ORIGINAL ARTICLE

Analysis of *CYP3A4* genetic polymorphisms in Han Chinese

Qing Zhou^{1,3}, Xiaomin Yu^{1,2,3}, Chang Shu^{1,3}, Yimei Cai¹, Wei Gong¹, Xumin Wang¹, Duen-mei Wang¹ and Songnian Hu¹

Our study aimed to comprehensively investigate the genetic polymorphisms of *CYP3A4* in Han Chinese. We sequenced the gene regions of *CYP3A4*, including its promoter, exons, surrounding introns and 3' untranslated region (3'UTR), from 100 unrelated healthy Han Chinese individuals. We detected 11 SNPs, three of which are novel. According to *in silico* functional prediction of novel variants, 20148 A>G in exon 10, resulting in substitution of Tyr319 with Cys (*CYP3A4**21), may induce dramatic alteration of protein conformation, and 26908 G>A in 3'UTR may disrupt post-transcriptional regulation. We identified five alleles in Han Chinese, the allele frequencies of *CYP3A4**1, *5, *6, *18 and *21 are 97, 0.5, 1, 1 and 0.5%, respectively. Haplotype inference revealed 14 haplotypes, of which the major haplotype *CYP3A4**1A constitutes 59% of the total chromosomes. We also examined the possible role of natural selection in shaping the variation of *CYP3A4* and confirmed a trend, consistent with the action of positive selection. We systematically screened the genetic polymorphisms of *CYP3A4* in Han Chinese, highlighted possible functional impairment of the novel allele and summarized the distinct allele and haplotype frequency distribution, with an emphasis on detecting the footprint of recent positive selection on the *CYP3A4* gene in Han Chinese.

Journal of Human Genetics (2011) 56, 415–422; doi:10.1038/jhg.2011.30; published online 17 March 2011

Keywords: *CYP3A4*; Han Chinese; genetic polymorphisms

INTRODUCTION

Cytochrome P450 3A4 (*CYP3A4*, MIM 124010), together with *CYP3A5* (MIM 605325), *CYP3A7* (MIM 605340) and *CYP3A43* (MIM 606534), are members of the *CYP3A* subfamily forming a gene cluster on chromosome 7q22. *CYP3A4*, as the most abundantly expressed P450 isoenzyme in human liver, contributes to the oxidative metabolism of approximately 50% of clinically-used drugs, with a broad range of substrate specificity. For instance, *CYP3A4* is involved in the oxidation of certain antibiotics, calcium channel blockers, antidepressants, immunosuppressants, HMG-CoA reductase inhibitors (HMGs), antihistamines and protease inhibitors, as well as some endogenous steroids, such as cortisol, testosterone and estradiol.

The enzyme activity of *CYP3A4* shows wide interindividual variability (up to 60-fold),¹ resulting in therapeutic failure, unpredictable adverse side effects or severe drug toxicity. The enormous variation observed in drug metabolism is mainly due to the combined effect of genetic polymorphisms, regulation of gene expression and interaction with drugs or environmental chemicals.² Genetic polymorphisms, accounting for as much as 90% of interindividual variability in *CYP3A4* activity,³ are therefore of clinical value in predicting an individual's ability to respond to certain therapeutic agents. Moreover,

CYP3A4 polymorphisms can also assist prediction of disease predisposition.⁴ *CYP3A4**1B allele (*CYP3A4*-V, rs2740574), a –392A>G transition in the promoter region, has been reported to be significantly associated with HIV infection;⁵ this allele can also result in an increased risk of estrogen negative breast cancer⁶ and prostate cancer;^{4,7,8} *CYP3A4**4 allele with the substitution of Ile118Val is significantly associated with the prevalence of type 2 diabetes mellitus;⁹ *CYP3A4**18 allele which contains a Leu293Pro substitution has also been shown to be associated with low bone mineral density and increased prostate cancer risk.^{4,10}

Given these important clinical implications, genetic polymorphisms of *CYP3A4* are extensively studied in different ethnic populations and exhibit remarkable interethnic difference in distribution and frequency spectrum. Thus far, a total of 20 allelic variants have been identified and characterized (<http://www.imm.ki.se/CYPalleles/cyp3a4.htm>), eight of which lead to reduced catalytic activity or functional deficiency *in vitro*. The most prevalent variant is *CYP3A4**1B, with allele frequency ranging from 0~4% in Asians and Europeans to 82% in Africans.^{11,12} The elimination of *CYP3A4**1B in non-African populations has been suggested to be a consequence of selective advantage.¹³ This variant might moderately enhance *CYP3A4* gene expression, by

¹CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China and ²Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

³These authors contributed equally to this work.

Correspondence: Dr Q Zhou or S Hu, CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, China.

E-mail: zhouqingnwu@gmail.com or husn@big.ac.cn

Received 27 July 2010; revised 9 January 2011; accepted 17 February 2011; published online 17 March 2011

virtue of reduced binding ability of a transcriptional repressor.¹⁴ CYP3A4*18 is the most frequent variant in Asian populations (about 1% in Chinese, 1.3% in Japanese and 2.1% in Malaysian),^{15–18} but absent both in Caucasians and Africans. One recently discovered rare variant, CYP3A4*20—with a premature stop codon yielding a truncated protein and completely devoid of functional activity—occurs less than 0.06% in Caucasians¹⁹ and has not been reported in Asians or Africans.

Han Chinese is the largest single ethnic group in the world, constituting over 92% of the population of mainland China and about 19% of the global human population. Han Chinese has significant internal difference as a result of ongoing demographic expansion, accompanied by assimilation of various regional ethnicities and tribes,^{20–23} which can be divided into two genetically differentiated groups, northern Han and southern Han, separated by the Yangtze River.^{20–22} Hitherto, only a few studies focused on single-nucleotide polymorphism (SNP) prevalence analysis of CYP3A4 in Han Chinese. Hsieh *et al.*²⁴ firstly screened SNPs in 102 Taiwan Chinese by PCR-single-strand conformation polymorphism and restriction fragment length polymorphism. Nevertheless, Taiwan Chinese are not appropriate to serve as a good representative for Han Chinese in mainland. Then, Wen and Liu *et al.*^{11,25} surveyed polymorphisms of CYP3A4 in Chinese by genotyping, a method that cannot detect novel alleles which might be specific to Han Chinese. Recently, Du *et al.*²⁶ used direct sequencing, but with a small amount of sample (20 Han Chinese). Therefore, the genetic polymorphisms of CYP3A4 gene in Han Chinese are far from being studied thoroughly, and more investigation of CYP3A4 polymorphisms in Han Chinese is imperatively needed.

Considering the ethnic-specific distribution of genetic polymorphisms and their important roles in optimizing therapeutic efficiency and modifying clinical presentation, we screened CYP3A4 variants in 100 Han Chinese individuals by direct sequencing. All of our samples are collected in Beijing, which is a good representative of northern Han. And we also estimated allele and genotype frequencies, reconstructed haplotypes and analyzed functional activity of novel polymorphisms.

MATERIALS AND METHODS

Subjects and genomic DNA extraction

A total of 100 cell lines of unrelated healthy individuals (51 male and 49 female) of Han Chinese ancestry were obtained from two sources: Han Chinese in Beijing from the HapMap Project²⁷ (42 cell lines) and blood samples of unrelated Han Chinese individuals in Beijing (remaining 58 cell lines).²⁸ The 58 participants (30 male and 28 female; mean age 26) were judged to be in healthy condition, based on a physical examination. Before blood and information collection, the participants signed informed consent forms. To ensure their Han Chinese origin, they were interviewed regarding their ancestral status for up to three previous generations. The project was approved by the Ethnical Committee at the Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

We used salting-out method to extract genomic DNA from Epstein-Barr virus-transformed lymphoblastoid cells, established from peripheral blood.²⁹

PCR amplification and sequencing

To screen the SNPs in CYP3A4, we sequenced gene regions, including its promoter, exons, surrounding introns and 3' untranslated region (3'UTR). The PCR was performed in a volume of 15 µl, containing 30 ng of genomic DNA, 1×PCR TaKara buffer (Mg²⁺ plus), 2.5 pmol of each primer, 25 pmol dNTPs and 1 unit of Taq DNA polymerase (TaKara Bio, Dalian, China). The cycling program was initiated with heat denaturation at 95 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 55–60 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 10 min, using a thermal cycler (GeneAmp PCR System

9700, Applied Biosystems, Foster City, CA, USA). The PCR products were enzymatically purified using shrimp alkaline phosphatase (SAP; Amersham Biosciences, NJ, USA), followed by incubation at 37 °C for 60 min, after which the shrimp alkaline phosphatase enzyme was inactivated at 85 °C for 15 min. Then the purified PCR products were subjected to direct sequencing by using the DYEnamic ET terminator cycle sequencing kit (GE Healthcare, Chalfont St Giles, UK) on the ABI Prism 3730xl DNA Analyzer (Applied Biosystems). All the PCR and sequencing primers were listed in Supplementary Table S1. Singleton mutations, occurring only once in 100 samples, were verified with a new PCR amplification, followed by bidirectional sequencing with coverage of the SNP.

Data analysis

We used Phred, Phrap, Consed and Polyphred programs (University of California; <http://elcapitan.ucsd.edu/hyper/polyphred.usage.html>) for base calling, quality assessment and polymorphism determination from DNA sequencing.^{30–33} Genetic polymorphisms of CYP3A4 gene were named according to the genomic reference sequence AF280107.1.

We employed Haploview4.1 (BROAD Institute; <http://www.broad.mit.edu/mpg/haploview/>) to estimate allele frequencies, test Hardy–Weinberg equilibrium and analyze linkage disequilibrium (LD) structure.³⁴ LD structure was displayed by GOLD heatmap color scheme,³⁵ and LD blocks were defined by the 'solid spine of LD' algorithm,³⁴ with a minimum D' value of 0.9.

We used PHASE2.1 (University of Washington; <http://www.stat.washington.edu/stephens/phase.html>) for haplotype reconstruction and haplotype frequency estimation.^{36,37}

We used PITA,³⁸ an miRNA target prediction algorithm based on hybridization energy and site accessibility, to predict potential miRNAs that target CYP3A4 and harbor certain SNP in their binding sites.

We combined two prediction programs, PANTHER (the National Institute of General Medical Sciences; <http://www.pantherdb.org/>)^{39,40} and PolyPhen (Brigham and Women's Hospital, Harvard University; <http://genetics.bwh.harvard.edu/pph/>),⁴¹ to predict possible impact of novel nonsynonymous SNP on protein function in the consideration of both sequence similarity and structural information. PANTHER subPSEC scores < −3.0 ($P_{\text{deleterious}} > 0.5$) and PolyPhen PSIC score difference > 2.0 indicate mutations as probably damaging protein function.^{39–41} PyMOL 0.99rc6 (Delano Scientific LLC; <http://www.py-mol.org/>) was utilized for visualization of CYP3A4 three-dimensional structure (PDB ID: 2J0D),⁴² which was downloaded from the RCSB Protein Data Bank (PDB; <http://www.rcsb.org/pdb/home/home.do>). CASTp (Computed Atlas of Surface Topography of proteins, <http://sts-fw.bioengr.uic.edu/castp/calculation.php>) was employed for protein pocket identification and measurement of pocket volume change influenced by novel amino acid substitution.⁴³

We performed multiple neutrality tests using DnaSP v5 (<http://www.ub.edu/dnasp/>)⁴⁴ to detect evidence of positive selection on the CYP3A4 gene. Tajima's *D* (TD),⁴⁵ Fu and Li's *F* (FL-F)⁴⁶ (with a chimpanzee sequence as outgroup), and Fay and Wu's *H* (FW-H)⁴⁷ (also with outgroup) were calculated by a sliding window of 25 bp to test for deviations from a neutral equilibrium frequency distribution at all loci. We additionally investigated evidence for natural selection by examining the integrated haplotype score (iHS)⁴⁸ that was obtained from the program haplotter (<http://haplotter.uchicago.edu/>).

RESULTS AND DISCUSSION

Allele and genotype frequency of CYP3A4 in Han Chinese

We detected 11 genetic variants in Han Chinese, with allele frequencies complied well with the Hardy–Weinberg equilibrium ($P > 0.05$), which include two in promoter, four in exons, three in introns and two in 3'UTR (Table 1). Of the 11 genetic variants, three novel polymorphisms were identified, they are 4231 A>C in intron2, 20148 A>G in exon 10 and 26908 G>A in 3'UTR. Most of the CYP3A4 polymorphisms in Han Chinese are relatively rare: nine SNPs occur below a frequency of 5%, except two commonly SNPs, rs28969391 (32.5%) and rs2242480 (25%). rs28969391 is a 26707T deletion in the 3'UTR with little chance to be involved in the post-transcriptional regulation, as it is situated in a region not targeted by

Table 1 Genetic polymorphisms of *CYP3A4* in Han Chinese

#	SNP ID	Position ^a	Region	Allele	Effect	Flanking sequence	HWpval ^b	MAF
1	rs28988569	-630 A>G	Promoter	/	/	AAAGAGGACA R TAGGATTGCA	1	0.005
2	rs2740574	-392 A>G	5' UTR	<i>CYP3A4</i> *1B	/	GACAAGGGCA R GAGAGAGGCG	1	0.005
3	new1	4231 A>C	Intron2	/	/	CACCCAGAAC M TGGCAATGTTT	1	0.01
4	rs55901263	15702 C>G	Exon7	<i>CYP3A4</i> *5	P218R	TTTTTGGATC S ATTCTTTCTC	1	0.005
5	rs4646438	17661 →A	Exon9	<i>CYP3A4</i> *6	277 Frame shift	TGATGATTGA N CTCTCAGAAT	1	0.01
6	rs56153749	17815 AT>-	Intron9	/	/	AGAACGACAC N GTTTGAATGT	1	0.025
7	rs28371759	20070 T>C	Exon10	<i>CYP3A4</i> *18	L293P	CTGTCCGATC Y GGAGCTCGTG	1	0.01
8	new2	20148 A>G	Exon10	<i>CYP3A4</i> *21	Y319C	TTCATTATGT R TGAAGTGGCC	1	0.005
9	rs2242480	20230 G>A	Intron10	/	/	TGAGTGGATG R TACATGGAGA	0.377	0.25
10	rs28969391	26707 T>-	3' UTR	/	/	GCCTACATGGT N GAAACCCCAT	1	0.325
11	new3	26908 G>A	3' UTR	/	/	TTTGAAGCAC R TACATTTTGTA	1	0.005

Abbreviations: MAF, minor allele frequency; SNP, single-nucleotide polymorphism; UTR, untranslated region.

^aPosition of genetic variants was named according to the genomic reference sequence AF280107.1.^bHWpval, Hardy-Weinberg equilibrium P-value, which is the probability that its deviation from H-W equilibrium could be explained by chance.

any known human miRNAs according to PITA prediction. Indeed, SNPs in miRNA target sites rarely reach such a high frequency, as they are usually subjected to more stringent functional constraints and would be removed by purifying selection eventually.^{49,50} rs2242480 is located in the 10th intron, 12 nt downstream of the donor site (IVS10+12). This polymorphism has been reported to increase the lipid-lowering efficacy of atorvastatin,⁵¹ although it remains unclear whether splicing efficiency would be influenced in the presence of a G>A mutant, adjacent to 5' splice site.

Moreover, five alleles were identified in Han Chinese, they are *CYP3A4**1, *5, *6, *18 and *21, with occurrence of 97, 0.5, 1, 1 and 0.5%, respectively (Table 2). We further compared allele frequencies of *CYP3A4* among different ethnic groups and demonstrated remarkable interpopulation difference, with regard to allelic distribution (Table 3). *CYP3A4**1B, the most frequent allele in Africans (82%),¹² has significantly lower incidence in Caucasians (4%)^{12,52} and is near extinction in Asians.^{53,54} This allele is present in about 0.5% of Han Chinese population, a higher incidence than in Malays¹⁸ and Japanese.⁵⁴ The top two frequent alleles estimated in Chinese by Du *et al.*, *CYP3A4**15 (16%) and *14 (7%), were neither detected in our samples. The significant differences of allele frequencies between the two investigations may result from involvement of ethnic minorities in Du's samples (consisting of 20 Han, 30 She and 10 Dong subjects) and substantial genetic diversity within Han Chinese. In our Han Chinese population, *CYP3A4**18 remains the most common coding variant, which is in agreement with previous investigations in Asian populations.^{15-18,54}

Among seven observed genotypes (Table 2), the wild type absolutely predominates in Han Chinese, accounting for approximately 94% of the populations; the other six genotypes—heterozygous *1/*1B, *1/*5, *1/*6, *1/*18, *1/*21 and *6/*18—each contributes an insignificant percentage (1%). The low prevalence of mutant genotypes indicates relative infrequency of poor metabolizer in Han Chinese, which is in line with previously reported low-frequency of *CYP3A4* poor metabolizer in Asians.⁵⁵

Haplotypes of *CYP3A4* in Han Chinese

Haplotypes were inferred with 11 detected SNPs by the PHASE program. Our reconstruction of haplotypes strengthened previous findings that a few common haplotypes of *CYP3A4* occupy a huge proportion of the total chromosomes examined,⁵⁴ whereas a large subset of haplotypes carries low levels of variation.⁵⁶ Of 14-phased

Table 2 Frequencies of *CYP3A4* alleles and genotypes in Han Chinese

Alleles	
*1	97%
*5	0.5%
*6	1%
*18	1%
*21	0.5%
Genotypes	
*1/*1	94%
*1/*1B	1%
*1/*5	1%
*1/*6	1%
*1/*18	1%
*1/*21	1%
*6/*18	1%

haplotypes (Figure 1), two major haplotypes, *CYP3A4**1A and *1v, constitute 59 and 19.5% of the total haplotype diversity, respectively, whereas 11 haplotypes occur at a frequency below 5%, bearing no more than three variants.

Seven haplotypes labeled with small alphabetical letters were newly assigned (Figure 1). Haplotype inference revealed the involvement of novel polymorphisms in *CYP3A4**1u, *1w and *21: haplotype of *1u contains novel substitution of A>C at nucleotide position 4231; the *1w haplotype is composed of 20230G>A, 26707T>- and new variant 26908G>A; and novel allele *21 harbors nonsynonymous SNP 20148A>G in exon 10, resulting in Cys taking the place of Tyr at amino acid 319 (Y319C).

According to haplotype combinations, *CYP3A4* haplotypes containing the IVS10+12G allele (such as *1A) show strong linkage with *CYP3A5**3, whereas 88% of the *CYP3A4* haplotypes with the IVS10+12A allele (such as *1G) are tightly linked to *CYP3A5**1 in Japanese.⁵⁴ The close linkage of haplotypes between *CYP3A4* and *CYP3A5* was also demonstrated in Chinese populations,⁵⁷ which indicates that genotyping at the IVS10+12 position in *CYP3A4* might be easy to predict the existence of *CYP3A5**3.⁵⁴ We observed about 75% of *CYP3A4* haplotypes carry the IVS10+12G allele, implying a large proportion of *CYP3A5**3 in the Han Chinese population.

Table 3 Ethnic differences in *CYP3A4* allele frequencies

Ethnicity	N	*1B	*2	*3	*4	*5	*6	*10	*11	*13	*14	*15	*16	*17	*18	*21	Ref
Han Chinese	100	0.5	/	/	/	0.5	1	/	/	/	/	/	/	/	1	0.5	This study
Chinese (Han, She, Dong)	60	1	/	/	/	/	/	1	/	1	7	16	/	2	1	/	26
Taiwan Chinese	102	/	/	/	2.9	1.9	0.5	/	/	/	/	/	/	/	/	/	24
Japanese	416	0	/	/	/	/	0.1	/	0.2	/	/	/	1.4	/	2.8	/	54
Malaysian	121	/	/	/	0	0	/	/	/	/	/	/	/	/	2.1	/	18
European	95	4	0	2	/	/	/	/	/	/	/	/	/	/	/	/	12
Portuguese	100	4	5	/	/	/	/	/	/	/	/	/	/	/	/	/	52
African	95	82	0	0	/	/	/	/	/	/	/	/	/	/	/	/	12

Figure 1 Haplotypes of *CYP3A4* in Han Chinese. AF280107.1 is defined as the genomic reference sequence of *CYP3A4* wild-type allele. Blue and yellow cells show common alleles and rare alleles, respectively. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

The novel nonsynonymous SNP altering enzyme function

The substitution of Cys for Tyr in codon 319 (Y319C) lies on the helix I,⁴² a domain which flanks the heme group and contains a substrate recognition site (SRS4).⁵⁸ As shown in the Supplementary Table S2, Tyr319 is a highly conserved residue in the cytochrome P450 family throughout eukaryote evolution from nematode to human. In order to predict the functional impact of the amino acid mutation, we used two *in silico* prediction algorithms, PolyPhen and PANTHER, which combine protein structure and sequence evolution analysis to achieve a more comprehensive result. According to PANTHER prediction, the Y319C variant stands a good chance of impairing the protein function with low subPSEC score (−6.90091) and significantly high $P_{\text{deleterious}}$ value (0.98018). Likewise, the variant was also considered to be probably damaging, because of the enlarging of a cavity based on PolyPhen analysis (PSIC score difference=3.453). The CASTp analysis provided us with a better understanding of the fatal transformation in cavity resulting from the nonsynonymous SNP (Figures 2a and b). The wild-type cavity (Pocket ID=144) covers 89.4 Å² of surface area and has a volume of 135.08 Å³, which is composed of atoms from 10 residues: Tyr319, Glu320, Leu366, Phe367, Tyr399, Arg403, Glu412, Pro474, Leu475 and Leu477 (Figure 2a). Surprisingly, the replacement of Tyr319 with Cys yields an increase in surface area of 270.1 Å² and a three-fold expansion in volume (Pocket ID=158, surface area=359.5 Å², volume=376.86 Å³). The dramatic change of cavity in the mutant is due to recruitment of atoms from additional 11 residues: Thr323, Pro467, Glu470, Thr471, Gln472, Ile473, Lys487, Pro488, Val489, Val490 and Leu491 (Figure 2b).

Furthermore, Tyr319 has a critical role in maintaining perfect position of the loop between β3-3 and β4-1. This β-β loop is held in place through four extra loop hydrogen bonds (Figure 2c). Two of them are formed from Tyr319 to the oxygen atom of Gln472 and the nitrogen atom of Leu475, with bond lengths of 3.01 and 3.05 Å, respectively, whereas the other two are donated by the bonding interactions between Ile369 and Leu483, and between Leu172 and Val489, respectively. The three-dimensional structure of the β-β loop in Figure 2c showed that the center of the loop (L477-P485) is available for approaching ligands, termed as a ligand-accessible region.⁵⁹ Both X-ray structure analysis and molecular modeling for *CYP3A4* have revealed ligand inducible of heme pocket and extreme flexibility of ligand-accessible regions, so that diverse substrates can be accommodated in active site by adopting multiple ligand-bound conformations.^{42,59} From this point of view, Tyr319 together with Leu172 and Ile369, might act as anchors of the β-β loop, not only keeping the flexibility of the loop for its participation in the regulation of the heme pocket size when interacting with various drugs, but also restrict the freedom of the loop to avoid ligand-accessible region losing connection from ligands. Thus, once Tyr319 is replaced by Cys, its interaction with Gln472 and Leu475 will be abolished (Figure 2d), leaving only two hydrogen bonds unable to adjust and stabilize the β-β loop.

Therefore, the highly conserved residue in important domain, together with striking alteration in cavity configuration as a result of this single amino acid substitution, raises the possibility of producing malfunction protein by the new variant.

The novel SNP in 3'UTR influencing miRNA post-transcriptional regulation

We detected a new polymorphism, which is a G-to-A transition at nucleotide position 26908 located in the 3'UTR region. Accumulating evidence revealed that variants in the 3'UTR could be functional impairment if they happen to be located in miRNA-binding sites,^{49,60} because SNPs will interfere with miRNA function and result in loss of miRNA-mediated regulation of drug metabolism genes. For this reason, we employed PITA to predict liver-expressed miRNAs,^{61–65} whose binding sites harbor this polymorphism site. A total of 14 miRNAs containing complementary site of the SNP in their seed sequences meet the threshold of interaction energy for miRNA:target pairing ($\Delta\Delta G \leq -5$ kcal/mol) (Table 4, Supplementary Figure S1). The strengthened or declined miRNA binding caused by SNP variations may have consequences of corresponding downregulation or upregulation of mRNA stability and translation.^{66,67} Therefore, the novel polymorphism in predicted miRNA target sites may be a candidate causal variant responsible for altering expression of *CYP3A4* protein.

Figure 2 The Tyr319Cys substitution enlarges cavity configuration (a, b) and disrupts two hydrogen bonds with Gln472 and Leu475 on the loop between β 3-3 and β 4-1 (c and d). (a) The pocket in white is formed by the atoms from 10 residues: Tyr319, Glu320, Leu366, Phe367, Tyr399, Arg403, Glu412, Pro474, Leu475, Leu477. (Pocket ID=144, surface area=89.4Å², volume=135.08Å³). And the atoms from Tyr319 are shown in yellow. (b) The pocket in green is formed by the atoms from 21 residues: Cys319, Glu320, Thr323, Leu366, Phe367, Tyr399, Arg403, Glu412, Pro467, Glu470, Thr471, Gln472, Ile473, Pro474, Leu475, Leu477, Lys487, Pro488, Val489, Val490, Leu491 (Pocket ID=158, surface area=359.5Å², volume=376.86Å³). And the atoms from Cys319 are shown in blue. (c) Tyr319 is shown in green. (d) Cys319 is shown in yellow.

Table 4 Prediction of novel SNP influencing miRNAs targeting

miRNA ID	miRNA sequences	Wild-type score ($\Delta\Delta G$: kcal/mol)	Mutation score ($\Delta\Delta G$: kcal/mol)	miRNAs expression in liver ⁶²			
				Tumor	Normal	Tumor	Normal
hsa-miR-210	cugugc gugugacagcgccuga	-13.72	-8.1	247.5399	141.4969	77.38133	270.0463
hsa-miR-517b	ucgugcau ccuuuagaguguu	-6.08	/	11.55175	11.29866	4.894332	8.796205
hsa-miR-493*	uuguaca ugguaggcuuucuu	-6.12	/	27.78663	24.98089	243.9701	24.93878
hsa-miR-22*	aguucuu caguggcaagcuuu	-9.34	-9.4	23.97158	26.48348	10.42277	18.12006
hsa-miR-520f	aagugcu uccuuuagaggguu	-12.88	-13.14	4.989946	11.31383	4.908601	8.818082
hsa-miR-302d	uaagugcu uccauguuagagugu	-11.68	-11.99	4.083453	9.26377	4.013037	13.53896
hsa-miR-373	gaagugcu ucgauuuuggggugu	-14.87	-15.18	4.986349	11.28935	4.897624	8.810456
hsa-miR-218	uugugcuu gaucuaaccaugu	-7.18	-8.14	38.21249	24.07236	17.54524	11.04794
hsa-miR-455-5p	uaugugccu uuggacuacaucg	-5.82	-6.53	164.741	131.1441	109.4137	115.2389
hsa-miR-636	ugugcuug cucgucccgccgca	-15.15	-15.21	3.521987	7.982176	3.464479	6.21753
hsa-miR-567	aguauguu cuccaggacagaaac	-9.28	-10.03	3.523162	7.98256	3.460543	6.218029
hsa-miR-297	auguaugu gugcaugugcaug	-10.22	-11	3.522004	7.982656	3.462658	6.225475
hsa-miR-632	gugucugc uuccuguggga	/	-7.32	4.085454	9.253369	4.012449	7.207448
hsa-miR-223*	cguguauu ugacaagcugaguu	/	-8.47	3.528426	8.017793	3.462842	6.241723

Seed sequences of miRNAs are highlighted in red and the complementary SNP sites in seed sequences are marked with green background. miRNA* is the antisense of miRNA. Normalized miRNAs expression data in liver (tumor/normal samples) was cited from Navon et al.⁶² A full color version of this table is available at the *Journal of Human Genetics* journal online.

Footprint of positive selection on CYP3A4 in Han Chinese

In non-African populations, the CYP3A locus was reported under strong positive selection,^{13,56} leading to sudden loss of neutral genetic variation when beneficial alleles arise nearby. The near-fixation frequency of derived alleles of both CYP3A4*1B and CYP3A5*3, coupled with their unusually geographic patterns correlating with latitude, were thought to be results of salt homeostasis and rickets, the underlying selecting factors in European, Caucasians and Asians.^{13,56} In Han Chinese, we also observed a number of characteristic features

in sequence variations: low polymorphism levels with a fraction of high-frequency-derived alleles, a marked skew toward rare variants and dramatic differences in allele frequency between Han Chinese and Africans. All of these distinct traits give us a clue towards selective sweep on CYP3A4 locus in Han Chinese. Thereby, both TD⁴⁵ and FL-F⁴⁶ tests were employed to test the observed mutation frequency distribution for deviations from neutral expectations. It is well known that the expected values of TD and FL-F are equal to zero under neutral equilibrium model. However,

Figure 3 The distribution of Tajima's D (TD) (a), Fu and Li's F (FL-F) (b), and Fay and Wu's H (FW-H) (c) along the *CYP3A4* gene.

Figure 4 Decay of haplotypes (above half) and decay of haplotype homozygosity (below half) in the population of Africans (YRI) (a), East Asians (ASN) (b) and Europeans (CEU) (c) using rs2242480 as core single-nucleotide polymorphism (SNP) in a region of 1 Mb. The allele marked in blue is the ancestral allele and in red is derived. For the decay of haplotypes, horizontal lines are haplotypes and SNP positions are marked below the haplotype plot.

balancing selection or population subdivision may cause an excess of intermediate frequency variants and positive TD and FL-F values, whereas positive selection or population growth may lead to an excess of low frequency variants and negative TD and FL-F values.^{45,46,68} For the Han Chinese population, although the obtained negative value of TD failed to reach statistical significance ($TD = -1.44935$, $P > 0.10$, Figure 3a), the bias in the spectrum of allele frequencies measured by FL-F test is significant ($FL-F = -2.43436$, $P < 0.05$, Figure 3b), which indicates that the *CYP3A4* locus is subject to positive selection in Han Chinese.

To obtain more evidence of positive selection, we used FW-H test⁴⁷ on the basis of the excess of high-frequency-derived alleles, immediately after a selective sweep. FW-H test statistic is also expected to be negative, whereas the target of selection is typically within the deepest valley of the statistic.⁴⁷ Indeed, the FW-H test statistic is negative across the *CYP3A4* locus ($FW-H = -2.65457$, normalized $FW-H = -2.19284$), further confirming a recent episode of positive selection on *CYP3A4*. Moreover, we observed two prominent valleys in the FW-H statistic, with the deeper one in the locus of *CYP3A4*1B* (Figure 3c), implying the derived allele of *CYP3A4*1B* that promotes interaction with transcription repressors and leads to a decrease in *CYP3A4* gene expression that has been favored by natural selection in the Han Chinese population. Our results are consistent with previous findings of detecting a selection against the *CYP3A4*1B* allele.¹³

Notably, the minor valley of the FW-H statistic occurs on the SNP rs2242480 (Figure 3c). To assess the evolutionary characteristics of this polymorphism, we conducted integrated haplotype score (iHS) statistic⁴⁸ by Phase II HapMap data,⁶⁹ a measure designed to detect extended haplotypes due to selective sweep. In principle, an extreme

Figure 5 Linkage disequilibrium (LD) in GOLD heatmap color scheme³⁵ by Haploview4.1. LD block was defined by the solid spine of LD method ($D' > 0.9$).

iHS value ($|iHS| > 2$) of an SNP allele can be taken as evidence for a powerful selection signal.⁴⁸ However, the iHS value is 0.443 in Africans (YRI), -0.756 in East Asians (ASN) and -1.591 in Europeans (CEU), respectively. The decay of extended haplotype homozygosity in YRI, ASN and CEU is shown in Figure 4. None of them display deviation

from neutral expectation. Hence, the SNP rs2242480 is possibly neutral or may be influenced by genetic hitchhiking effect.

In addition, according to the 'solid spine of LD' method,³⁴ the 11 SNPs formed a single LD block ($D' > 0.9$), spanning 27 kb of the CYP3A4 locus (Figure 5). As the presence of intense LD is a strong signature of recent selective sweep,⁷⁰ it also supports a role for positive selection on CYP3A4 in the Han Chinese population.

Taken together, our results confirm that the unusual low polymorphism levels and the skew towards rare variants of CYP3A4 gene in Han Chinese could be caused by the action of positive selection.

ACKNOWLEDGEMENTS

We thank Dr Jun Yu (Beijing Institute of Genomics, Chinese Academy of Sciences) for his valuable contribution. This work is supported by National Natural Science Foundation of China (No. 81001472) awarded to Qing Zhou.

- Nicolas, J. M., Espie, P. & Molimard, M. Gender and interindividual variability in pharmacokinetics. *Drug. Metab. Rev.* **41**, 408–421 (2009).
- Wilkinson, G. R. Genetic variability in cytochrome P450 3A5 and *in vivo* cytochrome P450 3A activity: some answers but still questions. *Clin. Pharmacol. Ther.* **76**, 99–103 (2004).
- Ozdemir, V., Kalow, W., Tang, B. K., Paterson, A. D., Walker, S. E., Endrenyi, L. et al. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. *Pharmacogenetics* **10**, 373–388 (2000).
- Perera, M. A. The missing linkage: what pharmacogenetic associations are left to find in CYP3A? *Expert. Opin. Drug. Metab. Toxicol.* **6**, 17–28 (2010).
- Kedmi, M., Maayan, S., Cohen, S. B., Hauzi, M. & Rund, D. MDR1 and CYP3A4 polymorphisms are associated with HIV seropositivity in Israeli patients but do not influence the course of HIV disease. *AIDS. Patient. Care. STDS.* **21**, 653–658 (2007).
- Rebbeck, T. R., Troxel, A. B., Norman, S., Bunin, G., DeMichele, A., Schinnar, R. et al. Pharmacogenetic modulation of combined hormone replacement therapy by progesterone-metabolism genotypes in postmenopausal breast cancer risk. *Am. J. Epidemiol.* **166**, 1392–1399 (2007).
- Rebbeck, T. R., Jaffe, J. M., Walker, A. H., Wein, A. J. & Malkowicz, S. B. Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. *J. Natl. Cancer. Inst.* **90**, 1225–1229 (1998).
- Tayeb, M. T., Clark, C., Sharp, L., Haites, N. E., Rooney, P. H., Murray, G. I. et al. CYP3A4 promoter variant is associated with prostate cancer risk in men with benign prostate hyperplasia. *Oncol. Rep.* **9**, 653–655 (2002).
- Yamada, Y., Matsuo, H., Watanabe, S., Kato, K., Yajima, K., Hibino, T. et al. Association of a polymorphism of CYP3A4 with type 2 diabetes mellitus. *Int. J. Mol. Med.* **20**, 703–707 (2007).
- Kang, Y. S., Park, S. Y., Yim, C. H., Kwak, H. S., Gajendrarao, P., Krishnamoorthy, N. et al. The CYP3A4*18 genotype in the cytochrome P450 3A4 gene, a rapid metabolizer of sex steroids, is associated with low bone mineral density. *Clin. Pharmacol. Ther.* **85**, 312–318 (2009).
- Liu, C. H., Peck, K., Huang, J. D., Lin, M. S., Wang, C. H., Hsu, W. P. et al. Screening CYP3A single nucleotide polymorphisms in a Han Chinese population with a genotyping chip. *Pharmacogenomics* **6**, 731–747 (2005).
- Garsa, A. A., McLeod, H. L. & Marsh, S. CYP3A4 and CYP3A5 genotyping by Pyrosequencing. *BMC. Med. Genet.* **6**, 19 (2005).
- Schirmer, M., Toliat, M. R., Haberl, M., Suk, A., Kamdem, L. K., Klein, K. et al. Genetic signature consistent with selection against the CYP3A4*1B allele in non-African populations. *Pharmacogenet. Genomics* **16**, 59–71 (2006).
- Amirimani, B., Ning, B., Deitz, A. C., Weber, B. L., Kadlubar, F. F. & Rebbeck, T. R. Increased transcriptional activity of the CYP3A4*1B promoter variant. *Environ. Mol. Mutagen.* **42**, 299–305 (2003).
- Hu, Y. F., He, J., Chen, G. L., Wang, D., Liu, Z. Q., Zhang, C. et al. CYP3A5*3 and CYP3A4*18 single nucleotide polymorphisms in a Chinese population. *Clin. Chim. Acta.* **353**, 187–192 (2005).
- Lee, S. J., Lee, S. S., Jeong, H. E., Shon, J. H., Ryu, J. Y., Sunwoo, Y. E. et al. The CYP3A4*18 allele, the most frequent coding variant in Asian populations, does not significantly affect the midazolam disposition in heterozygous individuals. *Drug. Metab. Dispos.* **35**, 2095–2101 (2007).
- Yamamoto, T., Nagafuchi, N., Ozeki, T., Kubota, T., Ishikawa, H., Ogawa, S. et al. CYP3A4*18: it is not rare allele in Japanese population. *Drug. Metab. Pharmacokinet.* **18**, 267–268 (2003).
- Ruzilawati, A. B., Suhaimi, A. W. & Gan, S. H. Genetic polymorphisms of CYP3A4: CYP3A4*18 allele is found in five healthy Malaysian subjects. *Clin. Chim. Acta.* **383**, 158–162 (2007).
- Westlind-Johnsson, A., Hermann, R., Huenemeyer, A., Hauns, B., Lahu, G., Nassr, N. et al. Identification and characterization of CYP3A4*20, a novel rare CYP3A4 allele without functional activity. *Clin. Pharmacol. Ther.* **79**, 339–349 (2006).
- Zhao, T. M. & Lee, T. D. Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Hum. Genet.* **83**, 101–110 (1989).
- Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T. et al. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* **95**, 11763–11768 (1998).
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y. et al. Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
- Gan, R. J., Pan, S. L., Mustavich, L. F., Qin, Z. D., Cai, X. Y., Qian, J. et al. Pinghua population as an exception of Han Chinese's coherent genetic structure. *J. Hum. Genet.* **53**, 303–313 (2008).
- Hsieh, K. P., Lin, Y. Y., Cheng, C. L., Lai, M. L., Lin, M. S., Siest, J. P. et al. Novel mutations of CYP3A4 in Chinese. *Drug. Metab. Dispos.* **29**, 268–273 (2001).
- Wen, S., Wang, H., Ding, Y., Liang, H. & Wang, S. Screening of 12 SNPs of CYP3A4 in a Chinese population using oligonucleotide microarray. *Genet. Test.* **8**, 411–416 (2004).
- Du, J., Xing, Q., Xu, L., Xu, M., Shu, A., Shi, Y. et al. Systematic screening for polymorphisms in the CYP3A4 gene in the Chinese population. *Pharmacogenomics* **7**, 831–841 (2006).
- International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
- Zhou, Q., Yu, X. M., Lin, H. B., Wang, L., Yun, Q. Z., Hu, S. N. et al. Genetic polymorphism, linkage disequilibrium, haplotype structure and novel allele analysis of CYP2C19 and CYP2D6 in Han Chinese. *Pharmacogenomics* **9**, 380–394 (2009).
- Lahiri, D. K. & Nurnberger, J. I. Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic. Acids. Res.* **19**, 5444 (1991).
- Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome. Res.* **8**, 175–185 (1998).
- Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome. Res.* **8**, 186–194 (1998).
- Nickerson, D. A., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic. Acids. Res.* **25**, 2745–2751 (1997).
- Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome. Res.* **8**, 195–202 (1998).
- Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- Abecasis, G. R. & Cookson, W. O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
- Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
- Stephens, M. & Scheet, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* **76**, 449–462 (2005).
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Davenport, R. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome. Res.* **13**, 2129–2141 (2003).
- Thomas, P. D. & Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc. Natl. Acad. Sci. USA* **101**, 15398–15403 (2004).
- Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic. Acids. Res.* **30**, 3894–3900 (2002).
- E Kroos, M. & Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. USA* **103**, 13682–13687 (2006).
- Binkowski, T. A., Naghibzadeh, S. & Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic. Acids. Res.* **31**, 3352–3355 (2003).
- Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
- Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
- Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
- Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
- Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS. Biol.* **4**, e72 (2006).
- Yu, Z., Li, Z., Jolicoeur, N., Zhang, L., Fortin, Y., Wang, E. et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic. Acids. Res.* **35**, 4535–4541 (2007).
- Chen, K. & Rajewsky, N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.* **38**, 1452–1456 (2006).
- Gao, Y., Zhang, L. R. & Fu, Q. CYP3A4*1G polymorphism is associated with lipid-lowering efficacy of atorvastatin but not of simvastatin. *Eur. J. Clin. Pharmacol.* **64**, 877–882 (2008).
- Cavaco, I., Gil, J. P., Gil-Berglund, E. & Ribeiro, V. CYP3A4 and MDR1 alleles in a Portuguese population. *Clin. Chem. Lab. Med.* **41**, 1345–1350 (2003).
- Chowbay, B., Zhou, S. & Lee, E. J. An interethnic comparison of polymorphisms of the genes encoding drug-metabolizing enzymes and drug transporters: experience in Singapore. *Drug. Metab. Rev.* **37**, 327–378 (2005).
- Fukushima-Uesaka, H., Saito, Y., Watanabe, H., Shiseki, K., Saeki, M., Nakamura, T. et al. Haplotypes of CYP3A4 and their close linkage with CYP3A5 haplotypes in a Japanese population. *Hum. Mutat.* **23**, 100 (2004).
- Mizutani, T. PM frequencies of major CYPs in Asians and Caucasians. *Drug. Metab. Rev.* **35**, 99–106 (2003).

- 56 Thompson, E. E., Kuttub-Boulos, H., Witonsky, D., Yang, L., Roe, B. A. & Di Rienzo, A. CYP3A variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.* **75**, 1059–1069 (2004).
- 57 Du, J., Yu, L., Wang, L., Zhang, A., Shu, A., Xu, L. *et al.* Differences in CYP3A41G genotype distribution and haplotypes of CYP3A4, CYP3A5 and CYP3A7 in 3 Chinese populations. *Clin. Chim. Acta.* **383**, 172–174 (2007).
- 58 Yamaguchi, Y., Khan, K. K., He, Y. A., He, Y. Q. & Halpert, J. R. Topological changes in the CYP3A4 active site probed with phenyldiazene: effect of interaction with NADPH-cytochrome P450 reductase and cytochrome b5 and of site-directed mutagenesis. *Drug. Metab. Dispos.* **32**, 155–161 (2004).
- 59 Ohkura, K., Kawaguchi, Y., Watanabe, Y., Masubuchi, Y., Shinohara, Y. & Hori, H. Flexible structure of cytochrome P450: promiscuity of ligand binding in the CYP3A4 heme pocket. *Anticancer. Res.* **29**, 935–942 (2009).
- 60 Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS. Biol.* **3**, e85 (2005).
- 61 Budhu, A., Jia, H. L., Forgues, M., Liu, C. G., Goldstein, D., Lam, A. *et al.* Identification of metastasis-related microRNAs in hepatocellular carcinoma. *Hepatology* **47**, 897–907 (2008).
- 62 Navon, R., Wang, H., Steinfeld, I., Tsalenko, A., Ben-Dor, A. & Yakhini, Z. Novel rank-based statistical methods reveal microRNAs with differential expression in multiple cancer types. *PLoS. One* **4**, e8003 (2009).
- 63 Simion, A., Laudadio, I., Prévot, P. P., Raynaud, P., Lemaigre, F. P., Jacquemin, P. *et al.* MiR-495 and miR-218 regulate the expression of the Onecut transcription factors HNF-6 and OC-2. *Biochem. Biophys. Res. Commun.* **391**, 293–298.
- 64 Pogribny, I. P., Tryndyak, V. P., Ross, S. A. & Beland, F. A. Differential expression of microRNAs during hepatocarcinogenesis induced by methyl deficiency in rats. *Nutr. Rev.* **66**, S33–S35 (2008).
- 65 Wong, Q. W., Lung, R. W., Law, P. T., Lai, P. B., Chan, K. Y., To, K. F. *et al.* MicroRNA-223 is commonly repressed in hepatocellular carcinoma and potentiates expression of Stathmin1. *Gastroenterology* **135**, 257–269 (2008).
- 66 Mishra, P. J., Humeniuk, R., Longo-Sorbello, G. S., Banerjee, D. & Bertino, J. R. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc. Natl. Acad. Sci. USA* **104**, 13513–13518 (2007).
- 67 Chen, K., Song, F., Calin, G. A., Wei, Q., Hao, X. & Zhang, W. Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. *Carcinogenesis* **29**, 1306–1311 (2008).
- 68 Tajima, F. DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* **123**, 229–240 (1989).
- 69 Consortium, T. I. H. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- 70 Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)