [21]. Genomic DNA samples of the 27 Chinese individuals were chosen from a sample collection assembled for the Chinese Human Genome Diversity Project through a coordinated effort of several institutes [22]. Genomic DNA samples for Europeans and Africans were obtained from the Coriell Cell Repositories (Camden, New Jersey, USA). Most SNPs are likely to represent a single mutational event as the mutation rate at a nucleotide position is estimated at $< 10^{-9}$ per year [23]. The nucleotide state in other hominoids (such as *Pan troglodyte*, the closest relative of *Homo sapiens*) at the homologous site can be used to infer the ancestral state for segregating sites [24]. And therefore, genomic DNA sample of a chimpanzee (*Pan troglodyte*) as outgroup was also obtained from the Coriell Cell Repositories.

PCR-resequencing-based single nucleotide polymorphism ascertainment

In the Chinese sample of 27 individuals, we not only covered all the 28 exons of *ABCB1* gene (Fig. 1a) for SNP screening, but also extended resequencing ~9.5 kb upstream from the ATG start codon to the distal enhancer, to which the pregnane X receptor (PXR), a promiscuous xenosensor binds [25]. We identified 45 SNPs in the Chinese, including 30 common SNPs with MAF > 0.05, three doubletons, and 12 singletons. With an intention to perform haplotype-based evolutionary analyses that were partly superimposed on sequence-based strategy, we sequenced in the other two populations, as well as the outgroup, the genomic regions that contain the 30 common SNPs in the Chinese using the

Fig. 1



Distribution of 62 single nucleotide polymorphisms (SNPs) across the *ABCB1* gene. (a) Genomic structure. The horizontal line represents its genomic entire length, and 28 exons are depicted by vertical bars above the line. The lines below indicate the resequenced regions in this study. (b) Distribution of SNPs. Each vertical line represents one SNP, and the label for each SNP is given below the line. The category of each SNP is presented in the parenthesis. 5F, 5′ flanking regions; 5UTR, 5′ untranslated regions (UTRs); I, intron; E, exon; 3UTR, 3′UTR; 3F, 3′ flanking regions. (c) Distribution of SNPs from Environmental Genome Project database: 43 SNPs were shared between (b) and (c).

same array of primer pairs (Fig. 1b). We thereby built, for the genomic region of *ABCB1* resequenced in each individual, a dataset of 62 SNPs (Table S1) for analyses, with only five singletons and one doubleton in the Chinese left out of this common dataset (data not shown). The ancestral states of each of the 62 SNP were inferred from the *Pan troglodyte* sequence. The PCR-resequencing-based SNP discovery platform has been described previously [26]. Details regarding PCR and sequencing conditions and PCR primers are available on request. As a measure of quality control, sequence segments of individuals presenting singletons or ambiguous polymorphisms were reamplified and resequenced. We assessed SNP data validity by repeating 10% of the genotype assays, on the basis of PCR and bidirectional resequencing. The error rate was relatively low (1.2%).

During our work on this project, the EGP released the SNP dataset of *ABCB1* for five populations [African-Americans, African (Yoruban), European, Hispanic, and Asian], in which all the exons, a much larger part of intronic sequence, about 2.5 kb upstream sequence from the ATG start codon, and about 3 kb sequence covering the PXR response element, were resequenced, and genotype data of 485 SNPs were deposited, 43 of which were shared with ours (Fig. 1c). The ancestral state for each SNP was obtained by aligning the FASTA sequence for each SNP (in the dbSNP database, http://www.ncbi.nlm.nih.gov/SNP/) to the draft genome build of chimpanzee sequence from the UCSC database, using the 'blat' program [27]. In this study, we analyzed genotype data of three of the five populations, African, European, and Asian, as one independent and comparable dataset.

Data analysis

We used the 'genetics' package implemented in *R* to perform population genetics analyses, including description of allele frequencies, and Fisher's exact test for Hardy–Weinberg equilibrium. The false discovery rate method was used to correct for multiple testing using the package 'QVALUE' in *R* (http://faculty.washington.edn/jstorey/qvalue) [28]. Two measures of LD between SNP pairs, the absolute D′ and the correlation coefficient r^2, were calculated for the three populations separately. The χ^2 test was used to assess statistical significance of LD, followed by false discovery rate correction for multiple testing. As rare alleles with frequencies < 0.05 do not have sufficient statistical power for LD detection [29,30], LD analysis and the following LD-based evolutionary inference were based on only common SNPs (MAF ≥ 0.05). Haplotype reconstruction was performed using the Bayesian method implemented in PHASE [31]. A median-joining haplotype network was constructed using NETWORK 4.2.0.0 (http://www.fluxus-technology.com) [32].

F_{ST}, a measure of population differentiation, quantifies variance of allele frequency between and within populations, and is used to detect local adaptation in the human genome [33]. We calculated an unbiased small-sample estimator of pairwise populations F_{ST} from pairwise population comparisons using the methods by Weir and Cockerham [34], which corrects the effect of small

sample size. A bootstrapping method (1000 bootstrap samples) was used to test the statistical significance of F_{ST} in each pairwise comparison between populations.

We calculated haplotype homozygosity (HH) in a stepwise manner – extended HH (EHH) – to explore how LD breaks down with increasing distance from a specific core SNP [35]. HH was calculated between a distance X and the specified core SNP for a chromosome population carrying a single allele of the core SNP (ancestral and derived allele, respectively). Distance X increases stepwise to the most outlying SNP. The patterns of HH were estimated on both sides of each allele for a specific core SNP. EHH is on a scale of 0 (no homozygosity, all extended haplotypes are different) to 1 (complete homozygosity, all extended haplotypes are the same). HH is evaluated as, $HH = (\sum p_i^2 - 1/n)/(1 - 1/n)$, with p_i being the relative haplotype frequency (for a core SNP, it is the allele frequency) and n the sample size. The variance of each HH [Var(HH)] is estimated as,

$$Var(HH) = \frac{2(n-1)}{n^3}\left\{2(n-2)\left[\sum p_i^3 - (\sum p_i^2)^2\right] + \sum p_i^2 - (\sum p_i^2)^2\right\}$$

according to Nei [36]. Relative EHH (REHH) is the ratio of the EHH on the tested core haplotype (core allele here) compared with the EHH of the grouped set of core haplotypes at the region not including the core haplotype tested [35]. We used sweep 1.0 software to visualize the decay of LD on ancestral and derived alleles for a given core SNP, drawing a haplotype bifurcation diagram (HBD) [35]. To test statistical significance of the long-range haplotype (LRH) analysis, we obtained the empirical distribution of core haplotype frequencies vs. REHH through screening the entire chromosome 7 HapMap data (*http://www.hapmap.org*, release #16) in Yoruban (YRI), European-descent populations (CEU), and Han Chinese in Beijing (HCB), respectively.

Results

Data summary and sequence variations

We sampled 74 unrelated human individuals in three populations and a chimpanzee as outgroup. By resequencing large segments of the genomic regions of *ABCB1* (total length of 11 326 bp) that include all known common SNPs in its exons, exon–intron boundaries, 3′ flanking regions, and especially in ~9.5 kb of the 5′ flanking region, we built a dataset of 62 SNPs for the three populations (Fig. 1). The homologous sites for all the 62 SNPs in *Pan troglodyte* were homozygous, and the ancestral states for each SNP were accordingly inferred. The SNP symbol, reference SNP accession number (rs, 47 SNPs) or submitter SNP accession number (ss, 15 SNPs), classes, and ancestral and derived allele frequencies (inferred by comparison with *Pan troglodyte*), are shown in Table S1. Only two SNPs (30 and 37) in Europeans, and two (53 and 57) in Africans, departed significantly from Hardy–Weinberg equilibrium (Fisher's exact test, $P < 0.05$). After correction for multiple testing, however, these departures were not statistically significant.

We analyzed sequence variation of *ABCB1* gene in the three populations, using subsets of all the screened 49, 41, and 39 SNPs in Africans, Europeans, and Chinese, respectively (Table 1). Two measures of nucleotide diversity, θ_w and π, were calculated for the resequenced regions. The average nucleotide diversity (π) for the three populations was comparable with that reported previously. The classical test of evolutionary neutrality test on the basis of nucleotide diversity-Tajima's D [37] did not show significant deviation from zero ($P > 0.05$) in any population. By using the PHASE program, 44, 28, and 31 distinct haplotypes were inferred for the Africans, Europeans, and Chinese, respectively. Significant inter-ethnic variability in the number and frequency of haplotypes was observed among populations, with haplotype diversity being 0.976, 0.957, and 0.900 for the Africans, Europeans, and Chinese, respectively.

Population differentiation

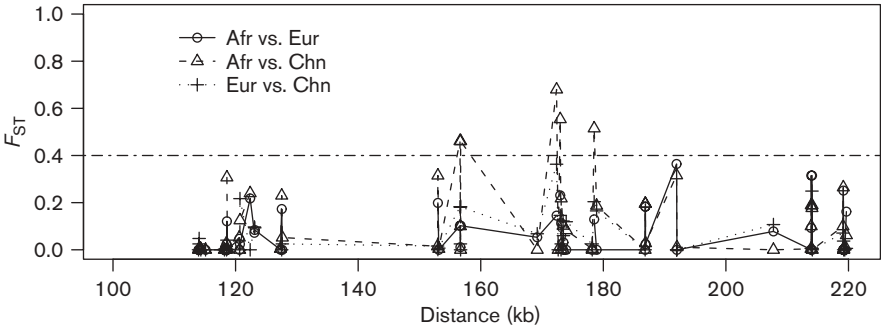
To test whether there are significant differences in allele frequency for *ABCB1* gene among populations, we calculated the measure of population differentiation (F_{ST}) in a pairwise manner. The *ABCB1* showed significantly high level of population differentiation: 19 out of 54 loci in Africans vs. Europeans, 24 out of 58 in Africans vs. Chinese, and 14 out of 50 in Europeans vs. Chinese, showed significantly high F_{ST} (Table S1). Figure 2 demonstrates the distribution of F_{ST} against physical distance. A higher F_{ST} was noted in Africans vs. Chinese,

Table 1 Nucleotide diversity in the resequenced regions of *ABCB1* among the three populations

Population	2N	Length (bp)	Number of SNP	Singletons	Nucleotide diversity ($\times 10^{-4}$)		Tajima's D (P value[a])
					π	θ_w	
African	46	11 326	49	11	8.27	7.56	0.318 (0.149)
European	48	11 326	41	8	8.65	7.96	0.297 (0.162)
Chinese	54	11 326	39	7	9.82	9.84	−0.008 (0.276)

[a]The P value was calculated from coalescent simulations under a standard neutral model given the recombination rate of 0.39 cM/MB in the *ABCB1* locus (D7S644). SNP, single nucleotide polymorphism. π, the average number of nucleotide differences per site between two sequences chosen from a randomly mating population. θ_w, the estimate of population mutation parameter based on the number of polymorphic sites in the sample.

Fig. 2



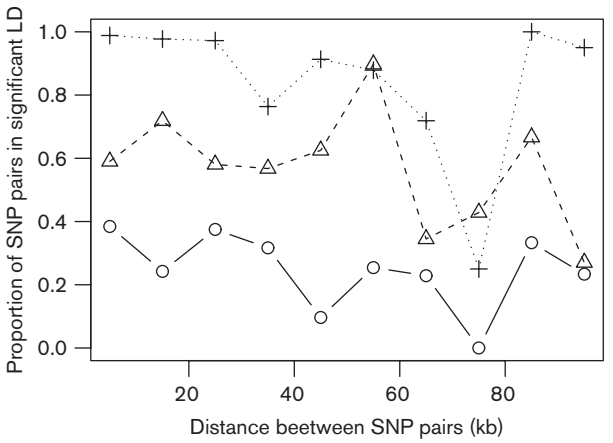
F_{ST} statistic (population differentiation) along the genetic distance, in pairwise populations for genotype data in Africans, Europeans, and Chinese.

as compared with Africans vs. Europeans and Europeans vs. Chinese. For example, eight SNPs showed F_{ST} (> 0.30) in African vs. Chinese, considerably higher than the well-documented genome-wide average of $0.10 \sim 0.15$ [33,38]. Population differentiation of the three well-known SNPs in coding regions (i.e. SNP 33: e12/1236; SNP 47: e21/2677; and SNP 52: e26/3435) was significantly higher (bootstrap P value < 0.05, Table S1). It should also be noted that two SNPs in the 5′ flanking regions/untranslated regions (UTR) (SNP 12: 5′f/-4489 and SNP 17: 5′f/-693) had significantly high F_{ST} values (0.31 and 0.24 in Africans vs. Chinese).

Variation of linkage disequilibrium pattern among populations

We estimated D′ and r^2 of common SNPs (MAF ≥ 0.05) for the *ABCB1* region to determine LD patterns of *ABCB1* gene in the three populations. Overall, the LD between SNP pairs was high but varied considerably among different populations. The relationship between LD and physical distance is shown in Figure S1. It showed significant and strong intragenic LD levels: 27.5% SNP pairs showed significant LD in Africans, 57.4% in Europeans, and 88.1% in Chinese. Some SNP pairs separated by ~ 100 kb showed nearly complete LD (e.g. $|D'| = 1$). We also observed obvious population variation in LD levels when plotting the proportion of SNP pairs in significant LD against physical distance for each population separately (Fig. 3). The Africans showed lower LD levels, whereas the Chinese manifested an extremely high degree of allelic association. The three core SNPs in coding region were in relatively strong LD, and the derived haplotype, which is phased on the three derived 'T' alleles, was significantly overrepresented in Europeans and Chinese than in Africans. This observation is in accordance with the previous studies [18,39]. In the regulatory regions, pronounced LD was also observed for SNPs 1, 6, 7, 12, 15, and 17. In the Chinese population, all of these variants were perfectly associated in phase. Among these segregating sites, SNPs 15 and 17 were reported to be also in complete LD in a Japanese population [15].

Fig. 3

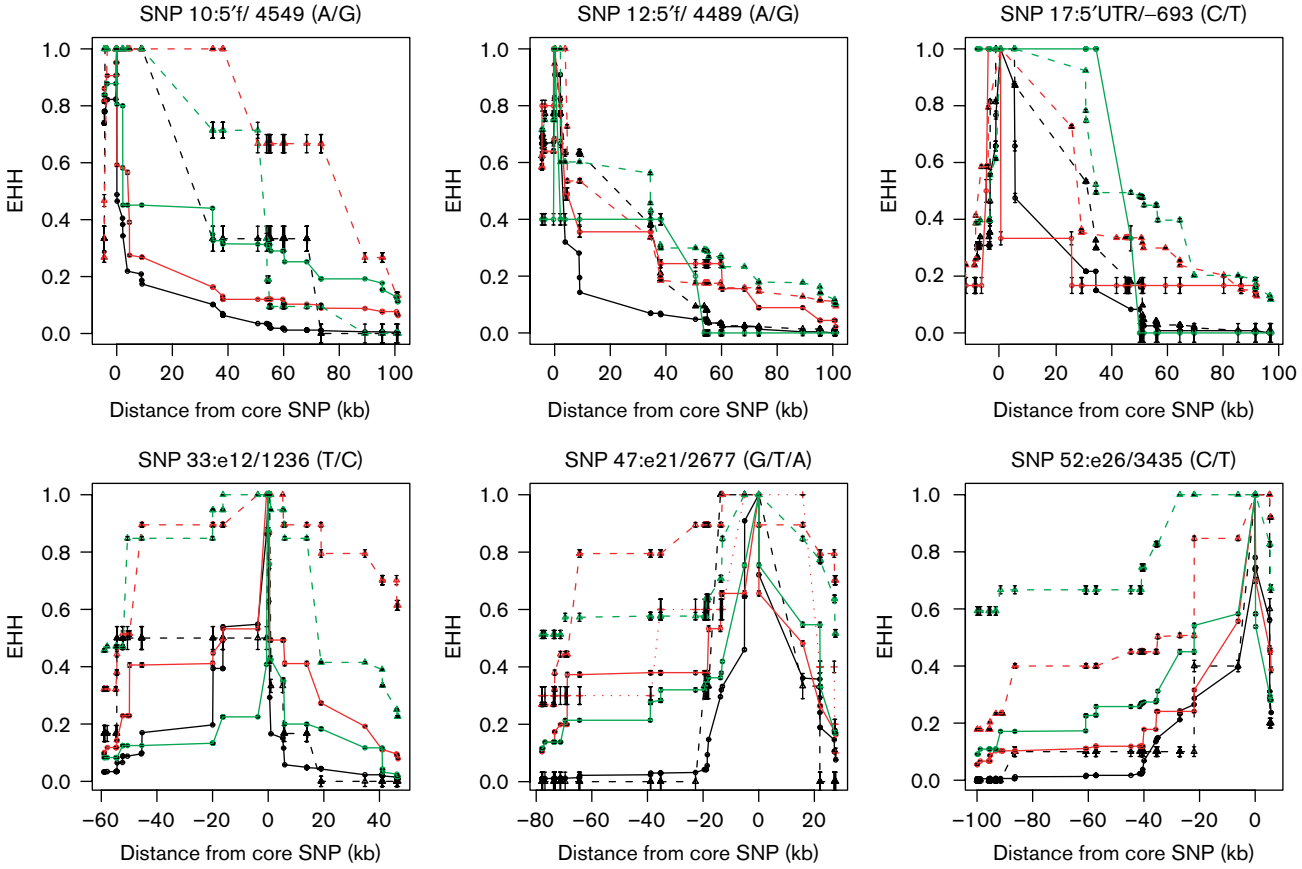


Proportion of single nucleotide polymorphism (SNP) pairs in significant linkage disequilibrium (LD) plotted against physical distance in the *ABCB1* ~ 100 kb region. SNP pairs were grouped into 10-kb bins. ○, Africans; Δ, Europeans; and +, Chinese.

Long-range haplotype test for positive selection

To identify mutations in the *ABCB1* locus potentially targeted by recent positive selection, and that might have had molded the population differentiation and LD pattern, we performed the LRH test. Six core SNPs that all cosegregated in each population were selected, including the three well-known variants in the coding regions (i.e. SNPs 33, 47, and 52), two SNPs in 5′ flanking regions/UTR with significantly high F_{ST} (SNPs 12 and 17), as well as one SNP in 5′ flanking regions with a derived allele frequency > 0.10 (SNP 10, 5′f/-4549). As revealed by the EHH plots (Fig. 4), the overall EHH of derived alleles decayed more slowly than that of ancestral alleles. For the core SNP 10 in the 5′ flanking region, the EHH of derived allele (broken lines) over ~ 51 kb were 0.68 and 0.71 in Europeans and

Fig. 4



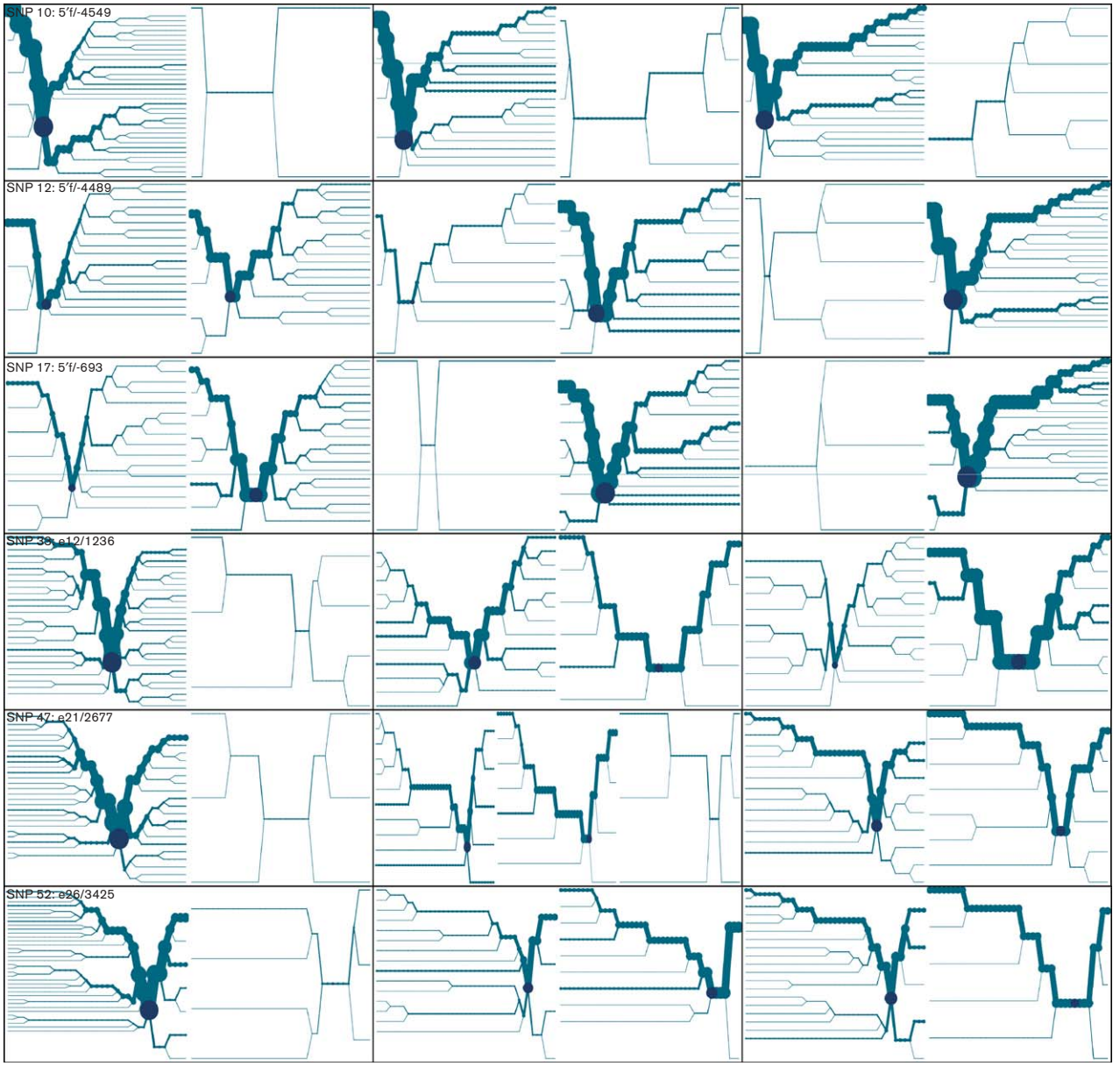
Plot of extended haplotype homozygosity (EHH) vs. physical distance for six core single nucleotide polymorphisms (SNPs) in the *ABCB1* gene: SNPs 10 and 12 in the 5′ flanking regions, SNP 17 in 5′ untranslated region (UTR), and SNPs 33, 47, and 52 in the coding regions. In each plot, different colors represent different populations: black, Africans; red, Europeans; green, Chinese. The solid and broken lines indicate the ancestral and derived allele, respectively.

Chinese, respectively, whereas the EHH of their ancestral allele over the same distance was 0.12 and 0.31. Similar decay of EHH with distance was also observed for the core SNPs 12 and 17. For the three SNPs in the coding regions, the EHH of derived allele decayed more slowly than their ancestral alleles, especially in Europeans and Chinese. Consistent with the EHH results, the haplotype bifurcation diagrams showed that the derived alleles of two core SNPs in 5′ flanking region/UTR (SNPs 12 and 17) had long-range LD across the three populations, especially in Europeans and Chinese, and the derived alleles of the three core SNPs in coding region demonstrated, in Europeans and Chinese, clear long-range LD as a predominance of one thick branch (Fig. 5). Furthermore, in the chromosome-wide empirical distribution of REHH obtained through screening the entire chromosome 7 HapMap data in three comparable populations (Fig. 6), a couple of outliers in both coding and regulatory regions deviated significantly from evolutionary neutrality in specific populations. In the 5′ flanking region, the derived allele (T) of SNP 17 in Africans ($P = 0.0066$), and the derived allele (G) of SNP 12 in Europeans ($P = 0.0000$) deviated significantly from neutrality. In the coding regions, the derived allele of the three core SNPs in Chinese also showed pronounced evidence of positive selection ($P = 0.0000$, 0.0050, and 0.0019 for SNPs 33, 47, and 52, respectively). Significant deviation from neutrality for the derived allele of SNP 33 was also observed in Europeans ($P = 0.0011$).

Haplotype structure and positive selection on *ABCB1* 5′ regulatory regions

As revealed by the LRH test, in addition to the three common SNPs in the coding region, whose phenotypic relevance and evolutionary implications have been well addressed in the literature, the derived allele of two SNPs in the 5′ flanking regions (SNP 17 in Africans and SNP 12 in Europeans) showed significant deviation from evolutionary neutrality, suggesting that positive selection might have operated on the regulatory regions of *ABCB1*. To further understand how these two SNPs and nearby variants affected haplotype structure in the regulatory regions among different populations, we selected

Fig. 5



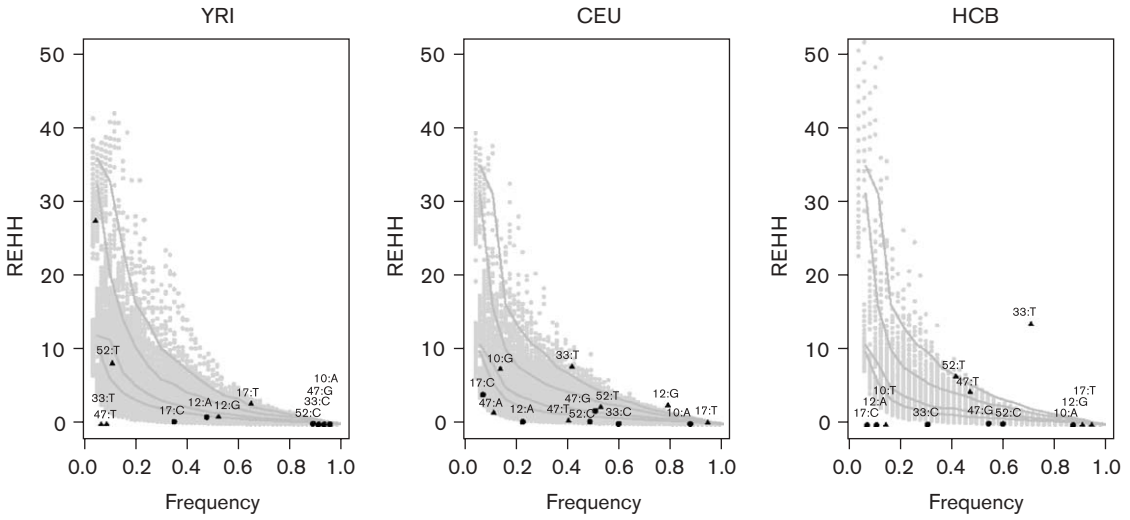
Haplotype bifurcation diagrams (HBDs) for six core single nucleotide polymorphisms (SNPs) (see legend for Fig. 4) in the three populations. In each plot, the root is an allele for the core SNP (left, ancestral allele; right, derived allele). Each marker is a node when moving in one direction. Column left, Africans; middle, Europeans; and right, Chinese.

12 common SNPs (MAF >0.05 at least in one population) out of the 19 SNPs located in the 5′ flanking region, to construct the median-joining network of haplotypes in *ABCB1* 5′ regulatory region (Fig. 7). The ancestral haplotype (e.g. haplotype A) was constituted mostly by the Africans. Two mutation events, that is SNPs 17 and 12, were needed to generate haplotype C and its derivatives (e.g. haplotypes D, G, and H), which were predominant as star-like phylogeny in non-Africans (Table 2). Therefore, the likely genealogical history of mutations in 5′ flanking regions further supported positive selection having led to an increase in the frequency of SNPs 17 and 12 in Europeans and Chinese and thereby shaped the haplotype patterns of *ABCB1* 5′ regulatory region.

Signatures of positive selection on *ABCB1* inferred from Environmental Genome Project data

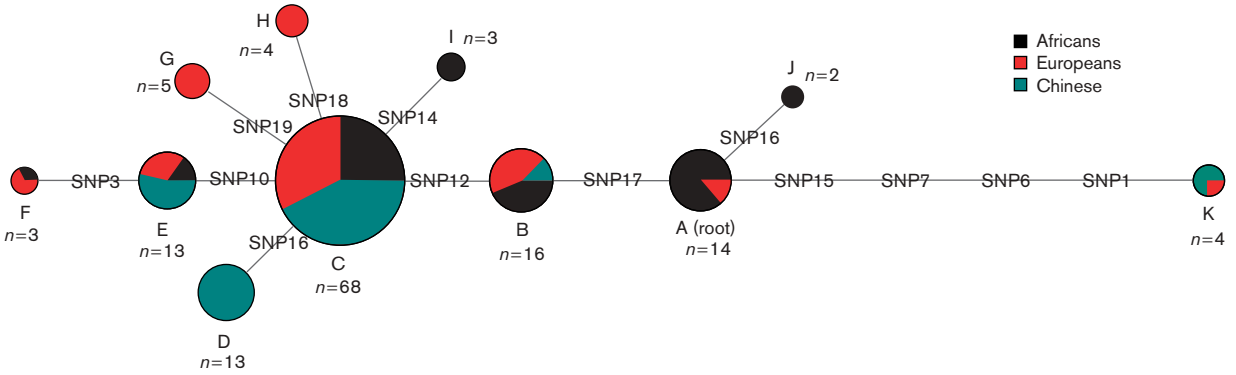
We also analyzed the resequenced genotype data for *ABCB1* in EGP database. There are 269, 187, and 181

Fig. 6



Relative EHH (REHH) tests for six core single nucleotide polymorphisms (SNPs) on the basis of empirical distribution in three populations. Relative EHH of each allele was plotted against its frequency. The points of REHH against allele frequency (light grey) were calculated from phased data of chromosome 7 in HapMap database using the 'sweep' program (see Materials and Methods section). The 50, 75, 95, and 99th percentile lines (grey) of the distribution of empirical data were also obtained by binning the empirical data by allele frequency into 20 bins of equal size with intervals of 5% (from bottom to upper). The circles and triangles (black) indicate the Relative extended haplotype homozygosity (REHH) against frequency of the ancestral and derived alleles, respectively. YRI, Yoruba; CEU, CEPH Europeans; and HCB, Han Chinese.

Fig. 7



A median-joining network of the 11 unique *ABCB1* 5′ regulatory region haplotypes (frequency ≥ 2) found in three populations. Haplotypes were labeled as 'A ~ K'. The size of each node is proportional to the haplotype frequency in the total samples, and the frequency of each haplotype within each population is indicated by the varied colors within each node. Branch lengths represent one nucleotide substitution. Haplotype A is the ancestral haplotype from Chimpanzee sequence.

SNPs in the Africans, Europeans and Asians, respectively. The patterns of population differentiation were similar to those in our dataset, with a notable fraction of pairwise $F_{ST} > 0.30$ (Figure S2). In the coding regions, the average F_{ST} among the three populations for the three well-known SNPs were 0.32 (e12/1236), 0.279 (e21/2677), and 0.21 (e26/3435), respectively. In the regulatory regions, $F_{ST} > 0.30$ was noted in six SNPs for Africans vs. Europeans, eight SNPs for Africans vs. Asians, and one SNP for European vs. Asians. We then focused on the three core SNPs in coding region to perform the LRH analysis. As in the prior analyses, the EHH of derived allele decayed more slowly than the ancestral allele, especially in Europeans and Asians (Figure S3). Results of the LRH test for the three core SNPs are shown in Figure S4. Compared with the results of our resequenced data in Fig. 6, the derived alleles of two SNPs (33 and 52) in Africans also appeared to deviate from neutrality. One

Table 2 Haplotypes and their frequency in the 5′ flanking region of *ABCB1*

	SNP												Frequency		
Haplotype	1	3	6	7	10	12	14	15	16	17	18	19	African	European	Chinese
A	C	T	G	G	A	A	A	T	G	C	G	A	0.26	0.04	–
B	·	·	·	·	·	·	·	·	·	T	·	·	0.15	0.15	0.04
C	·	·	·	·	·	G	·	·	·	T	·	·	0.37	0.46	0.54
D	·	·	·	·	·	G	·	·	A	T	·	·	–	–	0.24
E	·	·	·	·	G	G	·	·	·	T	·	·	0.04	0.08	0.13
F	·	C	·	·	G	G	·	·	·	T	·	·	0.02	0.04	–
G	·	·	·	·	·	G	·	·	·	T	·	G	–	0.10	–
H	·	·	·	·	·	G	·	·	·	T	A	·	–	0.08	–
I	·	·	·	·	·	G	T	·	·	T	·	·	0.07	–	–
J	·	·	·	·	·	·	·	·	A	·	·	·	0.04	–	–
K	G	·	A	A	·	·	·	C	·	·	·	·	–	0.02	0.06
L	G	·	·	A	·	·	·	·	·	·	·	·	0.02	–	–
M	G	·	A	A	·	G	·	C	·	·	·	·	0.02	–	–
N	·	·	·	·	·	G	·	·	·	T	·	·	–	0.02	–

Haplotype A is the ancestral haplotype inferred from the *Pan troglodytes*. SNP, single nucleotide polymorphism.

reason accounting for the discrepancy among the Africans in these studies might be the smaller sample size in the EGP data (i.e. 12 individuals), which led to a higher derived frequency for these two SNPs (0.208 and 0.250 for SNPs 33 and 52, respectively). In Europeans, the REHH of the three derived alleles were larger than 95th percentile given the frequency on the basis of HapMap chromosome 7 distributions. For each of the three derived alleles in Asians, the REHH was larger than 99th percentile.

Discussion

A key event in human population history is the dispersal of early humans from Africa to the other parts of the world with different climates, pathogens, and sources of food [40]. Current phenotypic differences between individuals/groups could be due to functional polymorphisms that facilitated survival of the ancestral populations in various geographic regions [41]. Therefore, resolving the underlying allelic architecture of our response to environmental factors and searching for signatures of natural selection in the human genome has been hailed as an attractive indirect strategy for implementation and interpretation of genetic analysis of xenobiotics/drug response and complex disorders [42,43]. Being one of the key components in the interaction between humans and xenobiotic environment, the ABCB1 transporter has a high priority for investigating phenotypic and evolutionary implications for variations of environmental response genes.

As different populations are subject to distinct selective environments (pressures), population-specific allelic architecture and population differentiation could be invoked to detect signal of Darwinian molecular selection at affected loci [33]. Differences in the ethnicity-dependent allelic spectrum for the three core SNPs in coding region of *ABCB1* locus have been noted in previous studies [18,39,44,45]. Few studies however, have focused on variations in the regulatory region of this gene. We observed significantly higher population differentiation (F_{ST} value) for a notable fraction of segregating sites, including not only the three well-known coding SNPs but also polymorphic sites (SNPs 12 and 17) in regulatory regions, suggesting that variations in the regulatory regions might also account for phenotypic differences of *ABCB1* across populations.

In addition to pronounced population differentiation, the case for positive selection on the *ABCB1* gene was further corroborated by the unusual extent and length of LD at this locus as revealed in the LRH test. As shown in Fig. 6, all three core SNPs (SNPs 33, 47, and 52) in the coding regions manifested significant departures from evolutionary neutrality in the Chinese, consistent with the previous study by Tang *et al.* [20]. We also observed that SNP 33 (e12/1236) departed from neutrality in the Europeans. We, however, did not observe the previously noted signal of positive selection on the ancestral allele of SNP 52 in Africans [20]. This could be because we used the empirical distribution rather than theoretical distributions under coalescent simulations utilized by Tang *et al.*, [20] who considered a series of population demography models. Notably, our novel findings was that, for the core SNPs in the 5′ flanking region/UTR, the common derived alleles of SNPs 12 and 17 are correlated with other alleles at nearby loci with a long-range association, and that these mutation events may have shaped the phylogenetically inferred star-like haplotype structure of the 5′ flanking region and conferred selective sweep in the *cis*-regulatory sequence of *ABCB1*.

A critical concern for our evolutionary inference of positive selection on both coding and regulatory regions of *ABCB1* locus is whether there are pharmacogenetic implications. For the three associated SNPs in coding

regions, the derived haplotype, e12/1236T-e21/2677T-e26/3435T, is reported to be associated with increased *in-vivo* P-gp activity and reduced initial oral drug absorption [39]. A pharmacogenetic investigation in acute myeloid leukemia patients demonstrated that the linked derived allele is associated with higher *ABCB1* mRNA expression in blast cell samples, and increased overall survival with a low probability of relapses [7]. Similarly, in a prospective study of cholesterol-lowering therapy, a central approach in the primary and secondary prevention of cardiovascular disease, carriers of the derived T-non-G-T haplotype also had better efficacy and are as well as lesser incidence of adverse drug reactions [46]. Furthermore, the derived haplotype e21/2677T-e26/3435T is associated with reduced risk of Parkinson's disease in two independent Chinese populations, putatively by conferring protection against xenobiotic insults at the blood–brain barrier [47,48]. These observations are compatible with the recent positive selection on the three common cSNPs and especially consistent with the pronounced signal of REHH in the Chinese population (Fig. 6).

Functional relevance of these variants, however, has been obscured by conflicting results from other molecular assays and population studies. The nonsynonymous mutation in exon 21 (Ala893Ser/Thr) was reported to result in P-gp with a cell surface distribution and function similar to wild-type transporter [49], and its derived allele was also reported to result in low expression of P-gp in human placenta, another important target tissue of *ABCB1* [6]. The derived e26/3435T allele was originally reported to correlate with lower P-gp expression and activity in the duodenum [3], and with increased susceptibility to ulcerative colitis [10,11]. The allele was also reported to affect mRNA secondary structure and decrease mRNA stability in the liver [50]. Notably, a recent report has revealed that the common synonymous SNP of 3435 C > T (rs1045642) in exon 26 has functional significance: the silent mutation can influence the dynamics of codon usage and affect the timing of cotranslational folding (translational kinetics and protein folding), and thereby result in alterations in protein conformation, substrate specificity, and P-gp function [51]. These discrepancies raise the possibility of differential gene regulation in different tissues and that some linked regulatory polymorphisms may account for the different regulatory kinetics.

It is becoming increasingly evident that the expression of *ABCB1* gene is under complex transcriptional control of many trans-acting factors such as p53, HIF-1, Sp1, NFKB, C/EBP, PXR, etc. [52,53]. Our study extends systemic identification of its common regulatory variants from the ATG start codon to the ~9 kb upstream PXR (a major xenobiotic receptor) response element, screening out five common regulatory variants (SNPs 10, 12, 16, 17, and 18) with MAF > 10% in at least one population. Although these common variants are not located within the typical binding sites for the above known transcriptional factors, SNP 12: 5′f/-4489 (A/G) and SNP 17: 5′UTR/-693(C/T) manifested striking population differentiation, derived alleles with very high frequency and significantly high REHH (SNP 12 in Europeans and SNP 17 in Africans), suggesting that positive selection may have operated on the regulatory regions of *ABCB1*. The SNP 17 is located just 8 bp downstream from the major transcription initiation site [54]. Its derived allele, −693T, is correlated with increased P-gp expression level in placenta [6] and colon [15], and is also associated with augmented expression of *ABCB1* mRNA both in colorectal adenocarcinomas and adjacent noncancerous colorectal tissues, possibly being an useful invasive marker predicting highly differentiated colorectal adenocarcinomas and thereby a better prognosis [55]. Although the functional role of SNP 17 in transcriptional activation of *ABCB1* gene remains unknown, a putative binding protein in HepG2 cells has been identified for the promoter fragment carrying the derived allele of −693T, which could not bind or bind only weakly to that carrying the ancestral −693C allele [15]. It is also interesting to note that the conversion from the ancestral −693G allele to the derived −693T allele breaches a GC site, which might potentially diminish the methylation status of the promoter CpG domain, and thereby have functional implications, given that the hypomethylation status of the *ABCB1* promoter region is associated with up-regulation, suppressed cell proliferation and increased cell differentiation, and pathogenesis and progression of diseases [55,56]. As to the SNP 12 presumably under positive selection, a clear functional role has not yet been established. The signature of positive selection on regulatory region of *ABCB1* gene indicates that some variants in this region may have phenotypic and clinical relevance, and that further functional analysis and association studies with relevant phenotypes are warranted.

It has become clear that populations and individuals have their own 'individual fingerprint' of unique allelic architecture coding the xenobiotic response system, and these genetic polymorphisms have functional relevance with respect to risk of environmentally caused toxicity and cancer [57,58]. It has been recently shown that positive selection has acted on xenobiotic response genes during the adaptation of humans to xenobiotic environment, such as *CYP1A2* [59] and *CYP3A5* [60] [phase I drug metabolizing enzyme (DME) gene], *NAT2* [61] (phase II DME gene), *ABCB1* [20] and *ABCC1* [62] (phase III transporter gene). It is interesting to note that positive selection on *cis*-regulatory variations has been reported for some xenobiotic response genes and disease candidate genes [59,62–64]. Notably, CYP3A4 (and CYP3A5), the most important and most abundant drug metabolizing enzymes in the liver, which are coded by the *CYP3A* gene cluster (7q21.3-q22.1) 119 kb away from the *ABCB1* locus, oxidize > 50% marketed drugs with a very

similar substrate spectrum as the ABCB1 transporter. In the liver, lung, kidney, and intestine, there is a close correlation between the expression of CYP3A4 and ABCB1, and their transcriptional factor, the orphan nuclear receptor (PXR) [65], which coordinately regulates drug metabolism and efflux through *trans*-activating the expression of the two genes [66]. The derived allele of 2677T in exon 21 of *ABCB1* is associated with enhanced constitutive CYP3A4 expression in the liver and intestine, suggesting that ABCB1 likely affect basal CYP3A4 expression by limiting the intracellular concentration of an endogenous regulator [67]. Intriguingly, the *CYP3A* locus might also be the target of positive selection. The nonexpressor allele of *CYP3A5*3* variant, which is likely to influence salt and water retention and risk for salt-sensitive hypertension, was under selective pressure resulting from an environmental variable correlated with latitude [60]. The shared positive selection on major xenobiotic response genes such as *ABCB1* and *CYP3A*, together with their closely linked genomic map, their finely coordinated *trans*-activation for xenobiotic disposition and clearance, and the overlapping tissue expression profile and substrate spectrum, may provide a good model in the context of systems biology of chemical homeostasis to decipher the integral and dynamic profile of the organization, regulation, and evolution of the xenobiotic disposition system.

In summary, this evolutionary population genetics study revealed complex signatures of natural selection on both coding and regulatory region of the human *ABCB1* gene, confirming that all the three common SNPs in coding regions may have cooperatively undergone (or have been undergoing) recent positive selection in the Chinese population, and observed pronounced manifestation of selective sweep for two common SNPs in its 5′ flanking regions in Europeans and Africans. Our findings point to the potential function of regulatory variants of *ABCB1*, and highlight that, in the context of evolving interplay between humans and xenobiotic environment, the evolutionary dynamics and mechanisms of transcriptional regulation may underlie the phenotypic variation of xenobiotic disposition system and varied predisposition to complex disorders, in which xenobiotics play a role.

Acknowledgements

The authors thank Professor Wei Huang at Chinese National Human Genome Center at Shanghai, Professor Jiujin Xu, and Professor Ruofu Du at Institute of Genetics and developmental biology, CAS, for their contribution to the collection and distribution of the Chinese DNA samples. They also thank Professor Yan Shen and Professor Zhijian Yao at the Chinese National Human Genome Center at Beijing for the support of the sequencing platform. They specially thank Professor Pengyuan Yang at the Institutes of Biomedical Sciences at Fudan University for his support of their work. Competing interest statement: The authors declare that they have no competing financial interests. This work was supported by the National High-Tech Research and Development Program of China (2002BA711A06), National Basic Research Program of China (2002CB512902), and Shanghai Science and Technology Developing Program (03DZ14024 and 04DZ14003). We acknowledge the technical support of the Supercomputing Institute of the University of Minnesota, Minneapolis.

References

1 Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 2001; **11**:1156–1166.
2 Gottesman MM, Fojo T, Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat Rev Cancer* 2002; **2**:48–58.
3 Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, Johne A, *et al.* Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci U S A* 2000; **97**: 3473–3478.
4 Thiebaut F, Tsuruo T, Hamada H, Gottesman MM, Pastan I, Willingham MC. Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. *Proc Natl Acad Sci U S A* 1987; **84**:7735–7738.
5 Schinkel AH, Wagenaar E, Mol CA, van Deemter L. P-glycoprotein in the blood–brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *J Clin Invest* 1996; **97**:2517–2524.
6 Tanabe M, Ieiri I, Nagata N, Inoue K, Ito S, Takahashi M, *et al.* Expression of P-glycoprotein in human placenta: relation to genetic polymorphism of the multidrug resistance (MDR)-1 gene. *J Pharmacol Exp Ther* 2001; **297**:1137–1143.
7 Illmer T, Schuler US, Thiede C, Schwarz UI, Kim RB, Gotthard S, *et al.* *MDR1* gene polymorphisms affect therapy outcome in acute myeloid leukemia patients. *Cancer Res* 2002; **62**:4955–4962.
8 Siddiqui A, Kerb R, Weale ME, Brinkmann U, Smith A, Goldstein DB, *et al.* Association of multidrug resistance in epilepsy with a polymorphism in the drug-transporter gene *ABCB1*. *N Engl J Med* 2003; **348**:1442–1448.
9 Fellay J, Marzolini C, Meaden ER, Back DJ, Buclin T, Chave JP, *et al.* Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 2002; **359**:30–36.
10 Ho GT, Soranzo N, Nimmo ER, Tenesa A, Goldstein DB, Satsangi J. *ABCB1/MDR1* gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum Mol Genet* 2006; **15**:797–805.
11 Schwab M, Schaeffeler E, Marx C, Fromm MF, Kaskas B, Metzler J, *et al.* Association between the C3435 T *MDR1* gene polymorphism and susceptibility for ulcerative colitis. *Gastroenterology* 2003; **124**:26–33.
12 Siegsmund M, Brinkmann U, Schaffeler E, Weirich G, Schwab M, Eichelbaum M, *et al.* Association of the P-glycoprotein transporter *MDR1* (C3435 T) polymorphism with the susceptibility to renal epithelial tumors. *J Am Soc Nephrol* 2002; **13**:1847–1854.
13 Marzolini C, Paus E, Buclin T, Kim RB. Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. *Clin Pharmacol Ther* 2004; **75**:13–33.
14 Chinn LW, Kroetz DL. *ABCB1* pharmacogenetics: progress, pitfalls, and promise. *Clin Pharmacol Ther* 2007; **81**:265–269.
15 Taniguchi S, Mochida Y, Uchiumi T, Tahira T, Hayashi K, Takagi K, *et al.* Genetic polymorphism at the 5′ regulatory region of multidrug resistance 1 (*MDR1*) and its association with interindividual variation of expression level in the colon. *Mol Cancer Ther* 2003; **2**:1351–1359.
16 Takane H, Kobayashi D, Hirota T, Kigawa J, Terakawa N, Otsubo K, *et al.* Haplotype-oriented genetic analysis and functional assessment of promoter variants in the *MDR1* (*ABCB1*) gene. *J Pharmacol Exp Ther* 2004; **311**:1179–1187.
17 Wang B, Ngoi S, Wang J, Chong SS, Lee CG. The promoter region of the *MDR1* gene is largely invariant, but different single nucleotide polymorphism haplotypes affect *MDR1* promoter activity differently in different cell lines. *Mol Pharmacol* 2006; **70**:267–276.
18 Tang K, Ngoi SM, Gwee PC, Chua JM, Lee EJ, Chong SS, *et al.* Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug

transporter gene locus in three ethnic Asian populations. *Pharmacogenetics* 2002; **12**:437–450.
19 Kroetz DL, Pauli-Magnus C, Hodges LM, Huang CC, Kawamoto M, Johns SJ, *et al*. Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. *Pharmacogenetics* 2003; **13**:481–494.
20 Tang K, Wong LP, Lee EJ, Chong SS, Lee CG. Genomic evidence for recent positive selection at the human *MDR1* gene locus. *Hum Mol Genet* 2004; **13**:783–797.
21 Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001; **27**:234–236.
22 Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, *et al*. Genetic relationship of populations in China. *Proc Natl Acad Sci U S A* 1998; **95**:11763–11768.
23 Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* 1996; **5**:182–187.
24 Iyengar S, Seaman M, Deinard AS, Rosenbaum HC, Sirugo G, Castiglione CM, *et al*. Analyses of cross species polymerase chain reaction products to infer the ancestral state of human polymorphisms. *DNA Seq* 1998; **8**:317–327.
25 Geick A, Eichelbaum M, Burk O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* 2001; **276**:14581–14587.
26 Wang H, Hao B, Zhou K, Chen X, Wu S, Zhou G, *et al*. Linkage disequilibrium and haplotype architecture for two ABC transporter genes (*ABCC1* and *ABCG2*) in Chinese population: implications for pharmacogenomic association studies. *Ann Hum Genet* 2004; **68**: 563–573.
27 Kent WJ. BLAT: the BLAST-like alignment tool. *Genome Res* 2002; **12**:656–664.
28 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003; **100**:9440–9445.
29 Lewontin RC. The detection of linkage disequilibrium in molecular sequence data. *Genetics* 1995; **140**:377–388.
30 Goddard KA, Hopkins PJ, Hall JM, Witte JS. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am J Hum Genet* 2000; **66**:216–234.
31 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**:978–989.
32 Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**:37–48.
33 Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002; **12**: 1805–1814.
34 Weir BS, Cockerham CC. Estimating F-statistics for the population structure. *Evolution* 1984; **38**:1358–1370.
35 Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, *et al*. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**:832–837.
36 Nei M. Genetic variability in natural populations. In: Tateno Y, Gojobori T, editors. Molecular population genetics and evolution. Amsterdam/New York: Elsevier; 1975.
37 Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**:585–595.
38 Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, *et al*. Genetic structure of human populations. *Science* 2002; **298**: 2381–2385.
39 Kim RB, Leake BF, Choo EF, Dresser GK, Kubba SV, Schwarz UI, *et al*. Identification of functionally variant *MDR1* alleles among European Americans and African Americans. *Clin Pharmacol Ther* 2001; **70**: 189–199.
40 Jin L, Su B. Natives or immigrants: modern human origin in East Asia. *Nat Rev Genet* 2000; **1**:126–133.
41 Di Rienzo A, Hudson RR. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet* 2005; **21**:596–601.
42 Bamshad M, Wooding SP. Signatures of natural selection in the human genome. *Nat Rev Genet* 2003; **4**:99–111.
43 Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, *et al*. Positive natural selection in the human lineage. *Science* 2006; **312**: 1614–1620.
44 Ameyaw MM, Regateiro F, Li T, Liu X, Tariq M, Mobarek A, *et al*. *MDR1* pharmacogenetics: frequency of the C3435T mutation in exon 26 is significantly influenced by ethnicity. *Pharmacogenetics* 2001; **11**:217–221.
45 Cascorbi I, Gerloff T, Johne A, Meisel C, Hoffmeyer S, Schwab M, *et al*. Frequency of single nucleotide polymorphisms in the P-glycoprotein drug transporter MDR1 gene in white subjects. *Clin Pharmacol Ther* 2001; **69**:169–174.
46 Fiegenbaum M, da Silveira FR, Van der Sand CR, Van der Sand LC, Ferreira ME, Pires RC, *et al*. The role of common variants of *ABCB1*, *CYP3A4*, and *CYP3A5* genes in lipid-lowering efficacy and safety of simvastatin treatment. *Clin Pharmacol Ther* 2005; **78**:551–558.
47 Lee CG, Tang K, Cheung YB, Wong LP, Tan C, Zhao Y, *et al*. MDR1, the blood–brain barrier transporter, is associated with Parkinson's disease in ethnic Chinese. *J Med Genet* 2004; **41**:e60.
48 Tan EK, Chan DK, Ng PW, Woo J, Teo YY, Tang K, *et al*. Effect of *MDR1* haplotype on risk of Parkinson disease. *Arch Neurol* 2005; **62**:460–464.
49 Kimchi-Sarfaty C, Gribar JJ, Gottesman MM. Functional characterization of coding polymorphisms in the human *MDR1* gene using a vaccinia virus expression system. *Mol Pharmacol* 2002; **62**:1–6.
50 Wang D, Johnson AD, Papp AC, Kroetz DL, Sadee W. Multidrug resistance polypeptide 1 (*MDR1*, *ABCB1*) variant 3435C>T affects mRNA stability. *Pharmacogenet Genom* 2005; **15**:693–704.
51 Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, *et al*. A 'silent' polymorphism in the *MDR1* gene changes substrate specificity. *Science* 2007; **315**:525–528.
52 Labialle S, Gayet L, Marthinet E, Rigal D, Baggetto LG. Transcriptional regulators of the human multidrug resistance 1 gene: recent views. *Biochem Pharmacol* 2002; **64**:943–948.
53 Scotto KW. Transcriptional regulation of ABC drug transporters. *Oncogene* 2003; **22**:7496–7511.
54 Kohno K, Sato S, Uchiumi T, Takano H, Kato S, Kuwano M. Tissue-specific enhancer of the human multidrug-resistance (*MDR1*) gene. *J Biol Chem* 1990; **265**:19690–19696.
55 Koyama T, Nakamura T, Komoto C, Sakaeda T, Taniguchi M, Okamura N, *et al*. *MDR1* T-129C polymorphism can be predictive of differentiation, and thereby prognosis of colorectal adenocarcinomas in Japanese. *Biol Pharm Bull* 2006; **29**:1449–1453.
56 Enokida H, Shiina H, Igawa M, Ogishima T, Kawakami T, Bassett WW, *et al*. CpG hypermethylation of *MDR1* gene contributes to the pathogenesis and progression of human prostate cancer. *Cancer Res* 2004; **64**:5956–5962.
57 Nebert DW, Dieter MZ. The evolution of drug metabolism. *Pharmacology* 2000; **61**:124–135.
58 Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; **286**:487–491.
59 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**:528–542.
60 Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. *CYP3A* variation and the evolution of salt-sensitivity variants. *Am J Hum Genet* 2004; **75**:1059–1069.
61 Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, *et al*. Deciphering the ancient and complex evolutionary history of human arylamine *N*-acetyltransferase genes. *Am J Hum Genet* 2006; **78**:423–436.
62 Wang Z, Wang B, Tang K, Lee EJ, Chong SS, Lee CG. A functional polymorphism within the *MRP1* gene locus identified through its genomic signature of positive selection. *Hum Mol Genet* 2005; **14**:2075–2087.
63 Ding K, Kullo IJ. Molecular evolution of 5′ flanking regions of 87 candidate genes for atherosclerotic cardiovascular disease. *Genet Epidemiol* 2006; **30**:557–569.
64 Hahn MW. Detecting natural selection on *cis*-regulatory DNA. *Genetica* 2007; **129**:7–18.
65 Miki Y, Suzuki T, Tazawa C, Blumberg B, Sasano H. Steroid and xenobiotic receptor (SXR), cytochrome P450 3A4 and multidrug resistance gene 1 in human adult and fetal tissues. *Mol Cell Endocrinol* 2005; **231**:75–85.
66 Synold TW, Dussault I, Forman BM. The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat Med* 2001; **7**: 584–590.
67 Lamba J, Strom S, Venkataramanan R, Thummel KE, Lin YS, Liu W, *et al*. *MDR1* genotype is associated with hepatic cytochrome P450 3A4 basal and induction phenotype. *Clin Pharmacol Ther* 2006; **79**:325–338.