buffer, separated by electrophoresis in 8% SDS-polyacrylamide gels, and transferred onto nitrocellulose membranes (Schleicher & Schuell Inc., Dassel, Germany). For immunostaining of CYP2C9, goat anti-CYP2C6 antiserum (diluted 1:1000, Daiichi Pure Chemical Co., Tokyo, Japan), which can cross-react with human CYP2C9, and horseradish peroxidase-conjugated rabbit anti-goat IgG (diluted 1:20,000, Jackson ImmunoResearch Laboratories, West Grove, Pennsylvania, USA) were used as the first and second antibodies, respectively. Calnexin was detected with rabbit anti-calnexin antiserum (diluted 1:2000, StressGen Biotechnologies Corp., Victoria, British Columbia, Canada) and donkey anti-rabbit IgG (diluted 1:4000, GE Healthcare, Piscataway, New Jersey, USA). Immunoreactive proteins were visualized with Western Lightning Chemiluminescence Reagent Plus (PerkinElmer Life Sciences, Boston, Massachusetts, USA), and the band densities were quantified with Diana III and Zero Dscan software (Raytest, Straubenhardt, Germany). Commercially available microsomes of human CYP2C9-expressing lymphoblast cells (BD Gentest, Woburn, Massachusetts, USA) were used as a calibration standard. Quantification was performed within a linear range of its fluorescent intensity. CYP2C9 levels from three independent transfection experiments were determined based on mg of microsomal protein, which was normalized to calnexin content.

Assay for CYP2C9-mediated enzymatic activity

CYP2C9 activities for the wild-type and seven variants were assessed by diclofenac 4′-hydroxylation according to the method of Leemann *et al.* [22] with some modifications. Diclofenac and 4′-hydroxydiclofenac were purchased from Sigma (St Louis, Missouri, USA) and BD Gentest, respectively. The incubation mixture contained diclofenac (2.5–100 μM), the microsomal fraction (400 μg protein) from COS-1 cells, human NADPH-P450 Reductase microsomes (150 μg protein, BD Gentest), a NADPH regenerating system (1.3 mM $NADP^+$, 3.3 mM glucose 6-phosphate, 3.3 mM $MgCl_2 \cdot 6H_2O$, and 0.4 U/ml glucose-6-phosphate dehydrogenase), and 100 mM Tris-HCl buffer (pH 7.5) in a final volume of 500 μl. Diclofenac was dissolved in methanol [final concentration of methanol in reaction medium was 1% (v/v)]. After preincubation at 37°C for 5 min, the reaction was started by the addition of microsomes and allowed to proceed at 37°C for 50 min. The reaction was stopped on ice by the addition of 100 μl of 96% acetonitrile acidified with 4% glacial acetic acid. The samples were then subjected to centrifugation at 6000*g* for 20 min at 4°C, and the resulting supernatants were filtered through 0.2 mM pore size polytetra-fluoroethylene membrane filters (Millipore, Bedford, Massachusetts, USA). The formation of the metabolite 4′-hydroxydiclofenac was linear under our conditions up to 50 min. Preliminary experiments also showed that the amount of exogenous NADPH-P450 reductase microsome (150 μg protein), which was equivalent to 0.345 U, was sufficient to support the optimum activity of CYP2C9 expressed in COS-1 cells.

The production of 4′-hydroxydiclofenac was determined by reverse-phase high-performance liquid chromatography (HPLC). HPLC analysis was performed using a JASCO GULLIVER system consisting of PU-980 pumps, UV-970 detector, AS-2057Plus autosampler, and CO-2065Plus column oven, which were all controlled by JASCO-BORWIN/HSS-2000 programs (Nihon Bunko, Tokyo, Japan). Aliquots of the supernatants (100 μl) were injected into an Inertsil ODS-3 column (5 μm, 250 × 4.6 mm i.d., GL Sciences, Tokyo, Japan) connected to a pre-column (GL Sciences). The mobile phase consisted of 30% acetonitrile containing 1 mM perchloric acid (A) and methanol (B), forming a 20-min linear gradient from 30% to 100% of B at a flow rate of 1 ml/min. The ultraviolet wavelength was fixed at 280 nm, and the column was maintained at 50°C. Under these conditions, the retention times of 4′-hydroxydiclofenac and diclofenac were 11 and 15 min, respectively. The standard curve was obtained within the range from 1.25 μM to 10.0 μM for 4′-hydroxydiclofenac. The detection limit of the metabolite was 10 pmol, which corresponded to the enzyme activity of 3.0 pmol/min/mg protein. The interday variation coefficients of diclofenac 4′-hydroxylation assay did not exceed 6% ($n = 4$). Kinetic parameters such as K_m and V_{max} were estimated using the computer program designed for non-linear regression analysis of a hyperbolic Michaelis–Menten equation (Prism v.3.0a, GraphPad Software, San Diego, California, USA). Data are presented as the mean ± SD for three microsomal preparations derived from separate transfections for each variant and analysed by one-way analysis of variance, and multiple comparisons were made with the Scheffe test.

Linkage disequilibrium (LD) and haplotype analysis

Hardy–Weinberg equilibrium and LD analysis was performed by SNPAlyse software (Dynacom Co., Yokohama, Japan), and a pairwise LD between variations was obtained for $|D'|$ and rho square (r^2) values. Some of the haplotypes were unambiguous from subjects with homozygous variations at all sites or a heterozygous variation at only one site. The diplotype configurations were inferred by LDSUPPORT software, which determines the posterior probability distribution of the diplotype for each subject based on the estimated haplotype frequencies [23]. We designated the haplotypes as follows: the group of haplotypes without amino acid changes was defined as **1*, and the groups with amino acid changes (Ile359Leu and Leu90Pro) were numbered (**3* and **13*, respectively) according to the designations by the Human Cytochrome P450 (CYP) Allele Nomenclature Committee. Six newly identified haplotypes (alleles) with novel non-synonymous variations (*CYP2C9*25* to **30*) have been registered to the Human Cytochrome P450 (CYP) Allele

Nomenclature Committee. Subtypes within each haplotype group were consecutively named with small alphabetical letters depending on their frequencies, except for **1A*, **1B* and **3B*, which were already assigned by the Committee. We could not detect the **1C* and **1D* haplotypes in our study. To avoid confusion with the nomenclature by the CYP Allele Nomenclature Committee, other **1* haplotypes were tentatively named (**1e* to **1al*). The haplotypes inferred in only one chromosome are shown as the haplotype name plus a question mark, since the predictability for these very rare haplotypes is known to be low in some cases.

Network analysis of unambiguous haplotypes was performed with Network 4.1.0.9 (www.fluxus-engineering.com) based on the algorithms of the reduced median network. The long-range haplotypes spanning *CYP2C19* and *CYP2C9* were inferred by Hapblock software (www.cmb.usc.edu/msms/HapBlock/).

Results

CYP2C9 variations found in a Japanese population

We found 62 variations including 32 novel ones by analysing the promoter regions, all nine exons and their flanking regions of *CYP2C9* from 263 Japanese subjects (Table 2). The distribution of the variations included 18 in the 5′ flanking region, 14 (five synonymous and nine non-synonymous) in the coding exons, 27 in the introns, two in the 3′-untranslated region (3′-UTR), and one in the 3′-flanking region. All of the detected variations were in Hardy–Weinberg equilibrium for two separate groups ($P \geq 0.37$ in diabetic patients and $P \geq 0.20$ in healthy volunteers) and for all subjects ($P \geq 0.44$). Since we did not find any significant differences in the frequencies between the healthy volunteers and diabetic patients ($P > 0.05$ by χ^2 test or Fisher's exact test), the data for all subjects were analysed as one group.

Ten novel variations were identified in the coding region, of which three, 336T > C (Ile112Ile), 798C > T (Cys266Cys) and 1137T > C (Tyr379Tyr), were synonymous, and seven, 49C > A (Leu17Ile), 353_362del AGAAATGGAA (Lys118ArgfsX9), 389C > G (Thr130Arg), 449G > T (Arg150Leu), 641A > T (Gln214Leu), 835C > A (Pro279Thr) and 1429G > A (Ala477Thr), were non-synonymous. All the non-synonymous variations were found as single heterozygotes (allele frequencies = 0.002), except for 449G > T (Arg150Leu), which was present in two subjects (frequency = 0.004). The six novel alleles were designated as **25* (Lys118ArgfsX9), **26* (Thr130Arg), **27* (Arg150Leu), **28* (Gln214Leu), **29* (Pro279Thr) and **30* (Ala477Thr) by the Human Cytochrome P450 Allele Nomenclature Committee. Another novel allele, 49C > A (Leu17Ile), was tentatively named **17I* because the subject carrying this allele was heterozygous also for **3* (1075A > C, Ile359Leu).

In the 5′ flanking region, nine novel SNPs (−3279T > C, −3070G > A, −2980C > T, −2787G > T, −2718G > A, −2711C > T, −1924A > G, −885A > G and −732T > G) were detected at an allele frequency of 0.002. Recent studies on functional nuclear receptor binding sites in the promoter region in *CYP2C9* revealed two constitutive androstane receptor-responsive elements (CAR-REs) between positions −2900 and −2841 [24] and between −1839 and −1824 [25] from the translation initiation site, one glucocorticoid receptor-responsive element (GRE) between −1697 and −1683 [25] and two proximal HNF4α binding sites between −152 and −129 and between −185 and −163 [14,26]. However, novel SNPs were not found in these established elements.

The remaining 13 novel variations were found either in the intronic region or 3′-UTR with allele frequencies less than 0.01. These variations seem unlikely to affect splicing because they were not located in the exon-intron splicing junctions or branch sites.

Thirty variations were already reported or publicized in the dbSNP database and/or the Human Cytochrome P450 Allele Nomenclature Committee homepage. Their ID numbers in the databases and references are provided in Table 2. We found two reported non-synonymous SNPs, 269T > C (**13*, Leu90Pro) and 1075A > C (**3*, Ile359Leu), at allele frequencies of 0.002 and 0.03, respectively. However, 430C > T (**2*, Arg144Cys) was not found in our Japanese population. As for the promoter SNPs, the substitution −2875C > G, located within distal CAR-RE (−2900 to −2841), was found [24]. The relatively frequent SNPs −3089G > A, −2665_−2664delTG, −1565C > T and −1188T > C were also detected.

Expression levels of the seven CYP2C9 variants

To functionally characterize the non-synonymous variations, the wild-type and each of the seven variant proteins (Leu17Ile, Lys118ArgfsX9, Thr130Arg, Arg150Leu, Gln214Leu, Pro279Thr and Ala477Thr) were transiently expressed in transfected COS-1 cells. First, expression levels of CYP2C9 proteins were assessed by Western blotting. Representative data obtained from three independent transfections are shown in Fig. 1. Expression levels of the endoplasmic reticulum-resident protein, calnexin, were almost constant in microsomes from the transfected cells. No CYP2C9 protein expression was detected, even as proteolytic fragments, for the Lys118ArgfsX9 variant (The residues 118–126, Lys-Lys-Trp-Lys-Glu-Ile-Arg-Arg-Phe, were changed to Arg-Arg-Ser-Gly-Val-Ser-Pro-Ser-Stop.) and the empty vector (Fig. 1a, b). The expression level of wild-type CYP2C9 was calculated as 12.0 ± 2.5 pmol/mg microsomal protein. No significant differences were shown in the expression levels of CYP2C9 protein among the wild-type and the other six variants ($P = 0.84$, Fig. 1b).

Table 2 Summary of *CYP2C9* single nucleotide polymorphisms detected in a Japanese population

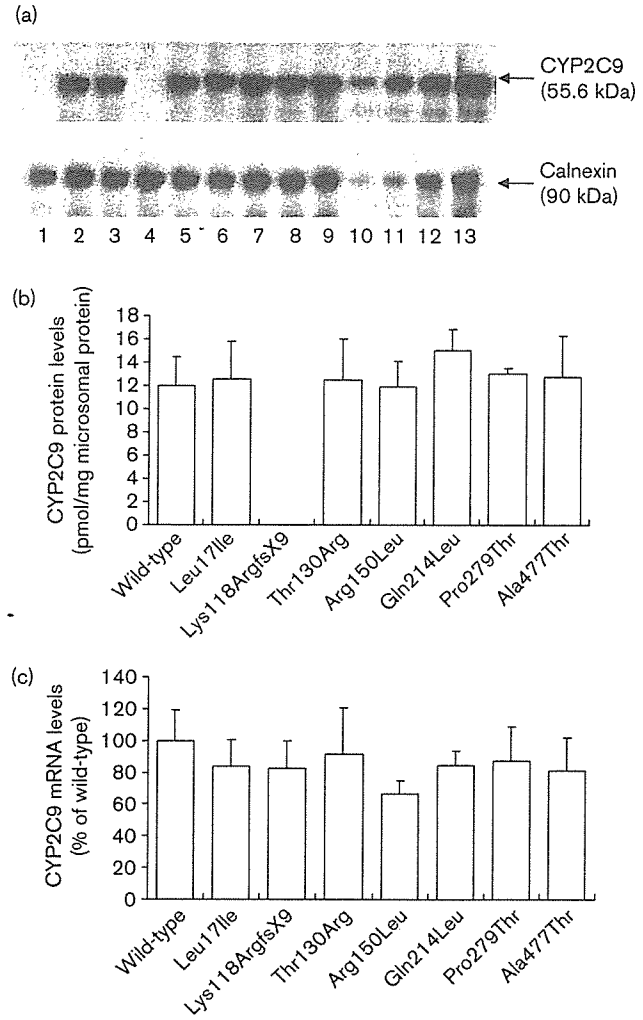
Location	Position		Nucleotide change and flanking sequences (5′–3′)	Amino acid change	Number of subjects			Allele Frequency	dbSNP SNP ID (NCBI)	Reference[c]
	NT_030059.12	From the translational initiation site or from the end of nearest exon			Wild-type	Heterozygote	Homozygote			
5′-Flanking	15443687	−3279[a]	CATCA*T*/CTGTCA		262	1	0	0.002		
5′-Flanking	15443877	−3089	CAACC*G*/*A*TATTA		134	106	23	0.289	rs12782374	
5′-Flanking	15443896	−3070[a]	AAAAG*G*/*A*CAAAA		262	1	0	0.002		
5′-Flanking	15443986	−2980[a]	TAGGG*C*/*T*AGTAA		262	1	0	0.002		
5′-Flanking	15444091	−2875	ATGAG*C*/*G*TTTGG		262	1	0	0.002		[10]
5′-Flanking	15444179	−2787[a]	GTGCA*G*/*T*GGAAA		262	1	0	0.002		
5′-Flanking	15444248	−2718[a]	CAAAA*G*/*A*CCTAC		262	1	0	0.002		
5′-Flanking	15444255	−2711[a]	CTACT*C*/*T*TAATC		262	1	0	0.002		
5′-Flanking	15444301_15444302	−2665_−2664	GTGAC*TG*/−TGGAG		134	106	23	0.289		[13,16]
5′-Flanking	15445042	−1924[a]	GAGTC*A*/*G*GGGAC		262	1	0	0.002		
5′-Flanking	15445055	−1911	AGTTA*T*/CTGCTT		247	16	0	0.030	rs9332092	
5′-Flanking	15445081	−1885	AAAGG*C*/*G*TTCTC		247	16	0	0.030	rs9332093	
5′-Flanking	15445401	−1565	CATTC*C*/*T*GGAAA		202	56	5	0.125	rs9332096	
5′-Flanking	15445429	−1537	AAGCA*G*/*A*AGGTA		247	16	0	0.030		[10,12,13,15–18]
5′-Flanking	15445778	−1188	ATCTT*T*/CTATTG		78	136	49	0.445	rs4918758	
5′-Flanking	15445985	−981	ATGGA*G*/*A*AAGGG		247	16	0	0.030	rs9332098	
5′-Flanking	15446081	−885[a]	TTATG*A*/*G*ACAGA		262	1	0	0.002		
5′-Flanking	15446234	−732[a]	AGCTA*T*/GGAGCT		262	1	0	0.002		
Exon 1	15447014	49[a]	TTCTCC/ATTTCA	Leu17Ile	262	1	0	0.002		
Intron 1	15447216	IVS1+83	GTACA*T*/CGTTAC		257	6	0	0.011	rs9332104	
Intron 1	15447258	IVS1+125[a]	AGGCT*T*/CTTGTT		261	2	0	0.004		
Intron 1	15449977	IVS1−164[a]	AGTGG*C*/*A*AAGTA		262	1	0	0.002		
Intron 1	15450031	IVS1−110[a]	GCCTG*T*/*G*GTGG		260	3	0	0.006		
Intron 1	15450119	IVS1−22[a]	TGCTG*T*/CTATCT		260	3	0	0.006		
Exon 2	15450200	228	GTGGT*G*/*A*CTGCA	Val76Val	260	3	0	0.006		[10]
Exon 2	15450241	269	TGATC*T*/CTGGAG	Leu90Pro	262	1	0	0.002		[11]
Intron 2	15450376	IVS2+73	AGAGC*T*/CCCTCG		257	6	0	0.011	rs9332120	
Exon 3	15450479	336[a]	GGAAT*T*/CGTTTT	Ile112Ile	262	1	0	0.002		
Exon 3	15450496_15450505	353_362[a]	TGGAA*AGAAATGGAA*/−−−−GGAGA	Lys118Arg fsX9	262	1	0	0.002		
Exon 3	15450532	389[a]	CATGA*C*/*G*GCTGC	Thr130Arg	262	1	0	0.002		
Exon 3	15450592	449[a]	AGCCC*G*/*T*CTGCC	Arg150Leu	261	2	0	0.004		
Intron 3	15450821	IVS3+197	TGTGC*G*/*A*TACAG		188	67	8	0.158	rs2860905	
Intron 3	15450863	IVS3+239	TTCTC*C*/*T*TGAAC		133	109	21	0.287		[13]
Intron 3	15450889	IVS3+265	TCAAA*T*/CAAGAA		247	16	0	0.030		[13]
Intron 3	15455728	IVS3−334	CCTTG*C*/*T*TGTCT		188	68	7	0.156	rs4086116	
Intron 3	15455951	IVS3−111[a]	TTGCT*G*/*A*TTAAG		262	1	0	0.002		

Table 2 (Continued)

Location	Position		Nucleotide change and flanking sequences (5′–3′)	Amino acid change	Number of subjects			Allele Frequency	dbSNP SNP ID (NCBI)	Reference[c]
	NT_030059.12	From the translational initiation site or from the end of nearest exon			Wild-type	Heterozygote	Homozygote			
Intron 3	15455997	IVS3 – 65	TATCT*G*/*C*TTAAC		202	56	5	0.125	rs9332127	
Exon 4	15456221	641[a]	GATCC*A*/*T*GGTAA	Gln214Leu	262	1	0	0.002		
Intron 4	15456408	IVS4 + 186[a]	AAATG*C*/*T*CTCAA		262	1	0	0.002		
Intron 4	15457276	IVS4 – 115	TCAAG*A*/*G*TATAC		202	56	5	0.125	rs9332129	
Exon 5	15457546	798[a]	GATTG*C*/*T*TTCCT	Cys266Cys	262	1	0	0.002		
Intron 5	15457643	IVS5 + 76[a]	TACTA*A*/*G*GGATG		262	1	0	0.002		
Intron 5	15480314	IVS5 – 73	TGTGA*A*/*G*TAATT		189	67	7	0.154	rs9332172	
Exon 6	15480402	835[a]	ACCAAC/*A*CATCT	Pro279Thr	262	1	0	0.002		
Intron 6	15480623	IVS6 + 95	GCTTC*A*/*G*TTATT		257	6	0	0.011	rs9332174	
Intron 6	15480680	IVS6 + 152[a]	TTACA*A*/*G*TGGGA		261	2	0	0.004		
Intron 6	15489303	IVS6 – 163[a]	TTTGT*G*/*A*CATCTG		259	4	0	0.008		
Intron 6	15489329	IVS6 – 137[a]	TTAAG*T*/*A*TTGCA		258	5	0	0.010		
Exon 7	15489579	1075	ATAC*A*/CTTGAC	Ile359Leu	247	16	0	0.030	rs1057910	
Exon 7	15489641	1137[a]	AACTA*T*/CCTCAT	Tyr379Tyr	260	3	0	0.006		
Intron 7	15489691	IVS7 + 38	GTTTTC/*T*GAAGT		245	18	0	0.034		[10]
Intron 8	15494510	IVS8 + 53	TTTTG*A*/*T*TCCAT		247	16	0	0.030	rs9332230	
Intron 8	15494604	IVS8 + 147	CCCTG*C*/*T*TCATG		133	109	21	0.287	rs2298037	
Intron 8	15497017	IVS8 – 113[a]	TTCTAC/*T*GATAC		262	1	0	0.002		
Intron 8	15497018	IVS8 – 112	TCTACG/*A*ATACA		247	16	0	0.030	rs9332238	
Intron 8	15497021	IVS8 – 109	ACGAT*A*/*T*CACTG		75	136	52	0.456	rs1934969	
Intron 8	15497032	IVS8 – 98[a]	AACAG*T*/*A*TATTG		262	1	0	0.002		
Exon 9	15497263	1425	AATGG*A*/*T*TTTGC	Gly475Gly	247	16	0	0.030	rs1057911	
Exon 9	15497267	1429[a]	GATTT*G*/*A*CCTCT	Ala477Thr	262	1	0	0.002		
3′-UTR	15497318	*7[a,b]	GAAGA*G*/CCAGAT		262	1	0	0.002		
3′-UTR	15497579_15497580	*268_*269[b]	TTAAT*AT*/–GTTAT		261	2	0	0.004		[12]
3′-flanking	15497707	*396[b]	ATGCA*T*/*A*ATGT		247	16	0	0.030	rs9332245	

[a]Novel variations detected in this study.
[b]The nucleotide following the translation termination codon TGA is numbered +1.
[c]Refer to the reference section.

Fig. 1



Expression of wild-type and seven variant CYP2C9 in COS-1 cells. (a) A representative Western blot showing the immunoreactive proteins for CYP2C9 (upper panel) and for calnexin (lower panel). Lane 1, empty vector; lane 2, wild-type; lane 3, Leu17Ile; lane 4, Lys118ArgfsX9; lane 5, Thr130Arg; lane 6, Arg150Leu; lane 7, Gln214Leu; lane 8, Pro279Thr; lane 9, Ala477Thr; lanes 10–13, commercially available CYP2C9-expressed microsomes used for calibration standards, 0.1 pmol P450 (lane 10), 0.2 pmol P450 (lane 11), 0.5 pmol P450 (lane 12) and 0.75 pmol P450 (lane 13). (b) Quantitative results of immunoreactive CYP2C9 protein. Each bar represents the mean ± SD of three separate experiments. (c) Quantitative results of CYP2C9 mRNA expression assessed by TaqMan real-time reverse transcriptase-polymerase chain reaction. Each sample was normalized to β-actin mRNA content and expressed as a percentage of wild-type. Each bar represents the mean ± SD of four to six separate experiments.

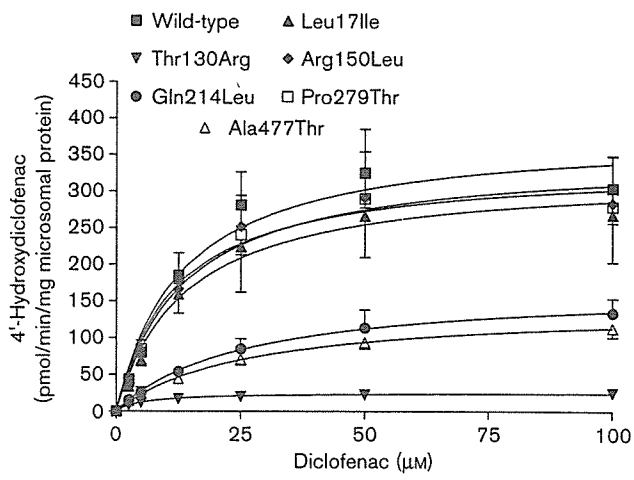
CYP2C9 mRNA expression levels were determined by TaqMan real-time RT-PCR (Fig. 1c). CYP2C9 mRNA was not detected from cells transfected with the empty plasmid. CYP2C9 mRNA levels in the seven variants were not significantly different from those in the wild-type ($P = 0.22$), suggesting that the transfection/transcription efficiencies of all plasmid constructs were nearly equal. Therefore, the truncated protein derived from the Lys118ArgfsX9 variant might be unstable and/or rapidly degraded in this expression system.

Diclofenac 4′-hydroxylation activities of the wild-type and seven variant CYP2C9s

Figure 2 shows the Michaelis–Menten curves for diclofenac 4′-hydroxylation by the wild-type and six variants. As expected, no 4′-hydroxylated metabolite was detected for the microsomal preparation derived from the Lys118ArgfsX9 variant (data not shown). The kinetic parameters are summarized in Table 3. The K_m value (13.2 ± 1.0 μM) of the wild-type was comparable to those of human liver microsomes (5.6–27.6 μM) [22,27], but slightly higher than those previously reported for yeast-expressed CYP2C9 (1.8–3.9 μM) [5,28] and CYP2C9 expressed in a baculovirus/insect system (4.8 μM) [8]. The relatively rapid depletion of diclofenac at lower concentrations (2.5–10 μM) might have occurred in our experimental conditions and resulted in the elevated K_m value of the wild-type (underestimation of the enzyme activity). The V_{max} value (32.1 ± 2.4 pmol/min/pmol CYP2C9) of wild-type was also comparable to that previously reported for yeast-expressed CYP2C9 (35.6 ± 1.3 pmol/min/pmol P450) [28].

As shown in Table 3, the Thr130Arg variant was the most defective in metabolizing diclofenac among the six variants. The V_{max} value and intrinsic clearance were significantly decreased by 93% and 83%, respectively, both on the mg microsomal protein basis and pmol CYP2C9 basis ($P < 0.01$ for all comparisons to the

Fig. 2



Michaelis–Menten curves for diclofenac 4′-hydroxylation from recombinant wild-type and variant CYP2C9. The solid line indicates fitting of the data to the Michaelis–Menten equation by non-linear regression. Each point represents the mean ± SD of three independent preparations derived from different transfections.

Table 3 Kinetic parameters for hydroxylation activities of wild-type and variant CYP2C9 against diclofenac

Amino acid alteration	K_m (μM)	V_{max} (pmol/min/mg protein)	V_{max} (pmol/min/pmol P450)	Clearance (V_{max}/K_m) (μl/min/mg protein)	Clearance (V_{max}/K) (μl/min/pmol P450)
Wild-type	13.2 ± 1.0	381.7 ± 59.9	32.1 ± 2.4	29.1 ± 5.3	2.4 ± 0.09
Leu17Ile	13.7 ± 2.6	326.0 ± 85.3	27.9 ± 11.5	23.6 ± 2.9	2.0 ± 0.52
Lys118ArgfsX9 (*25)	ND	ND	ND	ND	ND
Thr130Arg (*26)	5.9 ± 2.0	25.9 ± 6.0****	2.2 ± 0.7**	4.7 ± 1.8****	0.4 ± 0.22****
Arg150Leu (*27)	13.4 ± 3.2	351.2 ± 86.8	31.0 ± 11.8	26.1 ± 0.4	2.2 ± 0.43
Gln214Leu (*28)	25.9 ± 4.5***	170.6 ± 31.3*	11.5 ± 2.4	6.7 ± 1.3****	0.5 ± 0.13****
Pro279Thr (*29)	11.9 ± 2.7	338.1 ± 34.7	26.0 ± 3.0	29.0 ± 4.4	2.2 ± 0.37
Ala477Thr (*30)	25.9 ± 0.7***	142.9 ± 14.7**	11.7 ± 2.5	5.5 ± 0.6****	0.5 ± 0.10****

*$P<0.05$,
**$P<0.01$,
***$P<0.005$
****$P<0.001$ versus wild-type. One-way analysis of variance, post-hoc test; Scheffe. Data is represented by mean ± SD. ND, Not detected.

wild-type). By contrast, its K_m value was not significantly changed. The two variants, Gln214Leu and Ala477Thr, both of which were also defective compared to the wild-type, showed similar kinetics. The K_m values of these two variant enzymes were approximately two-fold higher than that of the wild-type ($P < 0.005$). They also exhibited 55% (Gln214Leu) and 63% (Ala477Thr) decrease in V_{max} and 77% (Gln214Leu) and 81% (Ala477Thr) decrease in intrinsic clearance (V_{max}/K_m) on the basis of mg microsomal protein, and these reductions were statistically significant. After normalization to the corresponding CYP2C9 protein content determined by Western blotting, the intrinsic clearance of both Gln214Leu and Ala477Thr was still significantly lower than that of the wild-type ($P < 0.001$). The kinetic parameters of other variants (Leu17Ile, Arg150Leu and Pro279Thr) were similar to those of the wild-type. Thus, these results demonstrate that in addition to the Lys118ArgfsX9 variant, the three amino acid substitution variants, Thr130Arg, Gln214Leu and Ala477Thr, cause markedly decreased catalytic activities towards diclofenac hydroxylation *in vitro*.

LD analysis

Using 26 common SNPs (more than 0.01 in their allele frequencies), LD analysis was performed by $|D'|$ and r^2 statistics, and their pairwise values are depicted with a 10-graded blue colour in Fig. 3. As already reported in various ethnic groups [10,12,13,16], five SNPs, −1911T > C, −1885C > G, −1537G > A, −981G > A, IVS3 + 265T > C and IVS8 + 53A > T, were in perfect LD ($r^2 = 1.0$) with 1075A > C (*3, Ile359Leu). Furthermore, we found that an additional three SNPs, IVS8-112G > A, 1425A > T (Gly475Gly) and *396T > A, were also perfectly associated with *3 in Japanese ($r^2 = 1.0$). Veenstra *et al.* [13] also reported that IVS8−112G > A was linked both with *2 and *3 in European-American warfarin patients; however, *2 (430C > T, Arg144Cys) was not detected in our study.

Perfect LD was observed in four other variation pairs: −3089G > A and −2665_−2664delTG; −1565C > T, IVS3−65G > C and IVS4−115A > G; IVS1 + 83T > C, IVS2 + 73T > C and IVS6 + 95A > G; and IVS3 + 239C > T and IVS8 + 147C > T. Close associations were observed among −1565C > T, IVS3 + 197 G > A, IVS3−334C > T, IVS3−65G > C, IVS4−115A > G and IVS5−73A > G ($r^2 = 0.76$) and among −3089G > A, −2665_−2664delTG, −1188T > C, IVS3 + 239C > T, IVS8 + 147C > T and IVS8−109A > T ($r^2 > 0.48$).

As for the $|D'|$ values between 26 common SNPs, 283 pairs (87%) and 311 pairs (95%) out of 325 pairs had $|D'| > 0.99$ and $|D'| > 0.80$, respectively, indicating that one LD block covers the entire gene. Recombination might have occurred between −2665_−2664delTG and −1911T > C in four out of 263 subjects (data not shown), but this event was too minor to divide the LD block. Thus, the haplotypes of *CYP2C9* were analysed as one LD block that spans at least 54 kb.

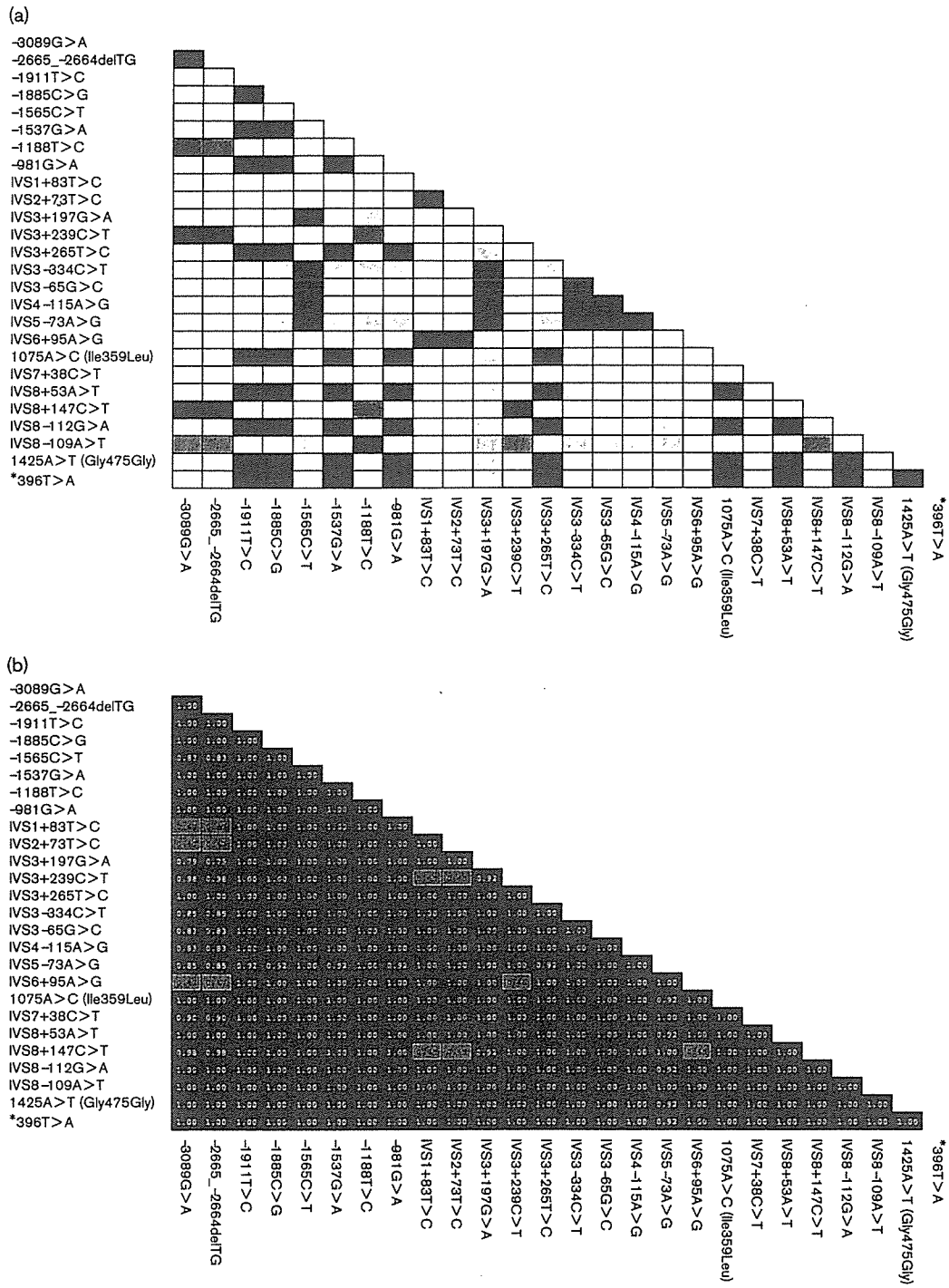
Haplotype estimation, selection of haplotype-tagging SNPs and network analysis

Using all 62 SNPs detected, 46 haplotypes were inferred. Although a large number of haplotypes were inferred, they were classified into only six major haplotype groups: *1A, *1B, *1e, *1f, *1h and *3 groups as shown in Fig. 4. The probability of diplotype configurations was over 0.99 except for 12 subjects.

The *1A group of haplotypes consisted of 15 *1 subtypes and three haplotypes with non-synonymous SNPs, *25a (Lys118ArgfsX9), *28a (Gln214Leu) and *30a (Ala477Thr). The most dominant haplotype *1A (frequency = 0.489) without nucleotide changes was considered the wild-type haplotype. The other haplotypes in this group differ from *1A in only one site and were found at frequencies of ≤ 0.01.

The *1B group was defined as the group of haplotypes that contained six linked non-coding SNPs, −3089G > A, −2665_−2664delTG, −1188T > C, IVS3 + 239C > T, IVS8 + 147C > T and IVS8−109A > T. The second

Fig. 3

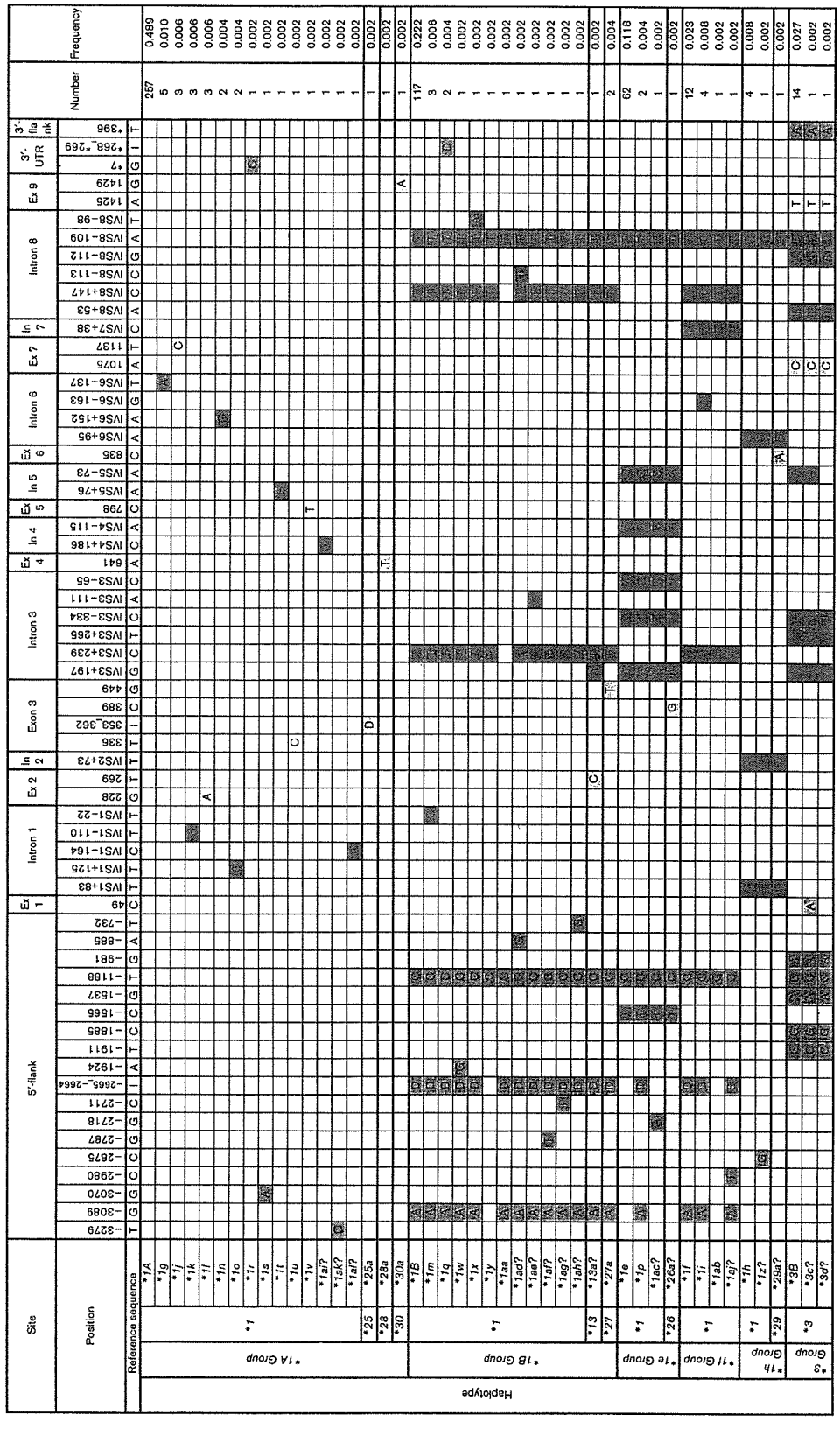


Linkage disequilibrium (LD) analysis of *CYP2C9*. Pairwise LD between 26 common single nucleotide polymorphisms (>0.01 in allele frequencies) is expressed as (a) r^2 and (b) |D'| by a 10-graded blue colour. The denser colour indicates higher linkage.

common haplotype in Japanese, **1B* (0.222 in frequency), was the most dominant haplotype in this group. This haplotype, first assigned as **1B* by King *et al.* [16], was reported by Veenstra *et al.* [13] as Haplotype 6 (**1B*). This group included the other 11 **1* subtypes, **13a* (Leu90Pro) and **27a* (Arg150Leu), with frequencies less than 0.01.

The **1e* group consisted of four haplotypes, **1e*, **1p*, **1ac* and **26a* (Thr130Arg), all of which were characterized by

Fig. 4



Haplotype distribution of *CYP2C9* in a Japanese population. The positions on the cDNA sequence (A of the translational start codon is +1, based on NT_030059.12) or positions from the nearest exon were used for the description of the single nucleotide polymorphisms. The haplotypes already assigned by Human Cytochrome P450 Allele Nomenclature Committee are shown by capital alphabetical letters. The other inferred haplotypes are shown with lower case letters. The haplotypes inferred in only one patient are indicated with a question mark. White boxes denote the reference sequence; the colored squares represent intronic (green), non-coding (blue), synonymous (yellow) and non-synonymous (pink) sites. D, Deletion; I, insertion.

the eight SNPs, −1565C > T, −1188T > C, IVS3 + 197 G > A, IVS3−334C > T, IVS3−65G > C, IVS4−115A > G, IVS5−73A > G and IVS8−109A > T. The third common haplotype in Japanese was **1e* (0.118 in frequency).

The **1f* group, which was closely related to the **1B* group, included **1f*, **1i*, **1ab* and **1aj*. The **1f* harbors IVS7 + 38C > T together with all six SNPs in **1B*. The **1f* was inferred at a frequency of 0.023, while others were less than 0.01.

Although the frequency of the **1h* group was low (< 0.01 in frequency), this group was quite different from any other group. All three haplotypes in this group, **1h*, **1x* and **29a* (Pro279Thr), contain three perfectly linked SNPs, IVS1 + 83T > C, IVS2 + 73T > C and IVS6 + 95A > G, together with the common SNP IVS8−109A > T. Except for **1h*, the other two haplotypes, **1x* and **29a*, were inferred in only one patient.

In the **3* group, three haplotypes harboring 1075A > C (Ile359Leu), **3B*, **3c* and **3d*, were inferred. King *et al.* [16] already assigned **3A* and **3B* in Caucasians; **3A* harbors the four SNPs in the promoter region, −1911T > C, −1885C > G, −1537G > A and −981G > A, together with 1075A > C (Ile359Leu), while **3B* harbors the additional SNP −1188T > C as well as the five SNPs described above. Because 1075A > C (Ile359Leu) was completely linked with −1188T > C, **3A* was not found in Japanese. The **3B* was found at a frequency of 0.027 and was the fourth frequent haplotype following **1A*, **1B* and **1e* and harbored IVS3 + 197G > A, IVS3 + 265T > C, IVS3−334C > T, IVS5−73A > G, IVS8 + 53A > T, IVS8−112G > A, IVS8−109A > T, 1425A > T (Gly475Gly) and *396T > A in addition to the five promoter SNPs and 1075A > C (Ile359Leu). The concurring SNPs in **3B* are in good agreement with the result of Veenstra *et al.* [13], who detected this haplotype in European-American patients and designated it Haplotype 3. Leu17Ile was assigned to haplotype **3c* concurring with 15 linked SNPs in **3B*. The **3c* and **3d* in this group were rare and ambiguous.

The network analysis of unambiguous haplotypes was performed to obtain a cladogram based on the sites and numbers of mutational events. In Fig. 5, the cladogram clearly discriminated six discrete haplotype groups, **1A*, **1B*, **1e*, **1f*, **1h* and **3*, which were connected to each other by several mutational sites. In the cladogram, each variation appears only once except for the two SNPs, −3089G > A and −2665_−2664delTG, which appeared twice, suggesting that recurrent recombination had hardly occurred throughout the analysed region. As a result, the haplotype structure of *CYP2C9* in Japanese is not complex, and only five common haplotypes were found in more than 10 subjects (at frequencies > 0.02), **1A*, **1B*, **1e*, **1f* and **3B*, account for 87% of all observed haplotypes. The haplotype-tagging SNPs (htSNPs) that resolved the five common haplotypes were the following four variations: −1565C > T, 1075A > C (Ile359Leu), IVS7 + 38C > T and IVS8 + 147C > T, which are indicated in red in Fig. 5.

Based on the haplotype structure in Caucasians reported by Veenstra *et al.* [13], we added **2* (430C > T, Arg144Cys) in the cladogram. The **2* was separated on the way from the long branch towards **3* and distinguished from **3* by multiple SNPs.

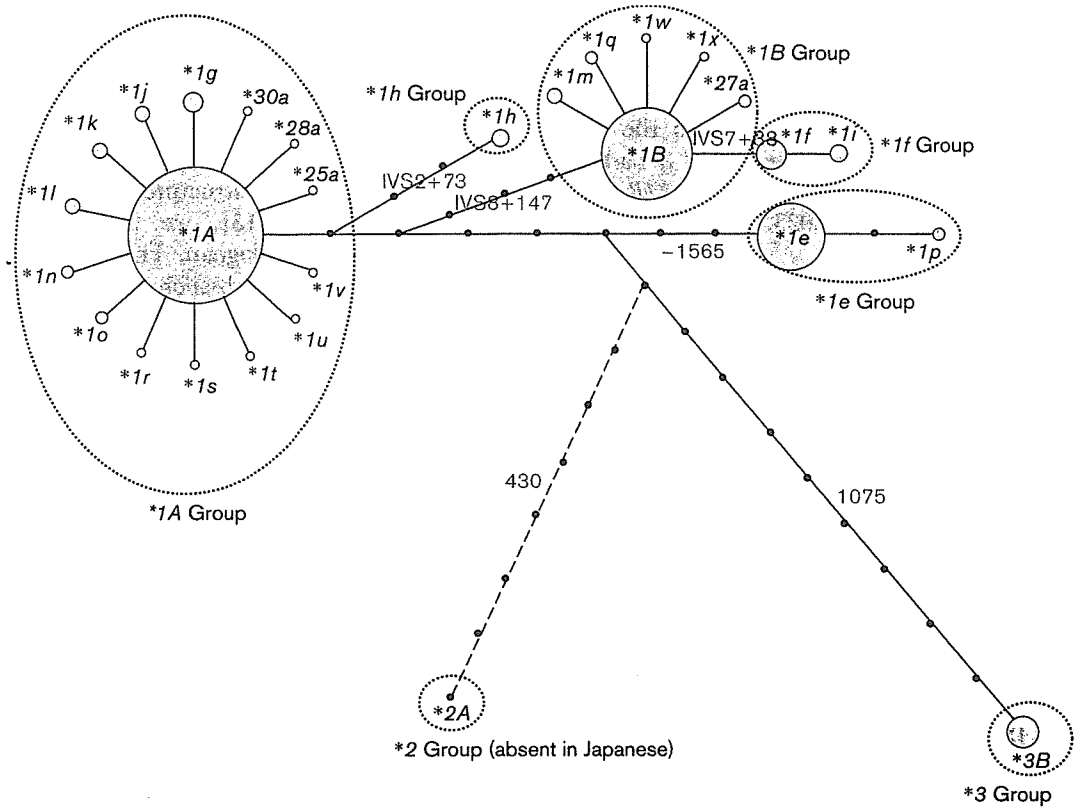
Ethnic differences in haplotype distribution

The haplotype distribution in Japanese was compared with those in other ethnic groups reported previously [10,13]. As shown in Table 4, frequencies of several *CYP2C9* haplotypes in Japanese were comparable to those in Asians, but quite different from those of corresponding haplotypes found in Caucasians and Africans. The frequency of the wild-type haplotype, **1A*, was higher in Japanese (0.489) than that in Caucasians (0.281) reported by Veenstra *et al.* [13]. The haplotype **1B*, which was tagged by IVS8 + 147C > T, was inferred at a frequency of 0.222 in Japanese. This frequency is comparable to that in Caucasians (0.175). The third dominant haplotype in Japanese, **1e* tagged by −1565C > T, was not found in Caucasians. According to Blaisdell *et al.* [10], the haplotypes harboring −1565C > T, were inferred with a frequency of 0.043 in Asians (Haplotype P) and with a frequency of 0.133 in African-Americans (Haplotype S), but not in Caucasians. Therefore, the absence of the haplotype **1e* appears to be characteristic of Caucasians. The haplotype **1f*, tagged by IVS7 + 38C > T, was not found in Caucasians. Blaisdell *et al.* [10] inferred Haplotype O harboring IVS7 + 38C > T with a frequency of 0.022 in Asians, which was comparable to that in Japanese (0.023). Therefore, the distribution of **1f* might be restricted in Asians.

A marked difference in haplotype frequencies of **1h* was observed between Japanese and Caucasians. Its frequency was approximately 20-fold higher in Caucasians than in Japanese. Blaisdell *et al.* [10] reported that the frequencies of Haplotype E harboring IVS2 + 73T > C was 0.043, 0.100 and 0.205 in Asians, African-Americans and Caucasians, respectively. Therefore, the haplotype **1h* appears to be more frequent in Caucasians and African-Americans than in Asians.

The frequency of haplotype **3B* harboring Ile359Leu in Japanese (0.027) was comparable to that in Asians (0.022; Haplotype U) reported by Blaisdell *et al.* [10], but was slightly lower than those in Caucasians reported by Veenstra *et al.* [13] (0.057; Haplotype 3), by King *et al.*

Fig. 5



Network analysis of unambiguous haplotypes in *CYP2C9*. The areas in the circles represent the approximate frequencies of each haplotype. The htSNPs that discriminate the common haplotypes are indicated in red. The **2A* haplotype found only in Caucasians was connected in this cladogram based on the report by Veenstra *et al.* [13].

Table 4 Ethnic differences in *CYP2C9* haplotypes

Haplotype tagging single nucleotide polymorphisms	Present study		Veenstra *et al.* [13]		Blaisdell *et al.* [10]			
	Haplotype	263 Japanese subjects	Haplotype	192 European-American patients	Haplotype	23 Asians	15 African-Americans	22 Caucasians
None	*1A	0.489	7 (*1A)	0.281	V (=*1A)	0.522	0.333	0.386
IVS8+147C>T	*1B	0.222	6 (*1B)	0.175	J	0.304	0.067	0.091
−1565C>T	*1e	0.118	–	–	S, P	0.043	0.133	ND
IVS7+38C>T	*1f	0.023	–	–	O	0.022	ND	ND
IVS2+73T>C	*1h	0.008	14, 15, 16, 17, 22	0.211	E	0.043	0.100	0.205
430C>T (Arg144Cys)	*2	ND	20 (*2A)	0.107	T (=*2)	ND	ND	0.068
1075A>C (Ile359Leu)	*3B	0.027	3	0.057	U (=*3B)	0.022	ND	0.159
IVS6-32T>C	–	ND	8	0.063	G	ND	ND	0.068

ND, Not detected.

[16] (0.062; Haplotype 5, **3B*), or by Morin *et al.* [18] (0.081). On the other hand, the distribution of the haplotype harboring IVS6–32T > C appears to be restricted in Caucasians. Its frequency in Caucasians was 0.063 (Haplotype 8) by Veenstra *et al.* [13] and 0.068 (Haplotype G) by Blaisdell *et al.* [10]. This result was consistent with the HapMap data where IVS6–32T > C was found in Europeans with an allele frequency of 0.067, but not in Yoruba, Han Chinese and Japanese.

Relationship between *CYP2C9* and *CYP2C19* haplotypes

The CYP2C subfamily members (*CYP2C18*, *CYP2C19*, *CYP2C9* and *CYP2C8*) are located on chromosome 10q24 as a cluster that spans approximately 400 kb. Previously, we reported the genetic variations and haplotype structure of *CYP2C19* in a Japanese population [29]. Because 253 out of the 263 subjects in this study were identical to those in the previous study on *CYP2C19*, we analysed LD patterns and the associations of haplotypes between *CYP2C9* and *CYP2C19*. Of all the 1225 pairwise $|D'|$ values between 50 common SNPs consisting of 24 in *CYP2C19* [29] and 26 in *CYP2C9* (>0.01 in their allele frequencies), 988 pairs (80%) had $|D'| > 0.90$ (data not shown), indicating an extended LD block covering both *CYP2C19* and *CYP2C9*. As shown in Table 5, 92% (228/249) of the wild-type haplotype **1A* was linked with *CYP2C19*1d*, which was the most dominant *CYP2C19* haplotype in Japanese harboring 99C > T (Pro33Pro) and 991A > G (Ile331Val). The majority of **1B* (83%, 94/113) was linked with *CYP2C19*2c*, which was the second dominant *CYP2C19* haplotype in Japanese harboring 681G > A (a splicing defect). The majority of **1e* (92%, 54/59) was linked with *CYP2C19*3b*, which was the third dominant *CYP2C19* haplotype in Japanese harboring 636G > A (Trp212X). The majority of **3B* (92%, 12/13) was linked with *CYP2C19*1e* without any non-synonymous amino acid change (Ile at codon 331). Namely, intergene haplotypes (*CYP2C19–CYP2C9* combinations) were found at the following frequencies (Table 6): *CYP2C19*1d–CYP2C9*1A* (0.451), *CYP2C19*2c–CYP2C9*1B* (0.186), *CYP2C19*3b–CYP2C9*1e* (0.107), *CYP2C19*1e–CYP2C9*3B* (0.024) and *CYP2C19*2c–CYP2C9*1f* (0.02). A strong linkage between *CYP2C19* and *CYP2C9* haplotypes indicated rare recombination between these two genes.

Discussion

The present study provides comprehensive data on genetic variations of *CYP2C9*, which encodes a clinically important enzyme that metabolizes numerous therapeutic drugs with a narrow therapeutic index. We found 62 variations, including seven novel non-synonymous ones, in 263 Japanese subjects. To assess the effects of these novel variations on both protein expression levels and enzymatic activity, we transiently expressed the recombinant protein in COS-1 cells. Although the expression levels of recombinant proteins in a mammalian expression system were low as compared with bacterial and baculovirus-mediated systems, the mammalian system is reliable in assessing functional significance of the CYP variants because of the correct protein folding assured by mammalian chaperone proteins. Indeed, the observed K_m value in our system was comparable to that in human liver microsomes when diclofenac was used as a probe drug [22,27].

The **25* (Lys118ArgfsX9), which was found in a heterozygous diabetic patient treated with glimepiride, produced an early termination codon within the C-helix. The variant protein for Lys118ArgfsX9 was not detected by Western blot analysis, suggesting that it was a null allele. Two null alleles with nonsense mutations in *CYP2C9* were already reported: **6* (Lys273ArgfsX34) [9] and **15* (Ser162X) [12]. An African-American woman with homozygous **6* showed severe phenytoin toxicity. On the other hand, an Indian woman with heterozygous **15* required a typical warfarin maintenance dose. We are now evaluating the clinical effects of Lys118ArgfsX9 on pharmacokinetic/pharmacodynamics profiles of glimepiride.

Table 5 Frequencies of common haplotype combinations of *CYP2C19* and *CYP2C9*

Haplotype Number (frequency)	*2C9*1A* 249 (0.492)	*2C9*1B* 113 (0.223)	*2C9*1e* 59 (0.117)	*2C9*3B* 13 (0.026)	*2C9*1f* 12 (0.024)
*2C19*1d* 249 (0.492)	**228 (0.451)**	–	–	–	–
*2C19*2c* 122 (0.241)	–	**94 (0.186)**	–	–	**10 (0.020)**
*2C19*3b* 58 (0.115)	–	2 (0.004)	**54 (0.107)**	–	–
*2C19*1e* 22 (0.043)	–	8 (0.016)	–	**12 (0.024)**	–
*2C19*1f* 11 (0.022)	7 (0.014)	–	–	–	–
*2C19*1g* 7 (0.014)	7 (0.014)	–	–	–	–

Number and frequencies (within parenthesis) of various haplotype combinations of *CYP2C19* and *CYP2C9* are shown as a matrix with *CYP2C19* haplotypes as rows and *CYP2C9* haplotypes as columns. Frequencies representing major combinations are shown in bold. *CYP2C19* haplotypes in a Japanese population are defined by Fukushima-Uesaka *et al.* [29].

The **26* (Thr130Arg) exhibited a drastic decrease in V_{max} against diclofenac hydroxylation without a significant change in the K_m value. Thr130 is not within the substrate recognition sites (SRSs), but is highly conserved in the CYP2C family [30]. The crystal structure of CYP2C9 showed that Thr130 in the C-helix is on the surface of the protein [31,32], suggesting little or no importance for this residue in substrate binding. Around Thr130Arg, two SNPs, **2* (Arg144Cys) in the D-helix and **14* (Arg125His) in the C-helix, were already reported, both of which are located on the exterior of the protein. The **2* causes a small decrement in V_{max} (0–35%) and little or no change in the K_m for catalysis of various substrates [5]. An *in vitro* experiment has suggested that the observed reductions in V_{max} of **2* might be associated with the altered interaction of the recombinant protein with cytochrome P450 reductase [33]. According to Delozier *et al.* [34] recombinant **14* exhibited a five-fold increase in the K_m value and a 65% decrease in the V_{max} towards tolbutamide. They suggested that loss of activity might reflect altered affinity of recombinant protein for the coenzyme because the previous site directed mutagenesis study demonstrated that the corresponding residue of CYP2B4 in the C-helix plays a prominent role in binding its redox partners, cytochrome b5 and P450 reductase [35]. Thus, we speculate that the substitution of Thr130 with the positively charged arginine might influence an electrostatic interaction between CYP2C9 and P450 reductase and/or cytochrome b5 as proposed in Arg144Cys [33] and Arg125His [34].

The **28* (Gln214Leu), which was found in a heterozygous diabetic patient, exhibited an approximately two-fold increase in the K_m value and a more than 50% decrease in the V_{max}, resulting in a 77% decrease in intrinsic clearance. Gln214, which is conserved in the CYP2C family, is located between the F- and G-helix and only five amino acids downstream of SRS 2 [30]. No site directed mutagenesis experiments were performed to examine the functional importance of the F-G loop of CYP2C9. Furthermore, there was significant inconsistency in the conformation in the F-G loop region between two crystal structures available for human CYP2C9, 1OG5 with warfarin [31] and 1R9O with flurbiprofen [32]. In the 1OG5 structure, in which seven mutations were introduced for crystallization, two short helices F′ and G′ were resolved, and Gln214 was located in the additional F′ helix. By contrast, helix F′ and helix G′ were not evident in the 1R9O structures, which exhibited a more extended conformation of the region between the F-G loop and helix A although seven amino acids (Gly214-Ser220) of the F-G loop were not included in the model. The structural difference in the F-G loop between the 1R9O and 1OG5 are likely to reflect the conformational flexibility indicative of an adaptive fit to the various substrates with different sizes, polarity and stereochemical features. In this regard, this is the first report experimentally demonstrating that Gln214 in the F-G loop affects the metabolism of diclofenac.

The **30* (Ala477Thr) was found in a heterozygous healthy volunteer. This substitution showed a similar extent of defective catalytic activity for **28* (Gln214Leu) with regard to the K_m, V_{max} and intrinsic clearance. Ala477 is within the substrate recognition site 6 (SRS 6) and forms a β4–2 sheet [30]. The importance of Phe476, which is next to Ala477, was previously proposed on the basis of a combined protein and pharmacophore model for CYP2C9 [36]. Supporting this model, Melet *et al.* [37] showed by site-directed mutagenesis that the Phe476Ile variant was two times less efficient than the wild-type CYP2C9 in terms of diclofenac 4′-hydroxylation without affecting the protein expression levels, which resulted from 3.7- and 2-fold increases in the K_m and K_{cat} values, respectively. Furthermore, the Phe476Ile variant exhibited a significant change in the regioselectivity of diclofenac hydroxylation; namely, this substrate was also 5-hydroxylated [37]. They concluded that Phe476 played a crucial role in substrate recognition and hydroxylation of diclofenac by CYP2C9, presumably via π-stacking interactions between its phenyl residues and substrate aromatic rings. Our results indicated that Ala477 located next to Phe476 might also be important for substrate recognition and hydroxylation of diclofenac. The substitution of small alanine to the nucleophilic residue, threonine, might influence the hydrophobic interaction of Phe476 with the substrates. Although we did not investigate the effects of Ala477Thr on the regioselectivity of diclofenac hydroxylation because of no availability of 3′- and 5′-hydroxy metabolites, further studies would be needed to provide detailed features of the CYP2C9 active site.

The other three novel alleles, Leu17Ile, **27* (Arg150Leu) and **29* (Pro279Thr), showed similar catalytic activities towards diclofenac hydroxylation as the wild-type. Because recombinant CYP2C proteins without their N-terminus have been found to be catalytically active, it is reasonable that a conservative substitution, Leu17Ile, in the membrane anchor region had no significant effects on both microsomal expression level and catalytic activity. Blaisdell *et al.* [10] also reported that an adjacent substitution, **7* (Leu19Ile) had no effect on catalytic activity towards tolbutamide *in vitro*. However, it should be noted that the diabetic patient carrying Leu17Ile was heterozygous for Ile359Leu, and both alleles were assigned to the same haplotype (**3c*) by an expectation-maximization algorithm. Because Ile359Leu itself is functionally definitive [5], the combined effects of Leu17Ile and Ile359Leu on clinical phenotype should be carefully estimated *in vivo*.

On the other hand, no apparent effects on the catalytic activity of **27* (Arg150Leu) were a little surprising because a substitution in the same position, **8* (Arg150His), exhibited a modest decrease in K_m and a two-fold increase in clearance of tolbutamide [10]. Because Arg150 is a surface residue of the D-helix, different electrostatic status near the substituted residues (His verses Leu) might differently influence the substrate-dependent catalytic behaviour. Pro279, located between helices H and I, is not conserved in the CYP2C family and is unlikely to play an important role in catalytic activity of CYP2C9. Accordingly, it was reasonable that **29* (Pro279Thr) did not show any functional changes.

The two previously reported alleles, **3* (Ile359Leu) and **13* (Leu90Pro), were also found in Japanese at allele frequencies of 0.03 and 0.002, respectively. Experiments performed *in vivo* and *in vitro* consistently demonstrated that **3* was associated with substantial loss of enzyme activity that resulted from decreased V_{max} and increased K_m for many CYP2C9 substrates [5]. The **13*, first detected in a Chinese poor metabolizer of lornoxicam, has been found at an allele frequency of 0.01 in a Chinese population [11]. Guo *et al.* [38] revealed that the Leu90Pro substitution markedly decrease the intrinsic clearance of lornoxicam *in vitro* and *in vivo*. Except for **3* and **13*, the other 21 *CYP2C9* alleles published on the Human CYP Allele Nomenclature Committee homepage were not found in the present study. Taken together, approximately 8% of Japanese individuals (21 out of the 263 subjects) carry one of the functionally defective alleles: **3* (Ile359Leu), **13* (Leu90Pro), **25* (Lys118ArgfsX9), **26* (Thr130Arg), **28* (Gln214Leu) and **30* (Ala477Thr). The existence of several rare but defective alleles in South-east Asian subjects was also confirmed by DeLozier *et al.* [34], such as **15* (Ser162X), **18* (Asp397Ala), **14* (Arg125His) and **16* (Thr299Ala). Therefore, defective alleles of *CYP2C9*, including **3*, occur more frequently than previously expected in Japanese and probably Asians. Thus, this could in part account for the interethnic and/or interindividual variability in metabolizing CYP2C9 substrate drugs.

Recently, Veenstra *et al.* [13] reported the first whole-gene high-resolution haplotype structures of *CYP2C9* in European-American patients administrated warfarin. They determined 23 haplotypes, only eight of which occurred at a frequency greater than 5%, indicating the overall haplotype structure of *CYP2C9* was not complex. Apart from distinctive ethnic differences in haplotype frequencies, the results of the present study are consistent with this report. The overall haplotype structure of *CYP2C9* in Japanese was also simple: only five common haplotypes with a frequency $> 2\%$ accounted for most of the haplotypes, and they can be distinguished by only four htSNPs.

As shown in Table 5, each of the five common *CYP2C9* haplotypes in Japanese (**1A*, **1B*, **1e*, **1f* and **3B*) was strongly linked with each of the four major *CYP2C19* haplotypes (**1d*, **2c*, **3b* and **1e*) reported by Fukushima-Uesaka *et al.* [29]. This result is in good agreement with a recent report by Ahmadi *et al.* [20] and Walton *et al.* [39], in which the long-range LD spanning *CYP2C19* and *CYP2C9* was identified. The *CYP2C19–CYP2C9* haplotype combination, *CYP2C19*1d–CYP2C9*1A* (0.451 in frequency), is prevalent in Japanese. It is the combination of the most dominant haplotypes of both *CYP2C19* and *CYP2C9* in Japanese that are associated with extensive metabolic phenotypes. There is no linkage between *CYP2C9*3B* and the two defective haplotypes of *CYP2C19*, *CYP2C19*2c* and *CYP2C19*3b*, suggesting that Japanese individuals do not have a haplotype simultaneously resulting in poor metabolism phenotypes for both CYP2C19 and CYP2C9. However, approximately 67% of Japanese individuals bear one or two copies of haplotypes harboring either *CYP2C19*2* (681G > A, splicing defect), *CYP2C19*3* (636G > A, Trp212X), or *CYP2C9*3* (1075A > C, Ile359Leu). The very close associations between *CYP2C19* and *CYP2C9* haplotypes could complicate pharmacogenetic studies on drugs such as phenytoin, tolbutamide and chlorpropamide that are metabolized by both CYP2C9 and CYP2C19.

In summary, we identified 62 variations, including 32 novel ones, in *CYP2C9* from Japanese subjects. Seven novel SNPs with non-synonymous substitutions were found, of which one, *CYP2C9*25* (Lys118ArgfsX9), was a null allele, and three, *CYP2C9*26* (Thr130Arg), *CYP2C9*28* (Gln214Leu) and *CYP2C9*30* (Ala477Thr), were functionally defective alleles towards diclofenac. Further clinical studies will be required to determine the clinical importance of the novel CYP2C9 alleles, including promoter variations, for the metabolism of CYP2C9 substrate drugs. Furthermore, remarkable differences in haplotype distributions among ethnic groups have highlighted the importance of ethnic-specific pharmacogenetic data.

Acknowledgements

We thank Drs Yasushi Kaburagi and Sachiko Honjo at the hospital, International Medical Center of Japan, for acquiring informed consent from the patients. We also thank Ms Chie Sudo for her secretarial assistance.

References

1 Rettie AE, Jones JP. Clinical and toxicological relevance of CYP2C9: drug–drug interactions and pharmacogenetics. *Annu Rev Pharmacol Toxicol* 2005; **45**:477–494.
2 Schwarz UI. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. *Eur J Clin Invest* 2003; **33 (Suppl 2)**:23–30.
3 Loebstein R, Yonath H, Peleg D, Almog S, Rotenberg M, Lubetsky A *et al.* Interindividual variability in sensitivity to warfarin–nature or nurture? *Clin Pharmacol Ther* 2001; **70**:159–164.
4 Grasela TH, Sheiner LB, Rambeck B, Boenigk HE, Dunlop A, Mullen PW, *et al.* Steady-state pharmacokinetics of phenytoin from routinely collected patient data. *Clin Pharmacokinet* 1983; **8**:355–364.

5 Lee CR, Goldstein JA, Pieper JA. Cytochrome P450 2C9 polymorphisms: a comprehensive review of the in-vitro and human data. *Pharmacogenetics* 2002; **12**:251–263.

6 Lee CR. CYP2C9 genotype as a predictor of drug disposition in humans. *Meth Find Exp Clin Pharmacol* 2004; **26**:463–472.

7 Imai J, Ieiri I, Mamiya K, Miyahara S, Furuumi H, Nanba E, *et al.* Polymorphism of the cytochrome P450 (CYP) 2C9 gene in Japanese epileptic patients: genetic analysis of the CYP2C9 locus. *Pharmacogenetics* 2000; **10**:85–89.

8 Dickmann LJ, Rettie AE, Kneller MB, Kim RB, Wood AJ, Stein CM, *et al.* Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans. *Mol Pharmacol* 2001; **60**:382–387.

9 Kidd RS, Curry TB, Gallagher S, Edeki T, Blaisdell J, Goldstein JA. Identification of a null allele of CYP2C9 in an African-American exhibiting toxicity to phenytoin. *Pharmacogenetics* 2001; **11**:803–808.

10 Blaisdell J, Jorge-Nebert LF, Coulter S, Ferguson SS, Lee SJ, Chanas B, *et al.* Discovery of new potentially defective alleles of human CYP2C9. *Pharmacogenetics* 2004; **14**:527–537.

11 Si D, Guo Y, Zhang Y, Yang L, Zhou H, Zhong D. Identification of a novel variant CYP2C9 allele in Chinese. *Pharmacogenetics* 2004; **14**:465–469.

12 Zhao F, Loke C, Rankin SC, Guo JY, Lee HS, Wu TS, *et al.* Novel CYP2C9 genetic variants in Asian subjects and their influence on maintenance warfarin dose. *Clin Pharmacol Ther* 2004; **76**:210–219.

13 Veenstra DL, Blough DK, Higashi MK, Farin FM, Srinouanprachan S, Rieder MJ, *et al.* CYP2C9 haplotype structure in European American warfarin patients and association with clinical outcomes. *Clin Pharmacol Ther* 2005; **77**:353–364.

14 Chen Y, Kissling G, Negishi M, Goldstein JA. The nuclear receptors constitutive androstane receptor and pregnane X receptor cross-talk with hepatic nuclear factor 4α to synergistically activate the human CYP2C9 promoter. *J Pharmacol Exp Ther* 2005; **314**:1125–1133.

15 Shintani M, Ieiri I, Inoue K, Mamiya K, Ninomiya H, Tashiro N, *et al.* Genetic polymorphisms and functional characterization of the 5′-flanking region of the human CYP2C9 gene: *in vitro* and *in vivo* studies. *Clin Pharmacol Ther* 2001; **70**:175–182.

16 King BP, Khan TI, Aithal GP, Kamali F, Daly AK. Upstream and coding region CYP2C9 polymorphisms: correlation with warfarin dose and metabolism. *Pharmacogenetics* 2004; **14**:813–822.

17 Takahashi H, Ieiri I, Wilkinson GR, Mayo G, Kashima T, Kimura S, *et al.* 5′-Flanking region polymorphisms of CYP2C9 and their relationship to S-warfarin metabolism in white and Japanese patients. *Blood* 2004; **103**:3055–3057.

18 Morin S, Bodin L, Loriot MA, Thijssen HH, Robert A, Strabach S, *et al.* Pharmacogenetics of acenocoumarol pharmacodynamics. *Clin Pharmacol Ther* 2004; **75**:403–414.

19 Solus JF, Arietta BJ, Harris JR, Sexton DP, Steward JQ, McMunn C, *et al.* Genetic variation in 11 phase I drug metabolism genes in an ethnically diverse population. *Pharmacogenomics* 2004; **5**:895–931.

20 Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, *et al.* A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005; **37**:84–89.

21 Ekins S, Maenpaa J, Wrighton SA. *In vitro* metabolism: subcelluar fractions. In: Woolf TF, editors. *Handbook of drug metabolism.* New York, NY: Marcel Dekker Inc.; 1999. pp. 363–399.

22 Leemann T, Transon C, Dayer P. Cytochrome P450TB (CYP2C): a major monooxygenase catalyzing diclofenac 4′-hydroxylation in human liver. *Life Sci* 1993; **52**:29–34.

23 Kitamura Y, Moriguchi M, Kaneko H, Morisaki H, Morisaki T, Toyama K, *et al.* Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm. *Ann Hum Genet* 2002; **66**:183–193.

24 Ferguson SS, LeCluyse EL, Negishi M, Goldstein JA. Regulation of human CYP2C9 by the constitutive androstane receptor: discovery of a new distal binding site. *Mol Pharmacol* 2002; **62**:737–746.

25 Gerbal-Chaloin S, Daujat M, Pascussi JM, Pichard-Garcia L, Vilarem MJ, Maurel P. Transcriptional regulation of CYP2C9 gene. Role of glucocorticoid receptor and constitutive androstane receptor. *J Biol Chem* 2002; **277**:209–217.

26 Ibeanu GC, Goldstein JA. Transcriptional regulation of human CYP2C genes: functional comparison of CYP2C9 and CYP2C18 promoter regions. *Biochemistry* 1995; **34**:8028–8036.

27 Yasar U, Eliasson E, Forslund-Bergengren C, Tybring G, Gadd M, Sjoqvist F, *et al.* The role of CYP2C9 genotype in the metabolism of diclofenac *in vivo* and *in vitro*. *Eur J Clin Pharmacol* 2001; **57**:729–735.

28 Takanashi K, Tainaka H, Kobayashi K, Yasumori T, Hosakawa M, Chiba K. CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. *Pharmacogenetics* 2000; **10**:95–104.

29 Fukushima-Uesaka H, Saito Y, Maekawa K, Ozawa S, Hasegawa R, Kajio H, *et al.* Genetic variations and haplotypes of CYP2C19 in a Japanese population. *Drug Metab Pharmacokinet* 2005; **20**:300–307.

30 Lewis DF. *Guide to cytochrome P450: structure and function.* New York, NY: Taylor & Francis; 2001.

31 Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* 2000; **5**: 121–131.

32 Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, *et al.* The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J Biol Chem* 2004; **279**:35630–35637.

33 Crespi CL, Miller VP. The R144C change in the CYP2C9*2 allele alters interaction of the cytochrome P450 with NADPH: cytochrome P450 oxidoreductase. *Pharmacogenetics* 1997; **7**:203–210.

34 Delozier TC, Lee SC, Coulter SJ, Goh BC, Goldstein JA. Functional characterization of novel allelic variants of CYP2C9 recently discovered in Southeast Asians. *J Pharmacol Exp Ther* 2005; **315**: 1085–1090.

35 Bridges A, Gruenke L, Chang YT, Vakser IA, Loew G, Waskell L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. *J Biol Chem* 1998; **273**: 17036–17049.

36 de Groot MJ, Alex AA, Jones BC. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J Med Chem* 2002; **45**:1983–1993.

37 Melet A, Assrir N, Jean P, Pilar Lopez-Garcia M, Marques-Soares C, Jaouen M, *et al.* Substrate selectivity of human cytochrome P450 2C9: importance of residues; 476,365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Arch Biochem Biophys* 2003; **409**:80–91.

38 Guo Y, Zhang Y, Wang Y, Chen X, Si D, Zhong D, *et al.* Role of CYP2C9 and its variants (CYP2C9*3 and CYP2C9*13) in the metabolism of lornoxicam in humans. *Drug Metab Dispos* 2005; **33**:749–753.

39 Walton R, Kimber M, Rockett K, Trafford C, Kwiatkowski D, Sirugo G. Haplotype block structure of the cytochrome P450 CYP2C gene cluster on chromosome 10. *Nat Genet* 2005; **37**:915–916.

doi: 10.1111/j.1469-1809.2006.00260.x

Genetic Variations and Haplotype Structures of the *ABCB1* Gene in a Japanese Population: An Expanded Haplotype Block Covering the Distal Promoter Region, and Associated Ethnic Differences

K. Sai[1,2,*], M. Itoda[1], Y. Saito[1,3], K. Kurose[1,4], N. Katori[1,5], N. Kaniwa[1,4], K. Komamura[7,8], T. Kotake[9], H. Morishita[9], H. Tomoike[7], S. Kamakura[7], M. Kitakaze[7], T. Tamura[10], N. Yamamoto[10], H. Kunitoh[10], Y. Yamada[11], Y. Ohe[10], Y. Shimada[11], K. Shirao[11], H. Minami[12], A. Ohtsu[13], T. Yoshida[14], N. Saijo[15], N. Kamatani[16], S. Ozawa[1,6] and J. Sawada[1,3]

[1]*Project Team for Pharmacogenetics,* [2]*Division of Xenobiotic Metabolism and Disposition,* [3]*Biochemistry and Immunochemistry,* [4]*Division of Medicinal Safety Science,* [5]*Division of Drugs,* [6]*Division of Pharmacology, National Institute of Health Sciences, Tokyo, 158-8501*

[7]*Division of Cardiology,* [8]*Department of Cardiovascular Dynamics Research Institute,* [9]*Department of Pharmacy, National Cardiovascular Center, Suita, 565-8565*

[10]*Thoracic Oncology Division,* [11]*Gastrointestinal Oncology Division, National Cancer Center Hospital,* [14]*Genetics Division, National Cancer Center Research Institute, Tokyo, 104-0045*

[12]*Division of Oncology/Hematology,* [13]*Division of GI Oncology/Digestive Endoscopy,* [15]*National Cancer Center Hospital East, Kashiwa, 277-8577 and*

[16]*Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, 162-0054, Japan*

Summary

As functional *ABCB1* haplotypes were recently reported in the promoter region of the gene, we resequenced the *ABCB1* distal promoter region, along with other regions (the enhancer and proximal promoter regions, and all 28 exons), in a total of 533 Japanese subjects. Linkage disequilibrium (LD) analysis based on 92 genetic variations revealed 4 LD blocks with the same make up as previously described (Blocks −1, 1, 2 and 3), except that Block 1 was expanded to include the distal promoter region, and that a new linkage between polymorphisms −1789G>A in the distal promoter region and IVS5 + 123A>G in intron 5 was identified. We re-assigned Block 1 haplotypes, and added novel haplotypes to the other 3 blocks. The reported promoter haplotypes were further classified into several types according to tagging variations within Block 1 coding or intronic regions. Our current data reconfirm the haplotype profiles of the other three blocks, add more detailed information on functionally-important haplotypes in Block 1 and 2 in the Japanese population, and identified differences in haplotype profiles between ethnic groups. Our updated analysis of *ABCB1* haplotype blocks will assist pharmacogenetic and disease-association studies carried out using Asian subjects.

Keywords: *ABCB1*, P-gp, haplotype

*Correspondence to: Dr. Kimie Sai, Division of Xenobiotic Metabolism and Disposition, National Institute of Health Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo, 158-8501, Japan. Tel. +81-3-3700-9478; Fax: +81-3-3707-6950, E-mail: sai@nihs.go.jp

Introduction

The *ABCB1* gene, encoding p-glycoprotein (P-gp)/multidrug resistance protein 1 (MDR1), is located on chromosome 7q21-q31 and consists of 28 exons. P-gp (1280 amino acids), a member of the

ATP-binding cassette (ABC) transporter superfamily, is a large transmembrane glycoprotein that consists of two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs). P-gp was initially identified as a component of the multidrug resistance phenotype in cancer cells (Riordan *et al.* 1985), but was later found to be widely expressed in normal epithelial cells of tissues such as the liver, intestine, kidneys, and the blood-brain and testis barriers, as well as in lymphocytes (Fojo *et al.* 1987; Cordon-Cardo *et al.* 1989). It is thought that P-gp plays a role in the protection of these tissues against structurally-unrelated toxic xenobiotics, and can modify the oral bioavailability and renal secretion of a variety of drugs (Hoffmann & Kroemer, 2004). Multiple other physiological functions of P-gp have also been suggested in lipid transport (van Helvoort *et al.* 1996), cholesterol metabolism (Debry *et al.* 1997), inhibition of ceramide-induced apoptosis (Liu *et al.* 2001), and the initiation of immune responses by cytokine release (Drach *et al.* 1996). Moreover, reduced P-gp expression has been linked to cancer (Siegsmund *et al.* 2002) and other diseases such as Parkinson's disease (Furuno *et al.* 2002) and ulcerative colitis (Schwab *et al.* 2003).

With recent advances in genomics research there has been an increasing number of pharmacogenetic studies focused on the *ABCB1* gene. Hoffmeyer *et al.* (2000) showed that a synonymous 3435C>T mutation in exon 26 was associated with reduced P-gp expression in the duodenum, and increased plasma levels of digoxin following its oral administration in healthy volunteers. Thus, the 3435C>T single nucleotide polymorphism (SNP) has become the focus of much attention. However, reports on the role of this common SNP have been very inconsistent, which suggests that other functional polymorphisms may be linked with 3435C>T (Kim, 2002). Further studies revealed that 3435C>T was closely linked to other common polymorphisms, such as 1236C>T (silent) at exon 12 and 2677G>T (Ala893Ser) at exon 21, and that the combinations of these SNPs (i.e. haplotypes) differed greatly between ethnic groups (Kim *et al.* 2001; Kroetz *et al.* 2003; Tang *et al.* 2002, 2004). While an *in vitro* functional study on the nonsynonymous 2677G>T (Ala893Ser) SNP at exon 21 showed that 2677G>T was associated with enhanced P-gp activity (Kim *et al.* 2001), other studies found no association (Kimchi-Sarfaty *et al.* 2002; Morita *et al.* 2003; Kroetz *et al.* 2003). One of these latter studies also revealed that another nonsynonymous SNP, 2677G>A (Ala893Thr), had no impact on P-gp function (Morita *et al.* 2003). Yet several clinical studies have shown that the haplotypes 2677T-3435T and 1236T-2677T-3435T are associated with reduced P-gp activity (Johne *et al.* 2002; Kurata *et al.* 2002; Chowbay *et al.* 2003; Wong *et al.* 2005), and that 2677A-bearing subjects exhibit higher P-gp activity (Yi *et al.* 2004). Studies that found no association between these *ABCB1* SNPs and P-gp expression levels (Goto *et al.* 2002), and other conflicting results, have been summarized in recent review articles (Kim, 2002; Ieiri *et al.* 2004).

Recently, *ABCB1* gene promoter region haplotypes were reported by two Japanese research groups, and revealed the existence of functional haplotypes that resulted in altered P-gp expression (Taniguchi *et al.* 2003; Takane *et al.* 2004). In these studies, haplotypes that included −1789G>A alone or in combination with −145C>G were associated with decreased P-gp expression. However, the reported effects of haplotypes carrying −129T>C and two other linked SNPs on P-gp expression were contradictory, showing reduction and enhancement.

From these findings it is clear that the establishment of detailed *ABCB1* gene haplotype profiles specific for each ethnic group is important. We previously conducted haplotype analysis on 145 Japanese subjects by dividing the *ABCB1* gene into 4 blocks, one of which included the proximal promoter region, and revealed that the *2 haplotype in Block 2, which harbours 1236C>T, 2677G>T and 3435C>T, showed a strong association with reduced renal clearance of irinotecan and its metabolites (Sai *et al.* 2003). However, recent findings on the functional distal *ABCB1* promoter region prompted us to identify the extended haplotypes that encompassed the above promoter region in a larger Japanese population.

In this study, we sequenced the distal *ABCB1* gene promoter regions from 533 Japanese subjects. This region covered approximately 2.5 kb upstream from the translational initiation site, adjacent to the previously described Block 1 region. We found that the promoter region SNPs were closely linked with SNPs located over a relatively wide range (up to intron 5) in Block 1, such

Table 1 Additional primers used for sequencing of the *ABCB1* gene promoter region

Primer name	Forward primer (5′ to 3′)	Primer name	Reverse primer (5′ to 3′)
First amplification[a]			
MDR1-1ZF1	CCTGCTCTGTTTTTCACCGT	MDR1-1ZR1	ATTGGTTTCCTCTATGCAGA
Second amplification			
MDR1-P1F	GAGAGGGACTACTGGTTAGC	MDR1-P1R	TGGTCCATCTGGGGTAAATG
MDR1-P2F	AAGGACTGTTGAAAGTAGCA	MDR1-P2R	TTTGAGACGGAGTCTTGCTT
MDR1-P3F	CAGAGATCATAGGCACAAAT	MDR1-P3R	AAACTTCAGACGTCAGATCA
MDR1-P4F	GAAACATCCTCAGACTATGC	MDR1-P4R	CAGGAGGAATGTTCTGGCTT
Sequencing			
MDR1-P5F	ATTTCTTTGAAGTGCTTGGC	MDR1-P5R	GCCACCACCACTTCTGTCAA
MDR1-P6F	GATCTTTACCTGATGCTCAA	MDR1-P6R	GTGCCTATGATCTCTGTTTT
MDR1-P7F	AGCTCACGCCTGTAATCCCT	MDR1-P1R	TGGTCCATCTGGGGTAAATG
MDR1-P4F	GAAACATCCTCAGACTATGC	MDR1-P8R	AGGAAAAGTACGTGCAATCT
MDR1-P9F	ACGTACTTTTCCTCAGTTTG	MDR1-P9R	ACACGTCTTTCAAAGTTCAC

Other primer sets used were as previously reported (Sai *et al.* 2003).
[a]The same set as previously used for the enhancer and promoter regions.

that it was necessary to re-evaluate the functional significance of Block 1 haplotypes. We also sequenced the same regions as covered by the previous study, including the enhancer region (Geick *et al.* 2001) and all exons and surrounding introns, for an additional 388 subjects. These results allowed us to add novel haplotypes to three other blocks. Lastly, we performed a network analysis on the haplotypes obtained in each block and compared the profile of *ABCB1* haplotypes in Japanese with those of other ethnic groups (Kroetz *et al.* 2003; Takane *et al.* 2004).

Materials and Methods

DNA Samples

All 533 Japanese subjects were patients with either ventricular tachycardia (121 subjects) who were administered an anti-arrhythmic drug (amiodarone) and/or β-blockers, or with various cancers (412 subjects) who were administered an anti-cancer drug (paclitaxel or irinotecan). Genomic DNA was extracted directly from blood leukocytes. This study was approved by the ethical review boards of the National Cardiovascular Center, the National Cancer Center, and the National Institute of Health Sciences. Written informed consent was obtained from all subjects.

DNA Sequencing

Amplification and sequencing of the *ABCB1* gene were performed as previously described (Sai *et al.* 2003), except that the region sequenced included the promoter region up to 2.5 kb upstream from the translational initiation site. For the promoter region, PCR amplification was first performed using the previous primer set that covered from 7 kb upstream of the transcription site to exon 3, and then new primer sets were used for the second PCR and sequencing (Table 1). Amplification and sequencing primers for the other regions and the PCR conditions used were the same as previously reported (Sai *et al.* 2003). Genbank NT_007993.14 was used as the reference sequence. Nucleotide positions were based on cDNA sequence as previously described, with the adenine of the translational initiation site at exon 2 numbered as +1. For 5′-flanking variations intron 1 was skipped for numbering nucleotide positions.

Haplotype and Network Analyses

Linkage disequilibrium (LD) analysis was performed using SNPAlyze software (Dynacom Co., Yokohama, Japan). According to the LD pattern we divided the *ABCB1* gene into 4 blocks following the previously described block partitioning, except for a changed border between Block 1 and Block 2 (IVS5 + 123A>G was shifted from Block 2 to Block 1). Diplotype configurations (combinations of haplotypes) in each block were inferred by LDSUPPORT software, which determined the posterior probability distribution of diplotype configurations for each subject based on estimated haplotype frequencies (Kitamura *et al.* 2002). As Block 1 was expanded we re-defined the Block 1 haplotypes.

For Block 2 haplotypes the previously defined *8c was deleted due to a shift of IVS5 + 123A>G to Block 1. For the rest of the haplotypes we followed the haplotype nomenclature used in our previous study (Sai *et al.* 2003) and added the newly-identified haplotypes consecutively. In our nomenclature the group of haplotypes without amino acid changes or marker SNPs in Block 2 (1236C>T, 2677G>T/A and 3435>T) was defined as *1, and haplotype groups bearing nonsynonymous SNPs or marker SNPs in Block 2 were consecutively numbered as described previously (Sai *et al.* 2003). Novel haplotypes within each haplotype group were designated in descending order of frequency. Haplotypes inferred in only one patient, or ambiguously defined, were described with "?", and some rare variations described as "Others" in Figures 3-5. To allow comparison with previous reports (Taniguchi *et al.* 2003; Takane *et al.* 2004) an additional classification for Block 1 haplotypes was given in Fig. 7, based on marker SNPs of the promoter region (−1789G>A, −1461_−1457delCATCC, −371A>G, −145C>G and −129T>C).

Network analysis of haplotypes was performed to obtain cladograms using Network 4.1.0.9 (www.fluxus-engineering.com.). Network calculations were based on algorithms of the reduced median network (for Blocks −1, 1 and 3) or the median joining network (for Block 2). Haplotypes inferred in only one patient were omitted from the network analysis due to their low predictability.

Results

Additional Genetic Variations

In this study we sequenced the distal promoter region covering approximately 2.5 kb upstream of the translational initiation site in exon 2 in 533 Japanese subjects. We also re-sequenced the enhancer region, and all 28 exons and surrounding regions (the same regions that were sequenced in the previous paper), in an additional 388 subjects. A total of 92 genetic variations were detected in the entire region sequenced in this study. All of the allelic frequencies were in Hardy-Weinberg equilibrium. Since we did not find any apparent differences in SNP frequencies between the two disease types ($P \geq 0.2233$; Fisher's exact test), the data from all subjects were analyzed as one group.

In addition to the variations reported in our previous study we detected 44 further variations, including 35 novel variations, as listed in Table 2. Novel variations included 8 nonsynonymous substitutions: 49T>C(F17L), 144G>T(K48N), 304G>C(G102R), 1342G>A(E448K), 1804G>A(D602N), 2359C>T (R787W), 2719G>A(V907I) and 3043A>G(T1015A); and 2 synonymous substitutions: 354C>T (Y118Y) and 447A>G(K149K); with frequencies ranging from 0.001 to 0.005. Other novel variations in the 5′-flanking region were 11 nucleotide substitutions and one deletion, while in the intronic regions there were 11 nucleotide substitutions, one deletion, and one insertion (Table 2).

The highly polymorphic variations 1236C>T, 2677G>T, 2677G>A, and 3435C>T were detected at frequencies of 0.572, 0.410, 0.183, and 0.440, respectively, which was consistent with our previous observations (Sai *et al.* 2003). In the newly-sequenced promoter region the reported polymorphic variations −1847T>C, −1789G>A, −1461'−1457del CATCC, and −1347T>C were found at frequencies of 0.084, 0.204, 0.030, and 0.084, respectively, which were comparable with frequencies in Japanese in previous reports (Taniguchi *et al.* 2003; Takane *et al.* 2004).

LD analysis was performed using the 92 detected genetic variations, and pairwise rho square (r^2) values for the representative 46 polymorphisms (alleles detected in 5 or more chromosomes), and the results are shown in Fig. 1. With the additional distal promoter region sequence close linkage relationships were observed between −1847T>C, −1347T>C, −371A>G, −129T>C, IVS3 + 36C>T and IVS5 + 76T>G. A close linkage was also detected between −1789G>A in the promoter region and IVS5 + 123A>G in intron 5 (formerly classified as Block 2). Based on these linkage relationships we changed the previous border between Block 1 and Block 2, such that IVS5 + 123A>G was now classified as part of Block 1. The other linkage profiles were the same as previously described, confirming the previous partitioning between Blocks 2 and 3. Similarly, the enhancer region at around 7 kb

Table 2 Additional *ABCB1* variations detected in Japanese

Block	SNP ID		Site	Position		Nucleotide change	Amino acid change	Frequency
	This study[a]	Reference		NT_007933.14	cDNA-based			
Block −1	MPJ6_AB1078	(novel)	5′-Flanking	12472468_12472461	−8128_−8121	GTAAAGTCAGATC**TAACCAA**/-CTGTTCATTGGT		0.002
	MPJ6_AB1079	(novel)	5′-Flanking	12466729	−2389	CTCCCATAGATA**C**/**T**ATATAGAACAGA		0.001
	MPJ6_AB1080	b)	5′-Flanking	12466680	−2340	ATGTTGTCAGAG**T**/**C**ATAGACAAGTTG		0.001
	MPJ6_AB1081	(novel)	5′-Flanking	12466659	−2319	GTTGGTGAAATG**G**/**T**CTACATGAGAGC		0.001
	MPJ6_AB1072	b,c)	5′-Flanking	12466187	−1847	GTTTAGGGAGGGT/**C**TTAAGGCCATTC		0.084
	MPJ6_AB1073	rs12720464[d]	5′-Flanking	12466129	−1789	AATGAAAGGTGA**G**/**A**ATAAAGCAACAA		0.204
	MPJ6_AB1082	(novel)	5′-Flanking	12466065	−1725	AAGATTAAAAACG/**A**CATGTAATGAAG		0.001
	MPJ6_AB1083	(novel)	5′-Flanking	12465983	−1643	CAGTGAACAATG**C**/**T**TGTACACTTGCA		0.001
	MPJ6_AB1084	(novel)	5′-Flanking	12465806	−1466	GGTCAGGAGATC**A**/**G**AGACCATCCTGG		0.002
	MPJ6_AB1085	c)	5′-Flanking	12465801_12465797	−1461_−1457	GGAGATCAAGAC**CATCC** /-TGGCTAACACAG		0.030
Block 1	MPJ6_AB1074	b,c)	5′-Flanking	12465687	−1347	GCAGGAGAATGGT/**C**GTGAACCCGGGA		0.084
	MPJ6_AB1086	(novel)	5′-Flanking	12465619	−1279	CCTGGGCGACAA**A**/**G**GCAAGACTCCGT		0.004
	MPJ6_AB1075	b,c)	5′-Flanking	12465494	−1154	AGAAAAAATTAT/**C**GGCTTTGAAGTA		0.001
	MPJ6_AB1087	(novel)	5′-Flanking	12465444	−1104	ATCCTCAGACTAT/**C**GCAGTAAAAAC		0.001
	MPJ6_AB1088	(novel)	5′-Flanking	12465421	−1081	ACAAAGTGATTTT/**C**CTTCTTCTAAAC		0.002
	MPJ6_AB1089	(novel)	5′-Flanking	12465405	−1065	CTTCTAAACTTAT/**C**GCAATAAACTGA		0.001
	MPJ6_AB1090	(novel)	5′-Flanking	12465326	−986	TCCTCTATGTTC**A**/**G**TAAGAAGTAAGA		0.001
	MPJ6_AB1091	(novel)	5′-Flanking	12464967	−627	TTATCATCAAATA/**G**AAGGATGAACAG		0.002
	MPJ6_AB1092	(novel)	Exon 2	12463728	49	AAGAAGAACTTTT/**C**TTAAACTGAACA	F17L	0.001
	MPJ6_AB1093	(novel)	Exon 4	12449246	144	TTGGCTTGACAA**G**/**T**TTGTATATGGTG	K48N	0.001
	MPJ6_AB1094	(novel)	Exon 5	12433798	304	ATCAATGATACA**G**/**C**GGTTCTTCATGA	G102R	0.005
	MPJ6_AB1095	(novel)	Exon 6	12430553	354	TGCCTATTATTA**C**/**T**AGTGGAATTGGT	Y118Y	0.001
	MPJ6_AB1096	(novel)	Exon 6	12430460	447	CAAAATTAGAAAA/**G**CAGTTTTTTCAT	K149K	0.002
	MPJ6_AB1097	(novel)	Exon 12	12413771	1342	TATGACCCCACA**G**/**A**AGGGGATGGTGA	E448K	0.001

Table 2 Continued.

Block	SNP ID: This study[a]	SNP ID: Reference	Site	Position: NT_007933.14	Position: cDNA-based	Nucleotide change	Amino acid change	Frequency
	MPJ6_AB1052	e)	Intron 12	12413746	IVS12 +17	GATGACCCATGC**G/A**AGCTAGACCCTG		0.006
	MPJ6_AB1098	(novel)	Intron 12	12413720	IVS12 +43	GGTGATCAGCAG**T/G**CACATTGCACAT		0.001
	MPJ6_AB1099	(novel)	Intron 13	12413353	IVS13 +90	CTACTATAAATC**G/A**GAAGAAGGGAAA		0.001
	MPJ6_AB1100	(novel)	Exon 15	12409538	1804	ATCGCTGGTTTC**G/A**ATGATGGAGTCA	D602N	0.002
	MPJ6_AB1101	(novel)	Intron 15	12408686	IVS15 −95	GTTACTAAACAA**A/G**TTGCTGTTTTCC		0.001
	MPJ6_AB1065	(novel)	Intron 16	12408363	IVS16 +52	CTGTGGTTCCCTA**A/C**GTTTGGTGGGCT		0.003
	MPJ6_AB1102	(novel)	Intron 16	12407939	IVS16 −72	TCCTTTACTAATT**A/T**TTGTGCGTATG		0.001
	MPJ6_AB1103	(novel)	Intron 18	12404862	IVS18 +87	AGTGTAATTGGCC**C/T**TTTAGTAGAAC		0.001
Block 2	MPJ6_AB1104	(novel)	Exon 19	12402898	2359	ATCCTCACCAAG**C/T**GGCTCCGATACA	R787W	0.001
	MPJ6_AB1105	(novel)	Intron 19	12400221	IVS19 −88	GGGGTATAAGTAT**T/C**AAACAAAACTGA		0.001
	MPJ6_AB1106	(novel)	Intron 20	12395242	IVS20 −153	TTCTTACTGTAG**A/G**AACTCAATAAAC		0.001
	MPJ6_AB1107	(novel)	Intron 20	12395172	IVS20 −83	GAATATAGTCTC**A/G**TGAAGGTGAGTT		0.001
	MPJ6_AB1108	(novel)	Intron 21	12384544_12384541	IVS21−73_−76	TTATTTTCATTA**GTCT/-**GTTTTATAGAAT		0.003
	MPJ6_AB1067	(novel)	Exon 22	12384435	2719	AACTTCCGAACC**G/A**TTGTTTCTTTGA	V907I	0.002
	MPJ6_AB1109	(novel)	Intron 22	12384359	IVS22 +9	ACAGGTAATAAC**C/T**GCTGAAGAGTGG		0.001
	MPJ6_AB1076	f)	Exon 24	12380229	2956	GTCTTTGGTGCC**A/G**TGGCCGTGGGGC	M986V	0.001
	MPJ6_AB1110	(novel)	Exon 24	12380142	3043	ATCATTGAAAAA**A/G**CCCCTTTGATTG	T1015A	0.001
	MPJ6_AB1111	(novel)	Intron 26	12372831_12372834	IVS26 +33_36	ACAGCCTGGGAG**-/CATG**TGGCAGCCTCTC		0.001
Block 3	MPJ6_AB1112	(novel)	Intron 26	12369713	IVS26 −78	ATATAGAATCGT**C/G**TATCCTACTTTC		0.001
	MPJ6_AB1077	rs2235051[d]	Exon 28	12367931	3747	GTTTCAGAATGG**C/G**AGAGTCAAGGAG	G1249G	0.002

All *ABCB1* genetic variations in the above list and detected in the previous study (Sai *et al.* 2003) were used for the haplotype analysis in this study.

[a]SNP ID assigned by our project team (MPJ-6).

[b]Taniguchi *et al.* 2003.

[c]Takane *et al.* 2004.

[d]NCBI dbSNP

[e]Itoda *et al.* 2002.

[f]Tanabe *et al.* 2001.