Table 2.

SNP ID		Pharm GKB[b]	Refe-rence	Location	Position		Nucleotide change	Amino acid change	Frequency		
This Study	dbSNP (NCBI)				NT_010393.15	From the translational initiation site or from the end of the nearest exon[c]			Total (n = 153)	Diabetic patients (n = 86)	Healthy volunteer (n = 67)
MPJ6_AC1058	rs28363989	#		Intron 16	7486627	IVS16 + 213	GATTTTGGTATCA > GTTTATTTCCATC		0.069	0.081	0.052
MPJ6_AC1059	rs4148356	#	11	Exon 17	7490354	2168	ATGATTCTCTCCG > AAGAAAACATCCT	R723Q	0.065	0.076	0.052
MPJ6_AC1060[a]				Intron 17	7490601	IVS17 + 123	GGCTCAGCTGGGC > TGGCTCTGCTGCA		0.007	0.012	0.000
MPJ6_AC1061		#	23	Intron 18	7497302_7497303	IVS18-39_-38	GCAGTCTCACAC/delAT/GTGCACTCACGT		0.065	0.076	0.052
MPJ6_AC1062	rs2074087	#	11	Intron 18	7497311	IVS18-30	CACATGTGCACTG > CACGTGGCCGGGT		0.245	0.233	0.261
MPJ6_AC1063[a]				Exon 19	7497410	2530	GTCATGAGTGGCG > AGCAAGATCTCTG	G844S	0.003	0.000	0.007
MPJ6_AC1064[a]				Intron 19	7497577	IVS19 + 53	GCACCTTGAAGGG > CCCACATTGGCCT		0.003	0.006	0.000
MPJ6_AC1065	rs4148369		25	Intron 19	7509388	IVS19-175	GATACCACCTGCC > TCCACAACCAGAC		0.098	0.093	0.104
MPJ6_AC1066[a]				Intron 21	7513820	IVS21 + 11	AGGTGAGATTCGC > GTCCTTAAGTGAT		0.003	0.006	0.000
MPJ6_AC1067	rs4780592			Intron 21	7518220	IVS21-91	CAGCTGGGTGGCG > ACAGTGCTGGTGA		0.284	0.279	0.291
MPJ6_AC1068	rs4780593			Intron 21	7518222	IVS21-89	GCTGGGTGGCACG > AGTGCTGGTGAAG		0.284	0.279	0.291
MPJ6_AC1069	rs4238623			Intron 21	7518240	IVS21-71	GGTGAAGCCCCCA > GACCTTGTGGGGC		0.474	0.448	0.507
MPJ6_AC1070	rs11282335		25	Intron 21	7518268_7518269	IVS21-43_-42	GCTGGGGCTGGG/insGCTGGG/TGCGTGCATGTG		0.526	0.552	0.493
MPJ6_AC1071	rs3887893	#	25	Intron 22	7518580	IVS22 + 62	TTTGTCTAATTAT > CAGAAATGGATCC		0.480	0.500	0.455
MPJ6_AC1072	rs28363990	#	20	Intron 22	7521659	IVS22-43	GTGCCTGGTCAGC > TTCCCTCTCTGCA		0.049	0.041	0.060
MPJ6_AC1073			11	Exon 23	7521795	3173	ACAGCATCCTGCG > AGTCACCCATGAG	R1058Q	0.003	0.006	0.000
MPJ6_AC1074[a]				Intron 23	7522149	IVS23 + 137	TTTGTCAGTTTCG > AAATACCTAAATT		0.003	0.006	0.000
MPJ6_AC1075[a]				Intron 23	7528780	IVS23-131	CACCCCTGTGAGG > CGCAGCCCGGCTC		0.010	0.017	0.000
MPJ6_AC1076	rs4148377		25	Exon 24	7528970	3450	CAGCCGCTCCCCG > AGTCTATTCCCAT	P1150P	0.003	0.006	0.000
MPJ6_AC1077[a]				Exon 24	7529010	3490	CTGGGGGTCAGCG > ATCATTCGAGCCT	V1164I	0.003	0.000	0.007
MPJ6_AC1078[a]				Exon 24	7529070	3550	CTGAAGGTGGACG > AAGAACCAGAAGG	E1184K	0.003	0.000	0.007
MPJ6_AC1079			23	Intron 25	7531965	IVS25 + 114	ACTTGAGAGGTAC > TGGAGTTTGAGGA		0.016	0.023	0.007
MPJ6_AC1080[a]				Intron 26	7533038	IVS26 + 191	AAAATAGTTTACC > TGGCTTTACCCAA		0.003	0.006	0.000
MPJ6_AC1081	rs2270490			Intron 26	7538695	IVS26-30	GGACTGGAAATTC > GCTTACTCTCTCC		0.003	0.006	0.000
MPJ6_AC1082		#		Intron 26	7538701_7538711	IVS26-24_-14	GAAATTCCTTAC/delTCTCTCCCTTC/ACTGCGATCGAA		0.007	0.012	0.000
MPJ6_AC1083[a]				Exon 27	7538806	3901	AACTACTGCCTGC > TGCTACCGAGAGG	R1301C	0.003	0.006	0.000
MPJ6_AC1084[a]				Intron 27	7538969	IVS27 + 98	CCCAGTCACTCAC > TGGCTCCACACCT		0.003	0.000	0.007
MPJ6_AC1085	rs212081	#	25	Intron 27	7539050	IVS27 + 179	AGAGCGCATACAG > ACTTGCAGAAGTG		0.294	0.285	0.306
MPJ6_AC1086	rs2239330	#	7	Exon 28	7541321	4002	AGCTGGGAAGTCG > ATCCCTGACCCTG	S1334S	0.196	0.203	0.187
MPJ6_AC1087[a]				Intron 28	7541458	IVS28 + 14	TGGGGTCTGGGTG > ATGGCCCAGGGGG		0.003	0.000	0.007
MPJ6_AC1088	rs7198430			Intron 28	7543148	IVS28-266	TTTTACTAGAGAC > GAGGGTGTTGCCA		0.320	0.267	0.388
MPJ6_AC1089[a]				Intron 28	7543246	IVS28-168	ACAGGCGTGAACC > TACCGTACCTGGC		0.007	0.006	0.007
MPJ6_AC1090	rs212087	#	25	Intron 28	7543369	IVS28-45	ATCCATGTCAGCG > ATGACACAGGTGT		0.304	0.326	0.276
MPJ6_AC1091	rs4148379		7	Intron 29	7545287	IVS29-13	TCCTGGTTTTTT/delT/CTTCCGGTCAAG		0.314	0.267	0.373
MPJ6_AC1092	rs212088	#	9	Intron 30	7545512	IVS30 + 18	GCCACTGGCACAG > ATGGCCTCTAGGC		0.291	0.314	0.261
MPJ6_AC1093[a]				Exon 31	7548123	4502	TGATCGTCTTGGA > GCAAAGGAGAAAT	D1501G	0.003	0.006	0.000
MPJ6_AC1094	rs3743527		25	3′-UTR	7548760	*543[d] (5139)	ATCATTTTCTCCC > TCTTGGCAGTGTC		0.310	0.267	0.366
MPJ6_AC1095	rs129081	#	25	3′-UTR	7549018	*801[d] (5397)	CCCACCCACCCCC > GACTCCAGGCTTT		0.395	0.419	0.366
MPJ6_AC1096	rs212090		25	3′-UTR	7549083	*866[d] (5462)	CTGTTATTACTGT > ATCCCACCATGAT		0.255	0.267	0.239
MPJ6_AC1097[a]				3′-UTR	7549275_7549276	*1058_1059[d] (5654_5655)	TTGTTCTTTTTT/insT/CTTACCACCTCT		0.003	0.006	0.000

[a]Novel variations detected in this study.

[b]Variations included in the PharmGKB database were marked with "#".

[c]Exon-intron boundary and amino acid numbering were based on the isoform 1.

[d]Numbered from the termination codon TGA.

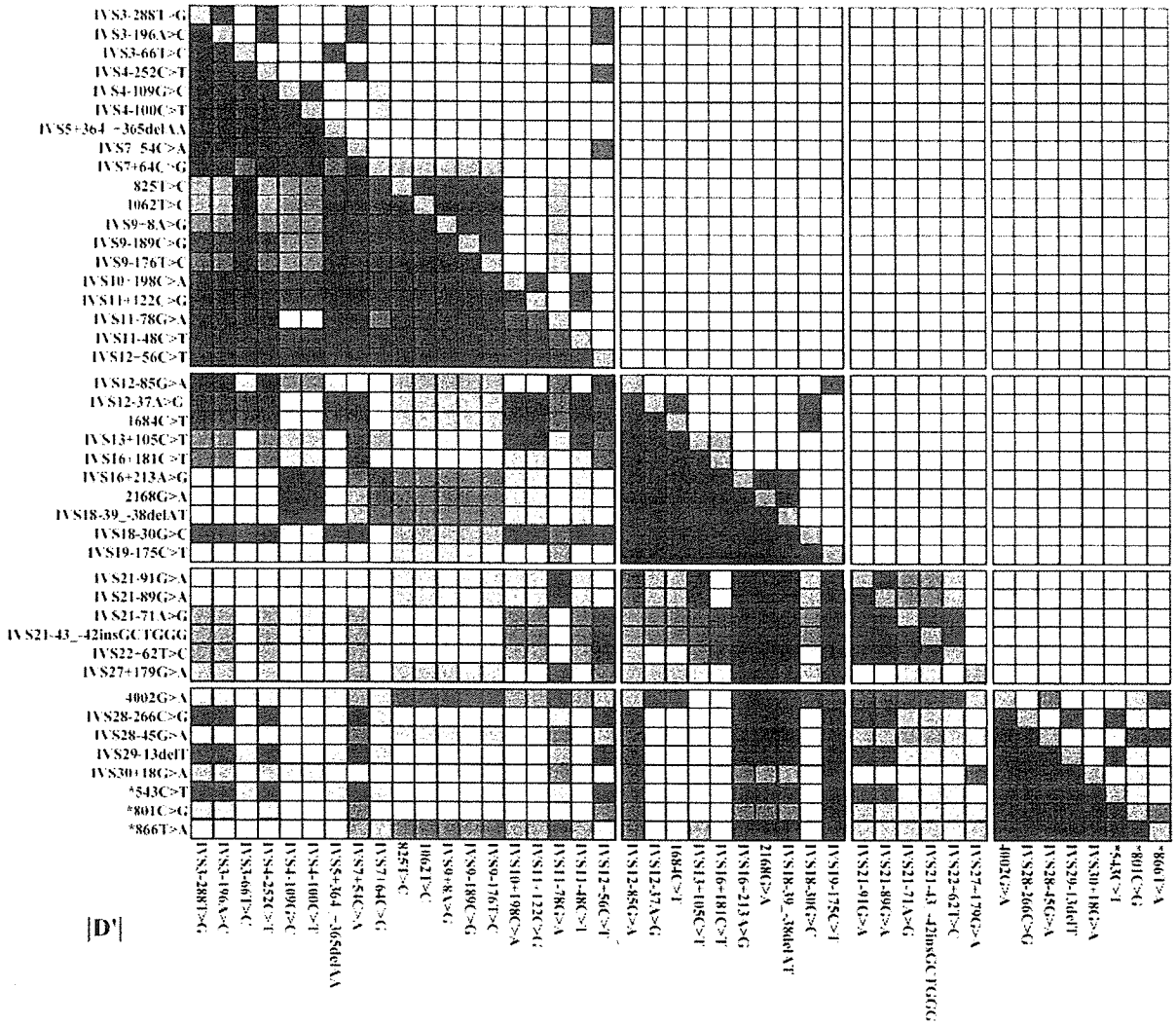

Fig. 1. Linkage disequilibrium (LD) analysis of *ABCC1*. Pairwise LD is expressed as r^2 (upper right) and |D'| (lower left) values (from 0 to 1) by 10-graded blue colors. Denser color represents closer linkage.

assay.[22] However, in Japanese and Chinese,[23] higher numbers of repeats were detected. The effects of these expanded repeats are currently unknown. We also detected one novel and eight known synonymous variations. Of these, 825T>C (Val275Val), 1684C>T (Leu562Leu), and 4002G>A (Ser1334Ser) were also detected in Caucasians and their frequencies were almost comparable to those in Japanese (**Table 2**).[24] Wang *et al.* (2003) sequenced the *ABCC1* gene of 27 Chinese subjects.[21] Of the 32 SNPs detected by them, 21 were also found in this study. The frequencies of common SNPs were almost equal between the two studies except for the following 3 SNPs: IVS22+62T>C (0.28 in Chinese vs. 0.48 in Japanese), 4002G>A (Ser1334Ser) (0.11 in Chinese vs. 0.20 in Japanese), and *866T>A (0.15 in Chinese vs. 0.26 in Japanese). These SNPs might provide population specificity within Asians.

Linkage disequilibrium (LD) analysis: Using the 43 genetic variations detected at ≥0.05 frequencies, LD analysis was performed with the r^2 and |D'| statistics, and the pairwise values for both are shown with 10-graded blue colors in **Fig. 1**.

For the r^2 values, perfect linkage was detected between IVS3-66T>C and IVS5+364_+365delAA, between IVS4-109G>C and IVS4-100C>T, among IVS10+198C>A, IVS11+122C>G and IVS11-48C>T, and between IVS21-91G>A and IVS21-89G>A. Strong linkages were observed among IVS3-288T>G, IVS3-196A>C, IVS4-252C>T, IVS7+54C>A and IVS12+56C>T ($r^2 \geq 0.65$), among 825T>C, 1062T>C, IVS9+8A>G, IVS9-189C>G and IVS9-176T>C

($r^2 \geq 0.95$), among IVS12-37A>G, 1684C>T and IVS18-30G>C ($r^2 \geq 0.93$), between IVS12-85G>A and IVS19-175C>T ($r^2 = 0.71$), among IVS16+213A>G, 2168G>A and IVS18-39_-38delAT ($r^2 \geq 0.95$), among IVS21-71A>G, IVS21-43_-42insGCTGGG and IVS22+62T>C ($r^2 \geq 0.83$), and among IVS28-266C>G, IVS29-13delT and *543C>T ($r^2 \geq 0.95$).

For the |D'| values, strong linkages (|D'| ≥0.8) were observed in 71.3% (122/171) of the pairs between 19 variations from IVS3-288T>G to IVS12+56C>T. In the region from IVS12-85G>A to IVS19-175C>T, very strong linkages were observed in |D'| values (≥0.92 in all the 45 pairs). Perfect linkages in |D'| (1.0 for all 10 pairs) were detected among the five variations from IVS21-91G>A and IVS22+62T>C. Strong linkages (≥0.91 in all the 28 pairs) were also observed among the eight variations from 4002G>A and *866T>A.

The multiallelic $(GCC)_{9\text{-}23}$ repeat was defined as Block -1 since no close linkages of these polymorphisms with other variations were detected with the PHASE program (data not shown). Based on the r^2 and |D'| values, we divided the rest of the analyzed *ABCC1* region into four LD blocks as indicated in **Fig. 1**. Block 1, spanning at least 48.9 kb, included 34 variations from IVS1-371G>A in intron 1 to IVS12+56C>T in intron 12. Block 2, which included 18 variations, ranges from IVS12-85G>A to IVS19-175C>T (34.4 kb). Block 3 spanned 25.2 kb from intron 21 (IVS21+11C>G) to intron 27 (IVS27+179G>A) with 20 variations. The very rare variation IVS21+11C>G and the SNP IVS27+179G>A were tentatively included in Block 3. Block 4 contained the remaining 12 variations from 4002G>A to *1058_*1059insT, spanning at least 7.9 kb.

Haplotype estimation and selection of htSNPs: We analyzed haplotype structures of *ABCC1* for each block and identified the haplotype-tagging SNPs (htSNPs), which is sufficient to capture frequent haplotypes in Japanese. The haplotypes for Blocks 1 to 4 and their frequencies were shown in **Tables 3 to 6**. Using all of the 34, 18, 20 and 12 variations, 32, 23, 23 and 13 haplotypes were inferred in Blocks 1, 2, 3 and 4, respectively. The diplotype configurations were obtained at probabilities over 0.9 for 95% (Block 1), 98% (Block 2), 91% (Block 3) and 100% (Block 4) of the 153 subjects. The haplotypes without amino acid change were designated as *1. Of all the estimated haplotypes, 20 in Block 1, 10 in Block 2, 7 in Block 3, and 5 in Block 4 were ambiguously inferred in only one subject. Of these ambiguous haplotypes, the *1 haplotypes were grouped into "*others*" in **Tables 3 to 6**. The haplotypes detected on more than 10 chromosomes (3% frequency) were called common haplotypes in this paper.

In Block 1 (**Table 3**), 4 haplotype groups (*1 to *4) were inferred, and the *2 to *4 groups were represented by the nonsynonymous variations, 218C>T (Thr73Ile) (*2), 726G>T (Trp242Cys) (*3), and 1199T>C (Ile400Thr) (*4). The most dominant haplotype was *1a with a 0.255 frequency, which was followed by *1b (0.206), *1c (0.150), *1d (0.101), *1e (0.049), *1f (0.042), *1g (0.039), *1h (0.036), and *1i (0.033). These 9 common haplotypes (*1a to *1i) accounted for 91% of all the inferred haplotypes. To discriminate these 9 common haplotypes, genotyping of the 8 htSNPs, IVS3-196A>C, IVS3-66T>C, IVS4-109G>C, IVS7+64C>G, 825T>C (Val275Val), IVS10-117A>G, IVS11-78G>A, and IVS11-48C>T is sufficient. In addition to these 8 htSNPs, 3 nonsynonymous variations, 218C>T (Thr73Ile), 726G>T (Trp242Cys), and 1199T>C (Ile400Thr) may be included in the htSNPs in order to detect *2 to *4 haplotypes because they might have the functional significance.

In Block 2 (**Table 4**), 4 haplotype groups (*1 to *4) were inferred. The *2 to *4 haplotypes were defined by the nonsynonymous variations, 2168G>A (Arg723Gln) (*2), 1967G>C (Ser656Thr) (*3), and 2530G>A (Gly844Ser) (*4). The most frequent haplotype was *1a (frequency: 0.288), followed by *1b (0.209), *1c (0.127), *1d (0.098), *1e (0.092), *2a (0.065) and *1f (0.033). These 7 common haplotypes accounted for 91% of all the inferred haplotypes. To distinguish these 7 haplotypes, the 6 htSNPs, IVS12-85G>A, 1684C>T (Leu562Leu), IVS13+105C>T, 2007C>T (Pro669Pro), IVS16+181C>T, and 2168G>A (Arg723Gln), can be used. In addition to them, 2 nonsynonymous variations, 1967G>C (Ser656Thr) (*3) and 2530G>A (Gly844Ser) (*4), may be added to the htSNPs for Block2.

As for Block 3 (**Table 5**), the haplotypes with 3550G>A (Glu1184Lys), 3901C>T (Arg1301Cys), 3490G>A (Val1164Ile) and 3173G>A (Arg1058Gln) were defined as *2, *3, *4 and *5, respectively. The most frequent haplotype was *1a (frequency: 0.359), followed by *1b (0.193), *1c (0.111), *1d (0.082), *1e (0.078), *1f (0.042) and *1g (0.039). These 7 common haplotypes accounted for 91% of all the haplotypes. The selected htSNPs were IVS21-89G>A, IVS22+62T>C, IVS22-43C>T, and IVS27+179G>A. In addition, the variations 3550G>A (Glu1184Lys, *2), 3901C>T (Arg1301Cys, *3), 3490G>A (Val1164Ile, *4) and 3173G>A (Arg1058Gln, *5) could be included in the Block 3 htSNPs.

Regarding Block 4 (**Table 6**), the haplotype containing the nonsynonymous variation 4502A>G (Asp1501Gly) was designated as *2. The common haplotypes were *1a (frequency: 0.310), *1b (0.278), *1c (0.190), *1d (0.085), *1e (0.059), and *1f (0.052). These 6 haplotypes accounted for 97% of the inferred haplotypes. Five htSNPs were selected: 4002G>A (Ser1334Ser), IVS28-45G>A, IVS30+18G>A,

Table 3. *ABCC1* Block 1 haplotypes

Region			Exon 2	Intron 3			Intron 4			Intron 5		Exon 7	Intron 7			Exon 8	Exon 9		Intron 9			Intron 10		Intron 11			Intron 12	Number	Frequency	
Nucleotide change[a]			218 C>T	IVS3-288 T>G	IVS3-196 A>C	IVS3 -66 T>C	IVS4-252 C>T	IVS4-109 G>C	IVS4-100 C>T	IVS5+120 A>T	IVS5 +364_365 delAA	726 G>T	IVS7 +54 C>A	IVS7 +64 C>G	IVS7 +69 C>T	825 T>C	1062 T>C	1199 T>C	IVS9 +8 A>G	IVS9-189 C>G	IVS9-176 T>C	IVS10 +198 C>A	IVS10-117 A>G	IVS11 +122 C>G	IVS11 -78 G>A	IVS11 -48 C>T	IVS12 +56 C>T			
Amino acid change			T73I									W242C				V275V	N354N	I400T												
Haplotypes[b,c]	*1	*1a																										78	0.255	0.987
		*1b																										63	0.206	
		*1c																										46	0.150	
		*1d																										31	0.101	
		*1e																										15	0.049	
		*1f																										13	0.042	
		*1g																										12	0.039	
		*1h																										11	0.036	
		*1i																										10	0.033	
		*1j																										3	0.010	
		*1k																										2	0.007	
		others[d]																										18	0.059	
	*2	*2a																										2	0.007	0.007
	*3	*3a[e]																										1	0.003	0.003
	*4	*4a[e]																										1	0.003	0.003
																												306	1.000	1.000

[a]A of the translational start codon of *ABCC1* is numbered +1. NT_010393.15 was used as the reference sequence.
[b]Major allele, white; minor allele, gray.
[c]The haplotypes are described as numbers plus small alphabetical letters.
[d]The ambiguous **1* haplotypes inferred in only one subject are grouped into "*others*", and the variations found only in these ambiguous haplotypes are not shown.
[e]The haplotype was inferred in only one subject and concurrent variations are ambiguous.

Table 4. *ABCC1* Block 2 haplotypes

Region			Intron 12		Exon 13	Intron 13	Intron 14	Exon 15	Intron 15	Exon 16		Intron 16		Exon 17	Intron 18		Exon 19	Intron 19	Number	Frequency	
Nucleotide change[a]			IVS12 -85 G>A	IVS12 -37 A>G	1684 C>T	IVS13 +105 C>T	IVS14 +115 C>T	1967 G>C	IVS15 -99 C>G	2001 C>T	2007 C>T	IVS16 +181 C>T	IVS16 +213 A>G	2168 G>A	IVS18 -39_ -38 delAT	IVS18 -30 G>C	2530 G>A	IVS19 -175 C>T			
Amino acid change					L562L			S656T		S667S	P669P			R723Q			G844S				
Haplotypes[b,c]	*1	*1a																	88	0.288	0.928
		*1b																	64	0.209	
		*1c																	39	0.127	
		*1d																	30	0.098	
		*1e																	28	0.092	
		*1f																	10	0.033	
		*1g																	7	0.023	
		*1h																	6	0.020	
		*1i																	1	0.003	
		*1j																	1	0.003	
		*1k																	1	0.003	
		*1l																	1	0.003	
		others[d]																	8	0.026	
	*2	*2a												2					20	0.065	0.065
	*3	*3a[e]						3											1	0.003	0.003
	*4	*4a[e]															4		1	0.003	0.003
																			306	1.000	1.000

[a]A of the translational start codon of *ABCC1* is numbered +1. NT_010393.15 was used as the reference sequence.
[b]Major allele, white; minor allele, gray.
[c]The haplotypes are described as numbers plus small alphabetical letters.
[d]The ambiguous **1* haplotypes inferred in only one subject are grouped into "*others*", and the variations found only in these ambiguous haplotypes are not shown.
[e]The haplotype was inferred in only one subject and concurrent variations are ambiguous.

Table 5. *ABCC1* Block 3 haplotypes

Region			Intron 21					Intron 22		Exon 23	Intron 23	Exon 24		Intron 25	Intron 26		Exon 27	Intron 27		Number	Frequency	
Nucleotide change[a]			IVS21 +11 C>G	IVS21 -91 G>A	IVS21 -89 G>A	IVS21 -71 A>G	IVS21 -43_-42ins GCTGGG	IVS22 +62 T>C	IVS22 -43 C>T	3173 G>A	IVS23 -131 G>C	3490 G>A	3550 G>A	IVS25 +114 C>T	IVS26 -30 C>G	IVS26 -24_-14 del TCTCTC CCTTC	3901 C>T	IVS27 +98 C>T	IVS27 +179 G>A			
Amino acid change										R1058Q		V1164I	E1184K				R1301C					
Haplotypes[b,c]	*1	*1a																		110	0.359	0.987
		*1b																		59	0.193	
		*1c																		34	0.111	
		*1d																		25	0.082	
		*1e																		24	0.078	
		*1f																		13	0.042	
		*1g																		12	0.039	
		*1h																		8	0.026	
		*1i																		5	0.016	
		*1j																		3	0.010	
		*1k																		1	0.003	
		*1l																		1	0.003	
		*1m																		1	0.003	
		*1n																		1	0.003	
		others[d]																		5	0.016	
	*2	*2a											2							1	0.003	0.003
	*3	*3a															3			1	0.003	0.003
	*4	*4a[e]										4								1	0.003	0.003
	*5	*5a[e]								5										1	0.003	0.003
																				306	1.000	1.000

[a]A of the translational start codon of *ABCC1* is numbered +1. NT_010393.15 was used as the reference sequence.
[b]Major allele, white; minor allele, gray.
[c]The haplotypes are described as numbers plus small alphabetical letters.
[d]The ambiguous **1* haplotypes inferred in only one subject are grouped into "*others*", and the variations found only in these ambiguous haplotypes are not shown.
[e]The haplotype was inferred in only one subject and concurrent variations are ambiguous.

Table 6. *ABCC1* Block 4 haplotypes

Region			Exon 28	Intron 28				Intron 29	Intron 30	Exon 31	3'-UTR (Exon 31)				Number	Frequency	
Nucleotide change[a]			4002 G>A	IVS28 +14 G>A	IVS28 -266 C>G	IVS28 -168 C>T	IVS28 -45 G>A	IVS29 -13 delT	IVS30 +18 G>A	4502 A>G	*543 C>T	*801 C>G	*866 T>A	*1058_ *1059 insT			
Amino acid change			S1334S							D1501G							
Haplotypes[b,c]	*1	*1a													95	0.310	0.997
		*1b													85	0.278	
		*1c													58	0.190	
		*1d													26	0.085	
		*1e													18	0.059	
		*1f													16	0.052	
		*1g													2	0.007	
		*1h													1	0.003	
		others[d]													4	0.013	
	*2	*2a[e]													1	0.003	0.003
															306	1.000	1.000

[a]A of the translational start codon of *ABCC1* is numbered + 1. NT 010393.15 was used as the reference sequence.
[b]Major allele, white; minor allele, gray.
[c]The haplotypes are described as numbers plus small alphabetical letters.
[d]The ambiguous **1* haplotypes inferred in only one subject are grouped into "*others*", and the variations found only in these ambiguous haplotypes are not shown.
[e]The haplotype was inferred in only one subject and concurrent variations are ambiguous.

*801C>G and *866T>A. The **2* marker, 4502A>G (Asp1501Gly), may also be included.

Recently, Wang *et al.* reported the haplotype structures of *ABCC1* in Chinese.[21] Although their variations used for block haplotyping were different from those used in this study, their positions for block partitioning were similar to ours. Furthermore, several differences in the haplotype frequencies were found between our Block 4 and their corresponding block (Block 3). Our Block 4**1d* and **1e* haplotypes were not shown in their study. The frequencies of our **1c* (0.190) and **1f* (0.052) were different from those of their corresponding haplotypes AAGGAT (0.093) and GAGGTT (0.130), respectively. These discrepancies partly reflect the differences in SNP frequencies of 4002G>A (Ser1334Ser) and *866T>A described above.

In conclusion, we identified 86 genetic variations including 31 novel ones in 153 Japanese subjects in *ABCC1* gene. Eight novel variations resulted in amino acid substitutions. Based on the LD profile, the analyzed region was divided into one multiallelic site and 4 blocks, and block haplotypes were inferred. We also identified the htSNPs that are sufficient to capture the common *ABCC1* haplotypes in Japanese. This is the first report on the comprehensive haplotype structures of *ABCC1* in Japanese. This information would be useful for pharmacogenetic studies to investigate the associations of the *ABCC1* haplotypes with interindividual differences of drug disposition.

Acknowledgments: We appreciate great help from Drs. Kei Fujimoto and Kanae Yasuda at Hiranuma Clinic. We also thank Ms. Chie Sudo for her secretarial assistance. The authors Hiromi Fukushima-Uesaka and Yoshiro Saito contributed equally to this work.

References

1) Leslie, E. M., Deeley, R. G. and Cole, S. P.: Toxicological relevance of the multidrug resistance protein 1, MRP1 (ABCC1) and related transporters. *Toxicology*, **167**: 3-23 (2001).
2) Leslie, E. M., Deeley, R. G. and Cole, S. P.: Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol. Appl. Pharmacol.*; **204**: 216-237 (2005).
3) Cole, S. P, Bhardwaj, G., Gerlach, J. H., Mackie, J. E., Grant, C. E., Almquist, K. C., Stewart, A. J., Kurz, E. U., Duncan, A. M. and Deeley, R. G.: Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science*, **258**: 1650-1654 (1992).
4) Grant, C. E., Kurz, E. U., Cole, S. P. and Deeley RG.: Analysis of the intron-exon organization of the human multidrug-resistance protein gene (MRP) and alternative splicing of its mRNA. *Genomics*, **45**: 368-378 (1997).
5) Flens, M. J., Zaman, G. J., van der Valk, P., Izquierdo, M. A., Schroeijers, A. B., Scheffer, G. L., van der Groep, P., de Haas, M., Meijer, C. J. and Scheper, R. J.: Tissue distribution of the multidrug resistance protein. *Am. J. Pathol.*, **148**: 1237-1247 (1996).
6) Payen, L., Delugin, L., Courtois, A., Trinquart, Y., Guillouzo, A. and Fardel, O.: The sulphonylurea glibenclamide inhibits multidrug resistance protein

(MRP1) activity in human lung cancer cells. *Br. J. Pharmacol.*, **132**: 778-784 (2001).
7) Le Saux, O., Urban, Z., Tschuch, C., Csiszar, K., Bacchelli, B., Quaglino, D., Pasquali-Ronchetti, I., Pope, F.M., Richards, A., Terry, S., Bercovitch, L., de Paepe, A. and Boyd, CD.: Mutations in a gene encoding an ABC transporter cause pseudoxanthoma elasticum. *Nat. Genet.*, **25**: 223-227 (2000).
8) Perdu, J. and Germain, D. P.: Identification of novel polymorphisms in the pM5 and MRP1 (ABCC1) genes at locus 16p13.1 and exclusion of both genes as responsible for pseudoxanthoma elasticum. *Hum. Mutat.* **17**: 74-75 (2001).
9) Conrad, S., Kauffmann, H. M., Ito, K., Deeley, R. G., Cole, S. P. and Schrenk, D.: Identification of human multidrug resistance protein 1 (MRP1) mutations and characterization of a G671V substitution. *J. Hum. Genet.* **46**: 656-663 (2001).
10) Conrad, S., Kauffmann, H. M., Ito, K., Leslie, E. M., Deeley, R. G., Schrenk, D. and Cole, S. P.: A naturally occurring mutation in MRP1 results in a selective decrease in organic anion transport and in increased doxorubicin resistance. *Pharmacogenetics* **12**: 321-330 (2000).
11) Ito, S., Ieiri, I., Tanabe, M., Suzuki, A., Higuchi, S. and Otsubo, K.: Polymorphism of the ABC transporter genes, MDR1, MRP1 and MRP2/cMOAT, in healthy Japanese subjects. *Pharmacogenetics*, **11**: 175-184 (2001).
12) Letourneau, I. J., Deeley, R. G. and Cole, S. P.: Functional characterization of non-synonymous single nucleotide polymorphisms in the gene encoding human multidrug resistance protein 1 (MRP1/ABCC1). *Pharmacogenet. Genomics*, **15**: 647-657 (2005).
13) Judson, R., Stephens, J. C. and Windemuth, A.: The predictive power of haplotypes in clinical response. *Pharmacogenomics*, **1**: 15-26 (2000).
14) Kitamura, Y., Moriguchi, M., Kaneko, H., Morisaki, H., Morisaki, T., Toyama, K. and Kamatani, N.: Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm. *Ann. Hum. Genet.*, **66**: 183-193 (2002).
15) Stephens, M., Smith, N. and Donnelly, P.: A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**: 978-989 (2001).
16) Stephens, M. and Donnelly, P.: A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**: 1162-1169 (2003).
17) Ren, X. Q., Furukawa, T., Aoki, S., Nakajima, T., Sumizawa, T., Haraguchi, M., Chen, Z. S., Kobayashi, M. and Akiyama, S.: Glutathione-dependent binding of a photoaffinity analog of agosterol A to the C-terminal half of human multidrug resistance protein. *J. Biol. Chem.*, **276**: 23197-23206 (2001).
18) Wu, P., Oleschuk, C. J., Mao, Q., Keller, B. O., Deeley, R. G. and Cole, S. P.: Analysis of human multidrug resistance protein 1 (*ABCC1*) by matrix-assisted laser desorption ionization/time of flight mass spectrometry: toward identification of leukotriene C4 binding sites. *Mol. Pharmacol.*, **68**: 1455-1465 (2005).
19) Bakos, E., Evers, R., Calenda, G., Tusnady, G. E., Szakacs, G., Varadi, A. and Sarkadi, B.: Characterization of the amino-terminal regions in the human multidrug resistance protein (MRP1). *J. Cell Sci.*, **113**: 4451-4461 (2000).
20) Haimeur, A., Conseil, G., Deeley, R. G. and Cole, S. P.: Mutations of charged amino acids in or near the transmembrane helices of the second membrane spanning domain differentially affect the substrate specificity and transport activity of the multidrug resistance protein MRP1 (*ABCC1*). *Mol. Pharmacol.*, **65**: 1375-1385 (2004).
21) Wang, H., Hao, B., Zhou, K., Chen, X., Wu, S., Zhou, G., Zhu, Y. and He, F.: Linkage disequilibrium and haplotype architecture for two ABC transporter genes (*ABCC1* and *ABCG2*) in Chinese population: implications for pharmacogenomic association studies. *Ann. Hum. Genet.*, **68**: 563-573 (2004).
22) Nicolis, E., Pasetto, M., Cigana, C., Pradal, U., Assael, B. M. and Melotti, P.: The GCC repeat length in the 5'UTR of MRP1 gene is polymorphic: a functional characterization of its relevance for cystic fibrosis. *BMC Med. Genet.*, **7**: 7 (2006).
23) Wang, Z., Sew, P. H., Ambrose, H., Ryan, S., Chong, S. S, Lee, E. J. and Lee, C. G.: Nucleotide sequence analyses of the MRP1 gene in four populations suggest negative selection on its coding region. *BMC Genomics*, **7**: 111 (2006).
24) Oselin, K., Mrozikiewicz, P. M., Gaikovitch, E., Pahkla, R. and Roots, I.: Frequency of MRP1 genetic polymorphisms and their functional significance in Caucasians: detection of a novel mutation G816A in the human MRP1 gene. *Eur. J. Clin. Pharmacol.*, **59**: 347-350 (2003).
25) Saito, S., Iida, A., Sekine, A., Miura, Y., Ogawa, C., Kawauchi, S., Higuchi, S. and Nakamura, Y.: Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABCC/MRP/CFTR). *J. Hum. Genet.*, **47**: 147-171 (2002).

Genetic Polymorphisms and Haplotypes of Major Drug Metabolizing Enzymes in East Asians and Their Comparison with Other Ethnic Populations

Y. Saito*, K. Maekawa, S. Ozawa and J. Sawada

Project Team for Pharmacogenetics, National Institute of Health Sciences, Tokyo 158-8501, Japan

Abstract: Remarkable ethnic differences in drug response are well known, and many of these can be attributed to differences in genetic backgrounds. Accumulating evidence has shown that genetic polymorphisms can cause the alteration or even loss of activity in drug metabolizing enzymes, transporters and receptors. Thus, genetic polymorphisms may be important in understanding these ethnic differences in drug response. Furthermore, haplotypes, linked combinations of genetic polymorphisms on a chromosome, have the advantage of providing more useful information on phenotype–genotype links than individual polymorphisms. In the past 6 years, mostly as a Japanese national project, we resequenced the exons and enhancer/promoter regions of more than 30 drug metabolizing enzymes, transporters and receptors using genomic DNA from 100 to 500 Japanese subjects, analyzed linkage-disequilibrium (LD), and estimated haplotype structures. Regarding *CYP2C9* and *2C19*, we found linkages between *CYP2C19*2* or **3* and *CYP2C9*1*, and between *CYP2C9*3* and *CYP2C19*1* haplotypes. Haplotype structures of *CYP2D6* are complicated by gene duplication or recombination. In contrast, the haplotype structure of *CYP3A4* was simple, but close linkages were observed with other *CYP3As*. As for *UGT1As*, the 8 first exons encoding active isoforms and common exons 2-5 were divided into 5 blocks by LD analysis, and intra- and inter-block haplotypes were estimated. Several linkages of haplotypes with functional importance were revealed, such as *UGT1A7*3 - UGT1A6*2 - UGT1A1*28* or **6*. In this review, we summarize polymorphisms and haplotype structures of these clinically important drug metabolizing enzymes in East Asians, mainly from our Japanese data, and compare them with those of other ethnicities.

INTRODUCTION

Remarkable ethnic differences in drug response are well known, and thus optimal drug dosages for prescription vary among or even within countries [Tate and Goldstein, 2004 for review]. For example, reduction of diastolic blood pressure by propranolol is more evident in Caucasians than in Africans [Cubeddu *et al.*, 1986]. Daily maintenance doses of warfarin, an anti-coagulant, are known to be different among Caucasians, Asians and Afro-Caribbeans [Blann *et al.*, 1999]. Many of the differences in drug response now can be attributed to genetic background. Development of DNA sequencing/genotyping technology and world-wide human genome projects has prompted the identification of clinically important genetic polymorphisms for diverse ethnic populations (see Grant 2005 for overview of genotyping technologies). As a result, accumulating data has shown that genetic polymorphisms specific for different ethnicities cause the alteration or even loss of activities in drug metabolizing enzymes, transporters and receptors [Evans and Relling, 1999, Chowbay *et al.*, 2005]. Thus, genetic polymorphisms are important in understanding ethnic differences in drug response. Furthermore, haplotypes, linked combinations of genetic polymorphisms on a chromosome, sometimes have the advantage of providing more useful information on phenotype–genotype links than individual polymorphisms [Judson *et al.*, 2000]. In addition, long-range haplotypes covering the gene clusters such as human Cytochrome P450 (CYP) *2Cs*, *CYP3As* and uridinediphosphoglucuronate glucuronosyltransferase (UGT) *1As* could help to elucidate the pharmacokinetics and pharmacodynamics of drugs with complicated metabolic pathways.

For the past 6 years, mostly as a Japanese national project to elucidate the genetic contribution to drug response in Japanese, we performed pharmacogenetic studies for more than 10 clinically important drugs. In these approaches, more than 30 genes encoding drug metabolizing enzymes, transporters and receptors were resequenced from genomic DNA from 100 to 500 Japanese subjects. Our studies cannot fully explain the interindividual or ethnic differences in drug response; however, identification of novel and/or known defective polymorphisms and haplotypes in Japanese suggests their involvement in such differences and highlights the necessity of ethnic-specific pharmacogenetic data.

In this review, we focus on four clinically important drug metabolizing enzyme groups: 1) CYP2C9 and CYP2C19, 2) CYP2D6, 3) CYP3A4, and 4) UGT1A1 and other UGT1As, and summarize the genetic polymorphisms and haplotype structures of these enzymes in East Asians, mainly from our Japanese data, and compare them with data from other ethnicities. Note that our sequence analysis to identify genetic polymorphisms focused on enhancer/promoter regions, exons and surrounding introns; thus, many intronic variations that were far from the exon-intron boundaries were excluded in the haplotype estimations. Haplotypes in this review are shown as a number plus alphabetical letters. The numbers are based on assignments by the Human Cytochrome P450 Allele Nomenclature Committee Home Page (http://www.cypalleles.ki.se/, as of July 11-15, 2006) or the UDP-

*Address correspondence to this author at the Project Team for Pharmacogenetics and Division of Biochemistry and Immunochemistry, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya, Tokyo 158-8501, Japan; Tel: +81-3-5717-3831; Fax: +81-3-5717-3832;
E-mail: yoshiro@nihs.go.jp

1570-1603/07 $50.00+.00

Glucuronosyltransferase (UGT) Alleles Nomenclature Home Page (http://galien.pha.ulaval.ca/alleles/alleles.html, as of July 14, 2006). Capital alphabetical letters were used when the corresponding haplotypes were already shown in the web sites listed above. The other inferred haplotypes in our study are provisionally shown with small alphabetical letters, basically depending on the frequencies, as shown in our original reports cited in each section.

CYP2C9 AND CYP2C19

Human CYP2C subfamily accounts for 20% of the total P450 content in human liver microsomes [Shimada *et al.*, 1994]. The four known human *CYP2C* genes are located in a cluster spanning 500 kb in chromosome 10q24 in the following order: *CYP2C18*, *CYP2C19*, *CYP2C9* and *CYP2C8*, each of which is estimated to be at least 50 kb. These genes have similar structures consisting of nine exons with conserved exon/intron boundaries and greater than 82% deduced amino acid homology [Goldstein and de Morais, 1994]. In this section, we focus on the genetic polymorphisms and haplotypes of *CYP2C9* and *CYP2C19*.

CYP2C9 Polymorphisms

More than 100 currently used drugs have been identified as substrates of CYP2C9, corresponding to about 10 to 20% of commonly prescribed drugs [Evans and Relling, 1999]. These include the narrow therapeutic index agents warfarin and phenytoin and other routinely prescribed compounds such as angiotensin II receptor blockers (losartan and irbesartan), sulfonylureas (tolbutamide, glibenclamide, glipizide and glimepiride), the diuretic torsemide, and various nonsteroidal anti-inflammatory agents (ibuprofen, diclofenac, piroxicam, flurbiprofen, and celecoxib) [Rettie and Jones, 2005 for review].

The possible genetic regulation of tolbutamide metabolism was first reported in 1979 [Scott and Poffenbarger, 1979]. *CYP2C9* polymorphisms had been recognized since multiple cDNA clones were isolated in the late 1980s and early 1990s. To date, more than 50 single nucleotide polymorphisms (SNPs) of *CYP2C9* including regulatory and coding SNPs have been identified (http://www.cypalleles.ki.se/cyp2c9.htm, as of July 15, 2006). Some SNPs have been reported to exhibit reduced catalytic activities compared with the wild-type by both *in vitro* functional studies and clinical pharmacokinetic/pharmacodynamic studies (Table 1).

A number of population genotyping studies also demonstrated that these SNPs were distributed with different frequencies among various ethnic populations. Two nonsynonymous SNPs, *CYP2C9*2* (430C>T, R144C) and *CYP2C9*3* (1075A>C, I359L), are found at allele frequencies of 10-15% and 5-10%, respectively, in Caucasians (American, European, Scandinavian, and Russian), Turkish, and Hispanic populations (Table 1) [Scordo *et al.*, 2001, Bravo-Villalta *et al.*, 2005, Garcia-Martin *et al.*, 2006]. In contrast, these SNPs are less prevalent in African and Asian populations. African-Americans and Ethiopians exhibit 2-4% and 1-2% allele frequencies for *CYP2C9*2* and *CYP2C9*3*, respectively [Scordo *et al.*, 2001; Bravo-Villalta *et al.*, 2005; Garcia-Martin *et al.*, 2006]. In East Asians, *CYP2C9*3* is found at 1-4% allele frequencies, while *CYP2C9*2* is hardly detected [Wang *et al.*, 1995; Nasu *et al.*, 1997; Yoon *et al.*, 2001]. In most *in vitro* studies, *CYP2C9*2* exhibited a small decrease in *V*max (0-35%) and little or no change in the *K*m for catalysis of various substrates [Lee *et al.*, 2002 for review]. The recombinant *CYP2C9*3* enzyme shows a greater *K*m and/or lower *V*max compared to wild-type for most CYP2C9 substrates although the magnitude of alterations in metabolic activity varies significantly among substrates [Takanashi *et al.*, 2000]. Both alleles, *CYP2C9*2* and *CYP2C9*3*, affect pharmacokinetics and/or the dose requirements of a number of substrates such as warfarin, phenytoin, losartan, and glimepiride [Kirchheiner and Brockmoller, 2005 for review].

Other reported alleles (*CYP2C9*4* to **24*) are mostly ethnic specific and/or relatively rare (Table 1). Due to the low frequencies of these alleles, *in vivo* elucidation of their functional significance is generally difficult. As for the defective alleles revealed by *in vitro* studies, *CYP2C9*5* (D360E) [Dickmann *et al.*, 2001; Tracy *et al.*, 2002; Yasar *et al.*, 2002a; Allabi *et al.*, 2004 and 2005; Takahashi *et al.*, 2006] and *CYP2C9*6* (K273RfsX34) with a null-activity mutation [Kidd *et al.*, 2001; Allabi *et al.*, 2004 and 2005; Takahashi *et al.*, 2006] were found only in Africans at allele frequencies around 0.017 and 0.006, respectively. *CYP2C9*11* (R335W) is present both in Africans and in Caucasians at allele frequencies around 0.01 [Higashi *et al.*, 2002; Blaisdell *et al.*, 2004; King *et al.*, 2004; Tai *et al.*, 2005; Veenstra *et al.*, 2005; Takahashi *et al.*, 2006], but is absent in Asians. Caucasians also carry two other rare defective alleles, *CYP2C9*12* (P489S) [Blaisdell *et al.*, 2004; Veenstra *et al.*, 2005] and *CYP2C9*14* (R125H) [Veenstra *et al.* 2005]. In Asians, 10 defective alleles have been identified: *CYP2C9*4* (I359T) [Ieiri *et al.*, 2000; Imai *et al.*, 2000], *CYP2C9*13* (L90P) [Si *et al.*, 2004], *CYP2C9*14* (R125H), *CYP2C9*15* (S162X), *CYP2C9*16* (T299A), *CYP2C9*18* (D397A+I359L) [Zhao *et al.*, 2004; Delozier *et al.*, 2005], *CYP2C9*25* (K118RfsX9), *CYP2C9*26* (T130R), *CYP2C9*28* (Q214L), and *CYP2C9*30* (A477T) [Maekawa *et al.*, 2006]. Especially, *CYP2C9*13* (L90P), an allele detected in a Chinese poor metabolizer (PM) of lornoxicam, has been found independently both in Chinese and Japanese at allele frequencies of 0.01 and 0.002, respectively [Si *et al.*, 2004; Maekawa *et al.*, 2006]. Guo *et al.* [2005a and 2005b] have revealed that the L90P substitution markedly decreased the intrinsic clearance of lornoxicam, tolbutamide and diclofenac *in vitro* and/or *in vivo*. Although further clinical investigation is required for these rare alleles, not only *CYP2C9*3* but also many other defective alleles described above would be at least partially responsible for highly variable interindividual and ethnic differences in the metabolism of CYP2C9 substrate drugs in Asians.

CYP2C9 Haplotypes

Recently, several groups reported comprehensive haplotype structures with high-density SNPs in *CYP2C9*, which will provide more useful information than single SNP genotyping in investigating interindividual or ethnic differences in the *in vivo* metabolic activity of CYP2C9. Veenstra *et al.* [2005] reported whole-gene high-resolution haplotype structures of *CYP2C9* in 192 European American patients administered warfarin. They determined 23 haplotypes, only 8 of

Table 1. Summary of *CYP2C9* Alleles

Allele	Nucleotide Change	Amino Acid Change	Allele Frequency			Functional Effect	Reference
			African	Caucasian	Asian		
*CYP2C9*2*	430C>T	R144C	0.02 - 0.04	0.10 - 0.15	ND	Decreased activity (*in vitro* and *in vivo*)	Lee *et al.* 2002, Schwarz 2003, Bravo-Villalta *et al.* 2005, Kirchheiner and Brockmoller 2005, Garcia-Martin *et al.* 2006
*CYP2C9*3*	1075A>C	I359L	0.01 - 0.02	0.05 - 0.10	0.01 - 0.04	Decreased activity (*in vitro* and *in vivo*)	Lee *et al.* 2002, Schwarz 2003, Bravo-Villalta *et al.* 2005, Kirchheiner and Brockmoller 2005, Garcia-Martin *et al.* 2006
*CYP2C9*4*	1076T>C	I359T	ND	ND	0.004 (1/264)	Decreased activity (*in vitro*)	Ieiri *et al.* 2000, Imai *et al.* 2000
*CYP2C9*5*	1080C>G	D360E	0.017	ND	ND	Decreased activity (*in vitro* and *in vivo*)	Dickmann *et al.* 2001, Yasar *et al.* 2002a, Tracy *et al.* 2002, Allabi *et al.* 2004 and 2005, Takahashi *et al.* 2006
*CYP2C9*6*	818delA	K273R fsX34	0.006	ND	ND	Decreased activity (*in vivo*)	Kidd *et al.* 2001, Allabi *et al.* 2004 and 2005, Takahashi *et al.* 2006
*CYP2C9*7*	55C>A	L19I	0.056 (1/18)	ND	ND	Unaltered activity (*in vitro*)	Blaisdell *et al.* 2004
*CYP2C9*8*	449G>A	R150H	0.036 (1/28)	ND	ND	Increased activity (*in vitro*)	Blaisdell *et al.* 2004, Allabi *et al.* 2004 and 2005
*CYP2C9*9*	752A>G	H251R	0.133 (4/30)	0.003	ND	Unaltered activity (*in vitro*)	Blaisdell *et al.* 2004, Allabi *et al.* 2005, Veenstra *et al.* 2005,
*CYP2C9*10*	815A>G	E272G				Unaltered activity (*in vitro*)	Blaisdell *et al.* 2004
*CYP2C9*11*	1003C>T	R335W	0.056 (1/18)	0.01	ND	Decreased activity (*in vitro* and *in vivo*)	Higashi *et al.* 2002, Blaisdell *et al.* 2004, King *et al.* 2004, Allabi *et al.* 2004 and 2005, Tai *et al.* 2005, Veenstra *et al.* 2005, Takahashi *et al.* 2006
*CYP2C9*12*	1465C>T	P489S		0.003		Decreased activity (*in vitro*)	Blaisdell *et al.* 2004, Veenstra *et al.* 2005
*CYP2C9*13*	269T>C	L90P	ND	ND	0.01	Decreased activity (*in vitro* and *in vivo*)	Si *et al.* 2004, Guo *et al.* 2005a and 2005b
*CYP2C9*14*	374G>A	R125H	ND	0.003	0.019	Decreased activity (*in vitro*)	Zhao *et al.* 2004, Veenstra *et al.* 2005, DeLozier *et al.* 2005
*CYP2C9*15*	485C>A	S162X	ND	ND	0.019	No holoprotein expression (*in vitro*)	Zhao *et al.* 2004, DeLozier *et al.* 2005
*CYP2C9*16*	895A>G	T299A	ND	ND	0.008	Decreased activity (*in vitro*)	Zhao *et al.* 2004, DeLozier *et al.* 2005
*CYP2C9*17*	1144C>T	P382S	ND	ND	0.008	Unaltered activity (*in vitro*)	Zhao *et al.* 2004, DeLozier *et al.* 2005
*CYP2C9*18*	1190A>C (+1075A>C)	D397A (+I359L)	ND	ND	0.019	No protein expression (D397A alone, *in vitro*)	Zhao *et al.* 2004, DeLozier *et al.* 2005
*CYP2C9*19*	1362G>C	Q454H	ND	ND	0.008	Unaltered activity (*in vitro*)	Zhao *et al.* 2004, DeLozier *et al.* 2005

(Table 1. Contd....)

Allele	Nucleotide Change	Amino Acid Change	Allele Frequency			Functional Effect	Reference
			African	Caucasian	Asian		
*CYP2C9*20*	208G>C	G70R	ND	ND	0.014		Zhao *et al.* 2004
*CYP2C9*21*	89C>T	P30L	ND	0.005	ND		Veenstra *et al.* 2005
*CYP2C9*22*	121A>G	N41D	ND	0.003	ND		Veenstra *et al.* 2005
*CYP2C9*23*	226G>A	V76M	ND	0.005	ND		Veenstra *et al.* 2005
*CYP2C9*24*	1060G>A	E354K	ND	0.002 (1/408)	ND		Herman *et al.* 2006
*CYP2C9*25*	353_362delAG AAATGGAA	K118R fsX9	ND	ND	0.002	No protein expression (*in vitro*)	Maekawa *et al.* 2006
*CYP2C9*26*	389C>G	T130R	ND	ND	0.002	Decreased activity (*in vitro*)	Maekawa *et al.* 2006
*CYP2C9*27*	449G>T	R150L	ND	ND	0.004	Unaltered activity (*in vitro*)	Maekawa *et al.* 2006
*CYP2C9*28*	641A>T	Q214L	ND	ND	0.002	Decreased activity (*in vitro*)	Maekawa *et al.* 2006
*CYP2C9*29*	835C>A	P279T	ND	ND	0.002	Unaltered activity (*in vitro*)	Maekawa *et al.* 2006
*CYP2C9*30*	1429G>A	A477T	ND	ND	0.002	Decreased activity (*in vitro*)	Maekawa *et al.* 2006

ND: not detected.

which occurred at frequencies greater than 5%, indicating that the overall haplotype structure of *CYP2C9* was not complex. In another study, 21 haplotypes were inferred from 92 individuals in three racial groups (Africans, Caucasians, and Asians) [Blaisdell *et al.*, 2004]. In our study, 46 haplotypes were assigned from 263 Japanese subjects, of which only 5 haplotypes with frequencies of >2% accounted for most (>87%) of the inferred haplotypes [Maekawa *et al.*, 2006], indicating that the haplotype structure of *CYP2C9* in Japanese is also simple. We determined 6 haplotype-tagging SNPs (htSNP), IVS8-109A>T (intronic variations are designated by "IVS" (intervening sequence), the intron number, and then positive numbers starting from the end of the preceding exon or negative numbers from the beginning of the proceeding exon), IVS8+147C>T, -1565C>T, IVS7+38C>T, IVS6+95A>G, and 1075A>C (I359L), which can distinguish the major haplotypes *CYP2C9*1A*, *CYP2C9*1B*, *CYP2C9*1e*, *CYP2C9*1f*, *CYP2C9*1h*, and *CYP2C9*3B*, respectively. Allele frequencies of these htSNPs exhibit interethnic differences between Japanese and other ethnicities publicized by the International HapMap Project (http://www.hapmap.org/index.html.ja, as of July 15, 2006) (Table 2).

Because HapMap data revealed substantial interethnic differences in the allele frequencies of htSNPs (Table 2), we then compared the precise haplotype frequency distribution in Japanese [Maekawa *et al.*, 2006] with those in other ethnic populations from previous reports in Caucasians [Veenstra *et al.*, 2005] and Africans [Blaisdell *et al.*, 2004]. The frequency of the wild-type haplotype *CYP2C9*1A* was higher in Japanese (haplotype frequency = 0.489; this frequency differs slightly from the allele frequency of htSNP shown in Table 2) than in Caucasians (0.281) as reported by Veenstra *et al.* [2005]. The haplotype *CYP2C9*1B*, first assigned by King *et al.* [2004], contained 6 linked noncoding SNPs, -3089G>A, -2665_-2664delTG, -1188T>C, IVS3+239C>T, IVS8+147C>T, and IVS8-109A>T, was found at comparable frequencies between Japanese (0.222) and Caucasians (0.175). Several studies on Caucasians and Asians showed that there was no association of the haplotype *CYP2C9*1B* or its promoter SNPs (-2665_-2664delTG and -1188T>C) with warfarin sensitivity [King *et al.*, 2004; Zhao *et al.*, 2004; Veenstra *et al.*, 2005] or acenocoumarol pharmacodynamics [Morin *et al.*, 2004]. The third dominant haplotype in Japanese, *CYP2C9*1e* (0.118) harboring the htSNP -1565C>T, was found at a frequency of 0.043 in Asians and at a frequency of 0.133 in African-Americans [Blaisdell *et al.*, 2004], but was absent in Caucasians [Veenstra *et al.*, 2005]. The fourth dominant haplotype in Japanese *CYP2C9*1f* (0.023), tagged by IVS7+38C>T, might be Asian-specific (0.022) [Blaisdell *et al.*, 2004]. These differences in the haplotype (*CYP2C9*1e* and *CYP2C9*1f*) between the various ethnicities might contribute to variance in CYP2C9 activity across populations. For example, East Asians require a lower maintenance dose of warfarin than Caucasians and Indians [Takahashi *et al.*, 2003; Zhao *et al.*, 2004]. In fact, Chern *et al.* [2006] reported that IVS3-65G>C, the *CYP2C9*1e*-tagging SNP linked perfectly with -1565C>T, is associated with an elevated warfarin sensitivity in Taiwan Chinese, leading to a lowered warfarin dose for patients who were heterozygous or homozygous carriers of this allele.

Table 2. Ethnic Differences in Allelic Frequencies of Haplotype-Tagging SNPs of *CYP2C9*

Haplotype-Tagging SNP in *CYP2C9*	dbSNP ID (NCBI)	Haplotype[#]	Our Study	HapMap*			
			Japanese (263 Subjects)	CEU (60 Subjects)	YRI (60 Subjects)	CHB (45 Subjects)	Japanese (45 Subjects)
IVS8-109A[§]	rs1934969	*CYP2C9*1A*	0.544		0.297[¶]	0.648	0.611
IVS8+147C>T	rs2298037	*CYP2C9*1B*	0.287	0.167	ND[¶]	0.267	0.330
-1565C>T	rs9332096	*CYP2C9*1e*	0.125	ND[¶]	0.183	0.033	0.044
IVS7+38C>T	rs17847029	*CYP2C9*1f*	0.034				
IVS6+95A>G[£]	rs9332174	*CYP2C9*1h*	0.011	0.225[¶]	0.267[¶]	0.023	ND
430C>T (R144C)	rs1799853	*CYP2C9*2*	ND	ND	ND	ND	ND
1075A>C (I359L)	rs1057910	*CYP2C9*3B*	0.030	0.058	ND		
IVS6-32T>C	rs9332197	-	ND	0.067[¶]	ND	ND	ND

ND: not detected.

[#]*CYP2C9* haplotypes in a Japanese population are defined by Maekawa *et al.* [2006].

*http://www.hapmap.org/index.html.ja (as of July 15, 2006). CEU, YRI and CHB are U.S. (residents with ancestry from Northern and Western Europe), Nigeria (Yoruba) and Chinese populations, respectively.

[¶]Significant differences ($P<0.01$, chi-square test) in allele frequencies between our Japanese population and each ethnic population. The multiple comparison was corrected by Bonferroni's method.

[§]The major allele, IVS8-109A, tags *CYP2C9*1A* (minor allele is IVS8-109T).

[£]In the previous papar [Maekawa *et al.*, 2006], we chose IVS2+73T>C as a htSNP of *CYP2C9*1h*, which was perfectly linked with IVS6+95A>G.

Further clinical studies are needed to evaluate the functional relevance of these Asian- (and/or African-) specific haplotypes, *CYP2C9*1e* and *CYP2C9*1f*, to the metabolism of CYP2C9 substrates. The minor Japanese haplotype, *CYP2C9*1h* (0.008) tagged by IVS6+95A>G, seems more frequent in Caucasians (0.205) and African-Americans (0.100) than in Asians (0.043) as reported by Blaisdell *et al.* [2004]. The frequency of the haplotype *CYP2C9*3B* harboring I359L in Japanese (0.027) was comparable to that in Asians (0.022) [Blaisdell *et al.* 2004], but was slightly lower than those in Caucasians (0.057-0.081) [King *et al.*, 2004; Morin *et al.*, 2004; Veenstra *et al.*, 2005].

A previous study in a Japanese population demonstrated that haplotypes harboring the promoter SNPs of *CYP2C9* (-1911T>C, -1885C>G, -1537G>A and -981G>A) resulted in a reduction of promoter activity [Shintani *et al.*, 2001]. However, the majority of the promoter SNPs are shown to be closely linked with *CYP2C9*2* (-1096A>G, -620G>T, -485T>A, -484C>A and R144C) and *CYP2C9*3* (-1911T>C, -1885C>G, -1537G>A, -981G>A and I359L) [Blaisdell *et al.*, 2004; King *et al.*, 2004; Veenstra *et al.*, 2005; Maekawa *et al.*, 2006]. It remains unclear whether these promoter SNPs contribute to the impaired activities of *CYP2C9*2* and *CYP2C9*3*.

CYP2C19 Polymorphisms

Another member of the human CYP2C subfamily, CYP2C19, accounts for only 1% of the total P450 in human liver microsomes [Inoue *et al.*, 1997]. However, it is responsible for the metabolism of clinically important drugs such as the anticonvulsant mephenytoin, proton-pump inhibitors (omeprazole, lansoprazole, rabeprazole and pantoprazole), the antimalarial proguanil and the anxiolytic diazepam [Desta *et al.*, 2002 for review]. CYP2C19 substrates are either neutral or weakly basic compounds, while CYP2C9 substrates are relatively acidic. In addition, CYP2C19 and CYP2C9 share a number of common substrates, but display different substrate stereospecificity and regioselectivity [Bajpai *et al.*, 1996; Lewis *et al.*, 1998].

Interindividual differences in the activity of CYP2C19 were first characterized by 4'-hydroxylation of *S*-mephenytoin [Andersson *et al.*, 1990]. The phenotypes of this enzyme are classified into two groups, extensive metabolizers (EMs) and PMs. The two genetic defects, *CYP2C19*2* (681G>A, splice defect) and *CYP2C19*3* (636G>A, W212X) are primarily responsible for the PM phenotype of mephenytoin [De Morais *et al.*, 1994a, 1994b]. The pharmacokinetics and/or pharmacodynamics of other CYP2C19 substrate drugs such as proton-pump inhibitors [Furuta *et al.*, 2005], diazepam [Inomata *et al.*, 2005] and antidepressants [Kirchheiner *et al.*, 2004] are also affected by *CYP2C19* genotypes. As shown in Table 3, the allele frequencies of *CYP2C19*2* (21-45%) and *CYP2C19*3* (5-13%) in Asian populations were higher than European-American populations (*CYP2C19*2*, 13-19%; *CYP2C19*3*, 0-0.3%) and Africans (*CYP2C19*2*, 11-25%; *CYP2C19*3*, 0-1.8%) [Bravo-Villalta *et al.*, 2005], resulting in significant interethnic differences in PM frequencies.

Both *CYP2C19*2* and **3* account for >99% of PM alleles in Asians and ~87% of Caucasian PM alleles [De Morais *et al.*, 1994a and 1994b]. Unequal distributions of these alleles among various ethnic groups are the primary cause of different population pharmacokinetics of CYP2C19 substrate

Table 3. Summary of *CYP2C19* Alleles

Allele	Nucleotide Change	Amino Acid Change or Effect	Allele Frequency			Functional Effect	Reference
			African	Caucasian	Asian		
*CYP2C19*2*	681G>A	Splice defect	0.11 - 0.25	0.13 - 0.19	0.21 - 0.45	No activity (*in vitro* and *in vivo*)	De Morais *et al.*, 1994a and 1994b, Desta *et al.* 2002, Kirchheiner *et al.* 2004, Furuta *et al.* 2005, Inomata *et al.* 2005
*CYP2C19*3*	636G>A	W212X	0 - 0.018	0 - 0.003	0.05 - 0.13	No activity (*in vitro* and *in vivo*)	De Morais *et al*, 1994a and 1994b, Desta *et al.* 2002, Kirchheiner *et al.* 2004, Furuta *et al.* 2005, Inomata *et al.* 2005
*CYP2C19*4*	1A>G	No translation	ND	0.006	0.004	No activity (*in vitro*)	Ferguson *et al.* 1998, Garcia-Barcelo *et al.* 1999
*CYP2C19*5*	1297C>T	R433W	ND	ND	0.0025		Xiao *et al.* 1997
*CYP2C19*6*	395G>A	R132Q	ND	0.003 (1/346)	ND	No activity (*in vitro*)	Ibeanu *et al.* 1998
*CYP2C19*7*	IVS5+2T>A	Splice defect	ND	0.002 (1/650)	ND		Ibeanu *et al.* 1999
*CYP2C19*8*	358T>C	W120R	ND	0.003	ND	Decreased activity (*in vitro*)	Ibeanu *et al.* 1999
*CYP2C19*9*	431G>A	R144H	0.17	ND	ND	Decreased activity (*in vitro*)	Blaisdell *et al.* 2002
*CYP2C19*10*	680C>T	P227L	0.03	ND	ND	Decreased activity (*in vitro*)	Blaisdell *et al.* 2002
*CYP2C19*11*	449G>A	R150H	ND	0.03	ND	Unaltered activity (*in vitro*)	Blaisdell *et al.* 2002
*CYP2C19*12*	1473A>C	X491C	0.03	ND	ND	No holoprotein (*in vitro*)	Blaisdell *et al.* 2002
*CYP2C19*13*	1228C>T	R410C	0.06	ND	ND	Unaltered activity (*in vitro*)	Blaisdell *et al.* 2002
*CYP2C19*14*	50T>C	L17P	0.06	ND	ND		Blaisdell *et al.* 2002
*CYP2C19*15*	55A>C	I19L	0.06	ND	ND		Blaisdell *et al.* 2002
*CYP2C19*16*	1324C>T	R442C	ND	ND	rare		Morita *et al.* 2004
*CYP2C19*17*	-3402C>T, -806C>T	Increased transcription	0.18	0.18	0.04	Increased activity (*in vitro* and *in vivo*)	Sim *et al.* 2006
*CYP2C19*18*	986G>A	R329H	ND	ND	0.002		Fukushima-Uesaka *et al.* 2005
*CYP2C19*19*	151A>G	S51G	ND	ND	0.002		Fukushima-Uesaka *et al.* 2005

ND: not detected.

drugs [Desta *et al.*, 2002 for review]. As summarized in Table 3, however, subsequent studies have revealed additional defective *CYP2C19* alleles. A null allele, *CYP2C19*4* (1A>G), was found in Caucasians and Chinese with 0.6% and 0.4% frequencies, respectively [Ferguson *et al.* 1998, Garcia-Barcelo *et al.*, 1999]. *CYP2C19*5* (1297C>T, R433W), located in the conserved heme-binding region, was found in one Chinese Bai subject who was a PM of mephenytoin

[Xiao *et al.*, 1997]. *CYP2C19*6* to **15* were found in Caucasians or Africans, but not in Asians [Ibeanu *et al.*, 1998 and 1999; Blaisdell *et al.*, 2002]. *CYP2C19*16* (1324C>T, R442C) located near the heme-binding region, was found in a Japanese subject with impaired mephobarbital 4'-hydroxylation activity [Morita *et al.*, 2004]. *CYP2C19*17* harboring -806C>T and -3402C>T in the 5'-flanking region was identified with frequencies of 0.18 in both Swedes and Ethiopians and 0.04 in Chinese [Sim *et al.*, 2006]. The **17* carriers had increased *in vivo* omeprazole metabolism, probably due to the mutated -806T site, which consistently increased the transcription of *CYP2C19* by luciferase reporter transfection experiments *in vivo* in mice. Recently, we identified 2 novel alleles, *CYP2C19*18* (986G>A, R329H) and *CYP2C19*19* (151A>G, S51G) in a Japanese population [Fukushima-Uesaka *et al.*, 2005], and their functional analysis is ongoing.

CYP2C19 Haplotypes

Although *CYP2C19*2* and *CYP2C19*3* polymorphisms were extensively studied in relation to the pharmacokinetics/pharmacodynamics of CYP2C19 substrate drugs, pharmacogenetic studies using haplotypes of *CYP2C19* in various ethnic groups are currently lacking. Recently, we performed a comprehensive haplotype analysis using 48 genetic variations obtained from 253 Japanese subjects, and inferred 31 haplotypes in *CYP2C19*, of which only 5 haplotypes (haplotype frequency in parentheses) had frequencies of >2%: *CYP2C19*1d* (0.492), *CYP2C19*2c* (0.241), *CYP2C19*3b* (0.115), *CYP2C9*1e* (0.043), and *CYP2C19*1f* (0.022) accounted for most (>91%) of the observed haplotypes [Fukushima-Uesaka *et al.* 2005]. The htSNPs that resolved the 6 common haplotypes were IVS7-106T>C (*CYP2C19*1d*), 681G>A (*CYP2C19*2c*), 636G>A (*CYP2C19*3b*), 991A>G (*CYP2C9*1e*), IVS7-201G>A (*CYP2C19*1f*) and -806C>T (*CYP2C19*17a*, originally designated *CYP2C19*1j* in Fukushima-Uesaka *et al.* [2005]). We compared the allele frequencies of these 6 htSNPs in Japanese with those of the International HapMap Project (http://www.hapmap.org/index.html.ja, as of July 15, 2006) (Table 4) although caution should be taken that Nigerian (Yoruba) may not necessarily represent Africans. The allele frequency of IVS7-106T>C tagging haplotype *CYP2C19*1d* in Japanese (0.530) was comparable to that of Caucasians (0.508), but was quite higher than that of Nigerians (0.183). 681G>A (splicing defect), the htSNP of *CYP2C19*2c* in Japanese, was found at an allele frequency of 0.267, which was comparable to that in Chinese (0.256) in the HapMap Project, but was slightly higher than those in Caucasians (0.150) and Nigerians (0.167). In agreement with previous reports [Bravo-Villalta *et al.*, 2005], 636G>A (W212X) tagging *CYP2C19*3b* was not found in Caucasians and Nigerians. The allele frequency of 991A>G (I331V), the htSNP of *CYP2C19*1e*, was comparable between Japanese and Caucasians. Marked differences in allele frequencies of -806C>T tagging the *CYP2C19*17a* haplotype were observed among East Asians, Nigerians, and Caucasians. Its frequency was about twenty times higher in Caucasians (0.217) and in Nigerians (0.275) than in Japanese (0.008). As described above, Sim *et al.* [2006] reported that -806C>T together with –3402C>T (*CYP2C19*17*) showed interracial differences in allelic frequency among Swedes, Ethiopians and Chinese, and was associated with the ultra-EM phenotype for omeprazole due to augmented expression of CYP2C19. They predicted that the omeprazole AUC (area under the plasma concentration-time curve) in subjects homozygous for *CYP2C19*17* would be 60% of that of subjects homozygous for *CYP2C19*1*. Thus, it is possible that *CYP2C19*17* (-806C>T, -3402C>T) and its representative haplotype *CYP2C19*17a* in Japanese cause therapeutic failures in treatment with proton-pump inhibitors and antidepressants. Further studies on comprehensive haplotype structures in *CYP2C19* of major ethnic groups and their associations with the metabolism of CYP2C19 substrate drugs are necessary.

Table 4. Ethnic Differences in Allelic Frequencies of Haplotype-tagging SNPs of *CYP2C19*

Haplotype-Tagging SNP in *CYP2C19*	dbSNP ID (NCBI)	Haplotype[#]	Our Study	HapMap*			
			Japanese (253 Subjects)	CEU (60 Subjects)	YRI (60 Subjects)	CHB (45 Subjects)	Japanese (45 Subjects)
IVS7-106T>C	rs4917623	*CYP2C19*1d*	0.530	0.508	0.183[¶¶]	0.602	0.593
681G>A (splicing defect)	rs4244285	*CYP2C19*2c*	0.267	0.150[¶]	0.167	0.256	0.284
636G>A (W212X)	rs4986893	*CYP2C19*3b*	0.128	ND[¶¶]	ND[¶¶]	0.033	0.045
991A (I331)[‡]	rs3758581	*CYP2C19*1e*	0.045	0.058	ND	0.056	0.091
IVS7-201G>A	rs17882222	*CYP2C19*1f*	0.024				
-806C>T	rs12248560	*CYP2C19*17a*	0.008	0.217[¶¶]	0.275[¶¶]	0.022	ND

ND: not detected.

[#]*CYP2C19* haplotypes in a Japanese population are defined by Fukushima-Uesaka *et al* [2005].

*http://www.hapmap.org/index.html.ja (as of July 15, 2006). CEU, YRI and CHB are U.S. (residents with ancestry from Northern and Western Europe), Nigeria (Yoruba) and Chinese populations, respectively. Significant differences ([¶]$P<0.05$, [¶¶]$P<0.01$, chi-square test) in allele frequencies between our Japanese population and each ethnic population. The multiple comparison was corrected by Bonferroni's method.

[‡]The minor allele, 991A (I331), tags *CYP2C19*1e* [(major allele is 991G (V331)].

LDs and Haplotype Structures of the CYP2C Cluster

It has recently become evident that alleles or haplotypes in the *CYP2C* subfamily gene (*CYP2C18*, *CYP2C19*, *CYP2C9* and *CYP2C8*) are closely linked with each other. By genotyping 1468 subjects in Stockholm, Yasar *et al.* [2002b] showed a strong linkage of *CYP2C9*2* with *CYP2C8*3* harboring two SNPs, 416G>A (R139K) and 1196A>G (K399R). In their study, approximately 96% of the subjects with the *CYP2C8*3* alleles also carried *CYP2C9*2*, and 85% of the subjects that had *CYP2C9*2* also carried *CYP2C8*3*. A similar linkage has been reported between *CYP2C18* and *CYP2C19* variations [Mamiya *et al.* 1998]. A coding region polymorphism in *CYP2C18*, which generates a premature stop codon (204T>A, Y68X), was completely linked to the *CYP2C19*3* allele in a Japanese population, suggesting that individuals who lack CYP2C19 activity also lack CYP2C18 activity. In addition, an upstream *CYP2C18* polymorphism (-478T>C) was in complete linkage with the *CYP2C19*2* allele although the effect of this upstream polymorphism on gene expression is currently unknown.

The LD profiles of SNPs in the polygenic *CYP2C* region from two population samples (European and Japanese) indicated that the four *CYP2C* genes are possibly divided into two LD blocks (clusters): *CYP2C18* and *CYP2C19* in cluster 1 and *CYP2C9* and *CYP2C8* in cluster 2 [Ahmadi *et al.*, 2005]. Analysis using HapMap data from Europeans, Yoruba, Chinese, and Japanese suggested that a more extensive LD block is observed in *CYP2C* across populations: *CYP2C* cluster 1 spans *CYP2C18* and *CYP2C19* and also includes the exonic part of *CYP2C9*, and *CYP2C* cluster 2 includes *CYP2C8* and a small part of the *CYP2C9* 3'-flanking region [Walton *et al.*, 2005]. We analyzed LD patterns for 253 Japanese subjects and revealed the associations of haplotypes between *CYP2C9* and *CYP2C19*. As shown in Fig. (1), of all 1225 pairwise |D'| values between 50 common SNPs consisting of 24 in *CYP2C19* [Fukushima-Uesaka *et al.*, 2005] and 26 in *CYP2C9* (> 0.01 in their allele frequencies) [Maekawa *et al.*, 2006], 988 pairs (81%) had |D'|>0.90, indicating an extended LD block covering both *CYP2C19* and *CYP2C9*. The long-range haplotypes spanning *CYP2C19* and *CYP2C9* were inferred using 12 htSNPs (Fig. 2). The most dominant haplotype, *H1* (0.524 frequency), is the combination of the wild-type haplotypes of both *CYP2C19* (*CYP2C19*1d*) and *CYP2C9* (*CYP2C9*1A*) in Japanese that are associated with extensive metabolic phenotypes. The defective allele of *CYP2C19*, *CYP2C19*2* (681G>A, splicing defect), was assigned to either *H2* or *H4* with a frequency of

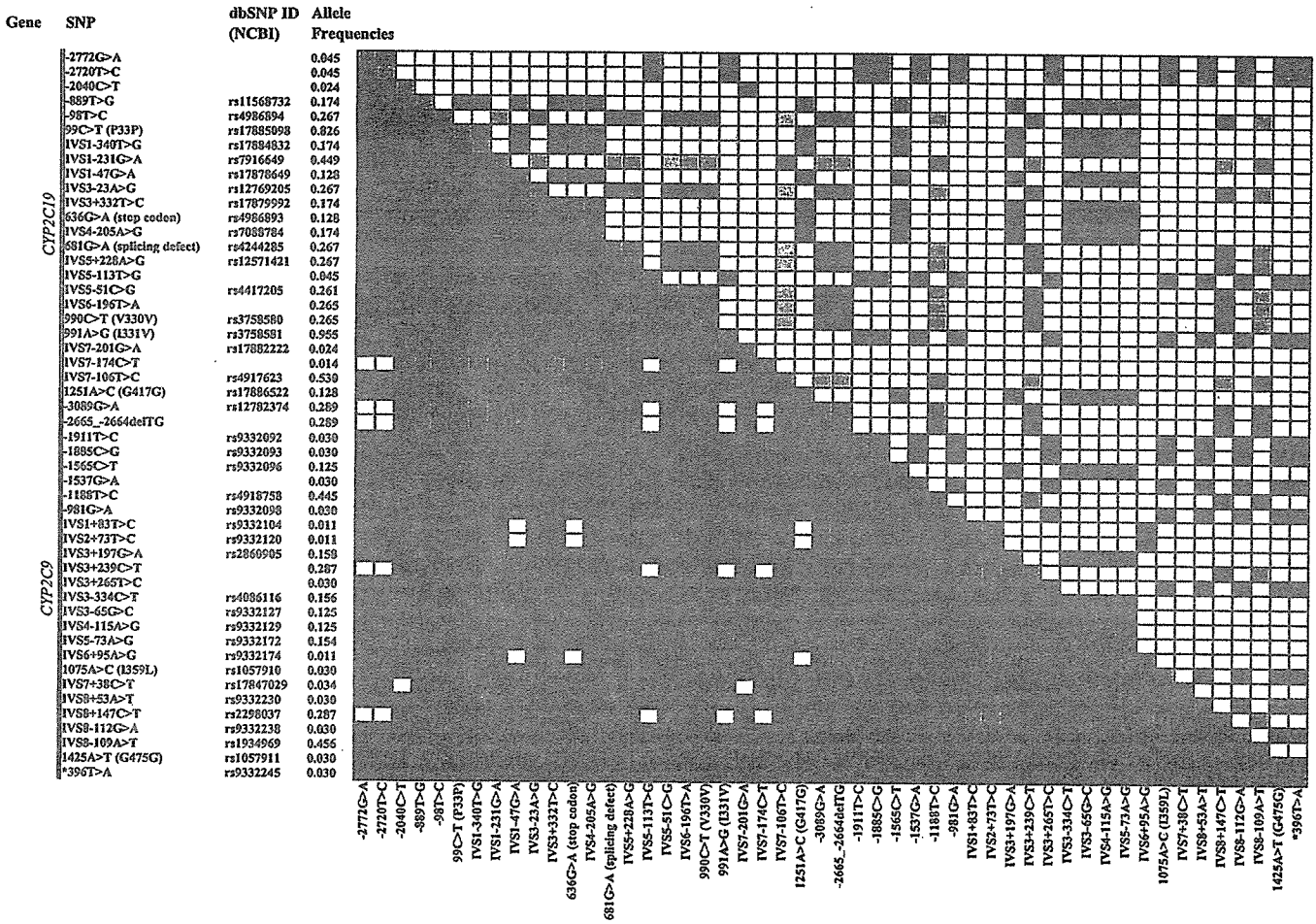

Fig. (1). Linkage disequilibrium (LD) analysis of *CYP2C19* and *CYP2C9* in a Japanese population (253 subjects). Pairwise LD between 50 common variations is expressed as |D'| (lower left) and r^2 (upper right) by 10-graded blue colors. The denser color indicates higher linkage. Allele frequencies of variations in Japanese are also shown.

Gene		*CYP2C19*						*CYP2C9*						Frequency	Haplotype of each gene	
Nucleotide change[#]		-806 C>T	636 G>A	681 G>A	991 A>G	IVS7 -201 G>A	IVS7 -106 T>C	-1565 C>T	IVS6 +95 A>G	1075 A>C	IVS7 +38 C>T	IVS8 +147 C>T	IVS8 -109 A>T			
Amino acid change or effect			W212X	splicing defect	I331V					I359L					*CYP2C19* haplotype[§]	*CYP2C9* haplotype[£]
Allele		*CYP2C 19*17*	*CYP2C 19*3*	*CYP2C 19*2*						*CYP2C 9*3*						
Combinatorial haplotype*	*H1*													0.524	*CYP2C19*1d*	*CYP2C9*1A*
	H2													0.231	*CYP2C19*2c*	*CYP2C9*1B*
	H3													0.123	*CYP2C19*3b*	*CYP2C9*1e*
	H4													0.034	*CYP2C19*2c*	*CYP2C9*1f*
	H5													0.030	*CYP2C19*1e*	*CYP2C9*3B*
	H6													0.024	*CYP2C19*1f*	*CYP2C9*1A*
	H7													0.016	*CYP2C19*1e*	*CYP2C9*1B*
	H8													0.008	*CYP2C19*17a*	*CYP2C9*1h*
	H9													0.005	*CYP2C19*3b*	*CYP2C9*1B*
	Others													0.007		

Fig. (2). Long-range haplotypes spanning *CYP2C19* and *CYP2C9* in a Japanese population (253 subjects). [#]A of the translational start codon of *CYP2C19* or *CYP2C9* is numbered +1. NT_030059.12 was used as the reference sequence. [*]Major allele, white; minor allele, gray. [§]Refer to Fukushima-Uesaka *et al.* [2005] for detailed *CYP2C19* haplotypes. [£]Refer to Maekawa *et al.* [2006] for detailed *CYP2C9* haplotypes.

0.231 and 0.034, respectively. Two haplotypes, *H3* and *H9* with frequencies of 0.123 and 0.005, respectively, contained another defective allele of *CYP2C19*, *CYP2C19*3* (636G>A, W212X). *CYP2C9*3* (1075A>C, I359L) was assigned to *H5* with a frequency of 0.030. There is no linkage among *CYP2C19*2*, *CYP2C19*3*, and *CYP2C9*3*, suggesting that statistically, Japanese individuals are unlikely to show PM phenotypes simultaneously for both *CYP2C19* and *CYP2C9*. However, the diplotype configurations showed that about 67% of Japanese individuals bear one or two copies of haplotypes harboring either *CYP2C19*2*, *CYP2C19*3* or *CYP2C9*3* (*H2*, *H3*, *H4*, *H5*, *H9*) (data not shown). *CYP2 C19*17* associated with the increased transcriptional activity [Sim *et al.*, 2006] and *CYP2C9*1h* were linked mutually and formed *H8* with frequencies of 0.008. This linkage might be conserved across populations because allele frequencies of both -806C>T in *CYP2C19* and IVS6+95A>G in *CYP2C9*, tagging *CYP2C19*17a* and *CYP2C9*1h*, respectively, was significantly different between Japanese (probably Asians) and the other ethnicities described above, but parallel within a population (Tables 2 and 4). Some CYP2C9 substrate drugs are also metabolized by CYP2C19 (phenytoin, tolbutamide, and chlorpropamide) or by CYP2C8 (troglitazone, pioglitazone, and rosiglitazone). The evaluation of LD profiles and long-range haplotype structures in the *CYP2C* gene region including *CYP2C18*, *CYP2C19*, *CYP2C9*, and *CYP2 C8* will facilitate pharmacogenetic studies aimed at detecting phenotypic differences of drugs with dual (complicated) metabolic pathways mediated by at least two enzymes.

CYP2D6

Cytochrome P450 (CYP) 2D6 metabolizes a number of clinically important drugs such as anti-arrythmics, psychiatrics, anti-histamines, and anti-depressants as well as endogenous substances [Ingelman-Sundberg, 2005 for review]. As for the major defective alleles **4* (1846G>A, splicing defect) and **5* (gene deletion), the frequency of **4* is relatively high in Caucasians but very low in the Chinese and Japanese [Ingelman-Sundberg, 2005; Bradford *et al.*, 2002]. Instead, the **10* allele, which confers a partially reduced enzymatic activity, is found at much higher allele frequencies in Japanese (38 to 43%) [Dahl *et al.*, 1995; Tateishi *et al.*, 1999; Nishida *et al.*, 2000; Kubota *et al.*, 2000], Chinese (40 to 50%) [Wang *et al.*, 1993; Johansson *et al.*, 1994; Dahl *et al.*, 1995; Droll *et al.*, 1998] and Koreans (35 to 50%) [Dahl *et al.*, 1995; Roh *et al.*, 1996; Yoon *et al.*, 2000] than in Caucasians (1 to 3%) [Sachse *et al.*, 1997; Droll *et al.*, 1998; Griese *et al.*, 1998].

A number of other *CYP2D6* variant alleles have been reported (http://www.cypalleles.ki.se/cyp2d6.htm, as of July 15, 2006, SNP positions were shown following this web site). Among them, relatively frequent alleles found in Caucasians and/or Africans are **2*, **3*, **6*, **9*, **17* [See Bradford 2002 for ethnic distributions], **29* and **41*. The **2* allele (R296C and S486T) is thought to be the second wild-type but may have slightly altered substrate specificity [Tsuzuki *et al.*, 2001; Marcicci *et al.*, 2002]. The **3* allele (2549delA, frame-shift) [Kagimoto *et al.*, 1990] found in Caucasians is rare in Africans. The **6* (1707delT, frame-shift) [Saxena *et al.*, 1994] is found in Caucasians and American Indians. The **9* allele (K281del) [Tyndale *et al.*, 1991; Broly and Meyer, 1993] is found in Caucasians and Malays [Teh *et al.*, 2001]. The **17* allele (T107I, R296C, and S486T), which has changed substrate specificity [Masimirembwa *et al.*, 1996; Wennerholm *et al.*, 2002], and **29* (V136M, R296C, V338M and S486T) [Marez *et al.*, 1997; Wennerholm *et al.*, 2001] are commonly found in black Africans. Except for **2* and **41*, these alleles were hardly found in East Asians.

Genetic Polymorphisms of CYP2D6 Found in East Asians

In addition to **2* and **10*, **41* is relatively frequently found in Japanese [Ikenaga *et al.*, 2005; our unpublished data] and Koreans [Lee *et al.*, 2006a] at allele frequencies around 0.02. This allele is a low-activity **2* variant with -1584C and intronic 2988G>A [Raimundo *et al.*, 2004; Toscano *et al.*, 2006], conferring the intermediate metabolizer phenotype to Caucasians [Raimundo *et al.*, 2000; Zanger *et al.*, 2001] and Mexicans [Luo *et al.*, 2005]. In

Japanese, -1584C and 2988A are perfectly linked to each other [our unpublished data].

The **2*-group minor alleles **14* (G169R, R296C and S486T) (Ji *et al.*, 2002a) and **21* (2573_2574insC, frameshift) [Chida *et al.*, 1999a; Yamazaki *et al.*, 2003] are found at frequencis of 0.001 to 0.02 [Nishida *et al.*, 2000; Soyama *et al.*, 2004; Ikenaga *et al.*, 2005; Ebisawa *et al.*, 2005; Ji *et al.*, 2002b; Lee *et al.*, 2006a]. The **1*-group **18* allele (468_470dupVPT) [Yokoi *et al.*, 1996] identified in a Japanese poor metabolizer has been found in Japanese at frequencies of 0.002 to 0.007 [Yokoi *et al.*, 1996; Chida *et al.*, 1999b; Soyama *et al.*, 2004] but not in Chinese [Garcia-Barcelo *et al.*, 2000a] or Koreans [Lee *et al.*, 2006a]. As described above, the **4* allele is rare (at allele frequencies of 0.002 to 0.008), and **3* is hardly found in East Asians [Wang *et al.*, 1993; Pang *et al.*, 1998; Garcia-Barcelo *et al.*, 2000a; Kubota *et al.*, 2000; Nishida *et al.*, 2000; Soyama *et al.*, 2004; Ebisawa *et al.*, 2005; Lee *et al.*, 2006a].

Our comprehensive resequencing of the gene in 263 Japanese subjects [Soyama *et al.*, 2002; Soyama *et al.*, 2004; our unpublished data] detected *CYP2D6*1A*, **2A*, and **10B*, which are known to exist with high frequencies in the Japanese, their known (**14*, **18*, **21*, **41*, and **44*) and novel (**47* to **51*) variant alleles and a number of intronic variations [Soyama *et al.*, 2002; Soyama *et al.*, 2004]. Ebisawa *et al.* [2005] have reported additional novel alleles **53* to **55* as well as **27* and **39* from a study with 286 Japanese subjects. Lee *et al.* [2006a] also resequenced the *CYP2D6* gene in 400 Koreans and detected the minor alleles **14*, **21*, **27*, **35*, **39*, and **47*.

Prevalence of *36-*10B in Japanese

The **10* allele was first reported as a single nucleotide polymorphism 100C>T (P34S) in exon 1 in a Japanese population [Yokota *et al.*, 1993]. Johansson *et al.* [1994] found two low-activity *CYP2D6* genes, $CYP2D6Ch_1$ (**10B*) and $CYP2D6Ch_2$ (**36*), in Chinese subjects who were intermediate metabolizers. These genes were tandemly organized downstream of *CYP2D7P* in the following order: *CYP2D8P-CYP2D7P-*$CYP2D6Ch_2$ (**36*)-$CYP2D6Ch_1$ (**10B*). This genomic organization confers the *Xba*I 44-kb haplotype. In addition, a single-type (*Xba*I 29-kb) **10B*, *CYP2D8P-CYP2D7P-*$CYP2D6Ch_1$ (**10B*), was also found. $CYP2D6*Ch_2$, originally designated **10C* and renamed **36*, is thought to be generated by recombination with the pseudogene *CYP2D7P* at a site upstream of exon 9, resulting in 13 nucleotide changes with six amino acid substitutions.

Although the tandem form of **36-*10B* was assumed to be a major form [Johansson *et al.*, 1994; Garcia-Barcelo *et al.*, 2000b; Nishida *et al.*, 2000], no detailed information has been published for its intervening and flanking regions. We first confirmed the presence of the tandem-type **36-*10B* utilizing long-range PCR with an intron 6-specific forward primer and an intron 2-specific reverse primer and then resequenced both genes. Our sequence data have shown that most (83%) of the **10*-positive haplotypes harbor the upstream **36* gene [Soyama *et al.*, 2006a]. Frequencies of the single-type **10B* and **36-*10B* were 0.055 and 0.278 [our unpublished data], respectively.

Since the regions between *CYP2D7P* and **36* and between **36* and **10B* have not been sequenced yet, the complete sequence of the entire **36-*10B* region was also obtained [GenBank DQ211353]. Our sequence data indicated the structure of *CYP2D6*36*-REP7-*CYP2D6*10*, and the downstream **10* was confirmed to be **10B* (or its variants). Moreover, the single-type **10B* was shown to have the structure of *CYP2D7P*-REP7-*CYP2D6*10B*-REP6, and the distance between the 3'-end of **10* and CYP-REP6 was 1.6-kb shorter than that between the 3'-end of **36* and CYP-REP7 [Soyama *et al.*, 2006a].

Gene Duplication in East Asians

The other type found in the Chinese by Johansson *et al.* [1994] was the *Xba*I 42-kb duplicated genes, which had the structure of *CYP2D8P-CYP2D7P-CYP2D6L2-CYP2D6L1* (*CYP2D6*2X2*). Several research groups have investigated duplicated *CYP2D6* genes in Asians and have found *CYP2D6*1X2*, *CYP2D6*10X2* [Roh *et al.*, 1996; Garcia-Barcelo *et al.*, 2000b; Nishida *et al.*, 2000; Ishiguro *et al.*, 2003; Mitsunaga *et al.*, 2002; Soyama *et al.*, 2006a; Lee *et al.*, 2006a], and *CYP2D6*36X2* [Chida *et al.*, 2002; Gaedigk *et al.*, 2006]. The allele frequencies were low (mostly around 0.005), and their detailed structures and functional relevance in Asian populations remains mostly unclear. Ishiguro et al. [2004a] have reported that **1X2/*1* and **2X2/*1* subjects show an ultrarapid metabolizer phenotype for dextromethorphan *O*-demethylation, but that **10X2* does not show a gene-dose effect.

Novel CYP2D6 Haplotypes Containing Chimeric REP7/6

Recently, a novel **10*-related haplotype, named *CYP2D6*10D* (**10D*) [Ishiguro *et al.*, 2004b], was found with a frequency of approximately 0.003 in Japanese [Fukuda *et al.*, 2005]. The **10D* haplotype harbors a downstream *CYP2D7*-derived region and a chimeric repetitive sequence, CYP-REP7/6 (REP7/6). REP7/6 structures are also present in the deletion haplotype **5* [Steen *et al.*, 1995] and have been often utilized for **5*-typing [Hersberger *et al.*, 2000]. Thus, for Japanese and probably Chinese and Koreans, the typing of REP7/6 might have caused misplacement of **10D* as **5* [Ishiguro *et al.*, 2004b, Lee *et al.*, 2006a]. In addition to the single-type **10D*, we found an additional **10D*-bearing haplotype, **36-*10D*, at a frequency of 0.004. Moreover, a novel defective structure consisting of *CYP2D6*36* followed by 3'-flanking REP7/6 (single-type **36*-REP7/6) was also found in a Japanese population at a frequency of 0.004 [Soyama *et al.*, 2006b]. Gaedigk *et al.* [2006] have also found a single-type **36* in an Asian subject as well as in 9 African-Americans. The haplotype structures that we have found in Japanese are shown in Fig. (3).

Then, the REP7/6 sequences in **5*, **10D*, **36-*10D*, and **36*-REP7/6 were determined and classified into 5 types: types A to D for **5*, type E for **10D* and **36-*10D*, and type F for **36* [Soyama *et al.*, 2006b]. Comparisons of the sequences revealed that types A, C, and D were derived from the **1* sequence, and type B from the **2* sequence, and type E from the **10* sequence. These findings could be useful for accurate determination of the **5* and REP7/6-harboring aberrant *CYP2D6* haplotypes in Asian populations.

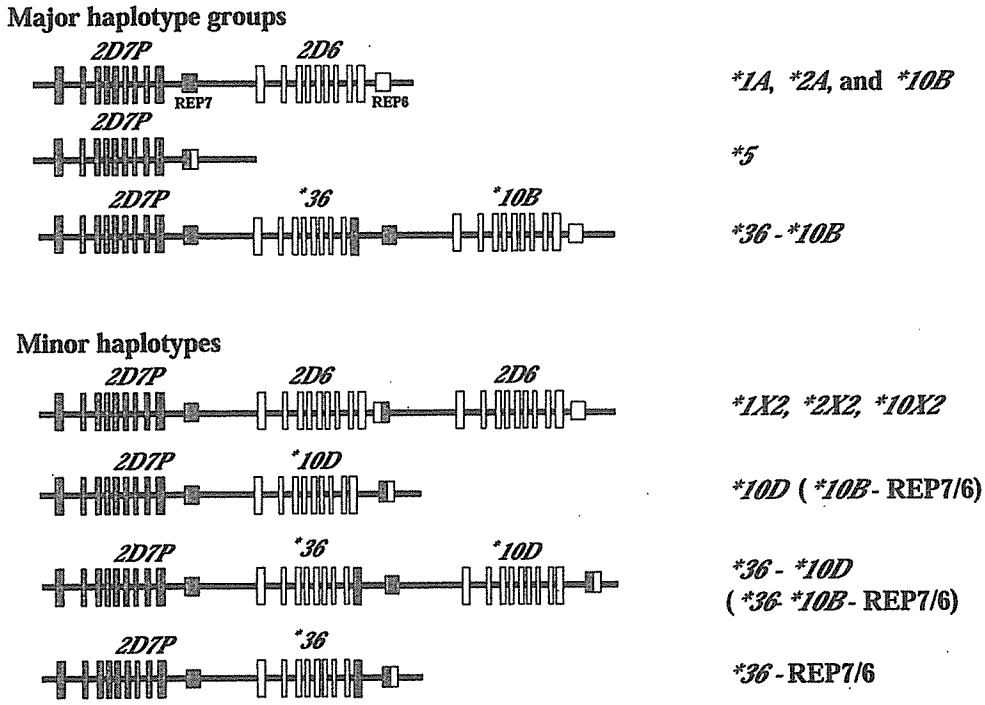

Fig. (3). Structures of *CYP2D6* haplotypes in a Japanese population. Exon and repetitive sequences derived from *CYP2D7P* are shown by dotted boxes.

Enzymatic activity of **10D* is considered almost the same as that of **10B* because sequences in the coding and proximal promoter regions of **10D* are identical to that of **10B* [Ishiguro *et al.*, 2004b; Soyama *et al.*, 2006b; our unpublished data]. Since CYP2D6.36 shows very low activities towards several drugs [Johansson *et al.*, 1994; Fukuda *et al.*, 2000; Hanioka *et al.*, 2006], **36-*10D* activity is considered similar to that of **10B* and **10D*. On the other hand, single-type **36* (*CYP2D6*36*-REP7/6) would be defective.

CYP3A4

The human cytochrome P450 (CYP) 3A subfamily has been estimated to be involved in the metabolism of 50% of the currently used therapeutic drugs [Wrighton *et al.*, 1996, Thummel and Wilkinson, 1998; Guengerich, 1999]. The *CYP3A5*, *CYP3A7*, *CYP3A4*, and *CYP3A43* genes consist of a cluster spanning 231 kb on chromosome 7 in the order listed above [Gellner *et al.*, 2001]. Overall, the CYP3A subfamily is the predominant P450 isoforms in human adult liver (approximately 30% of the total P450 content) [Shimada *et al.*, 1994]. The expression of CYP3A enzymes is differentially regulated in the developmental process: CYP3A7 levels are high in fetal liver, and CYP3A4 is abundant in adult liver. CYP3A5 is present in both fetal and adult livers, but its expression is known to be highly polymorphic. Since CYP3A43 is expressed at very low levels in several tissues including liver, it is believed not to play a substantial role in drug metabolism. In this review, we focus on the genetic polymorphisms of *CYP3A4*.

Among the subfamily members, CYP3A4 is the most predominant form in the adult human liver. This enzyme metabolizes a wide variety of substrates without structural similarity including steroids, fatty acids and xenobiotics (drugs, pesticides and carcinogens) [Wrighton *et al.*, 1996; Thummel and Wilkinson, 1998; Guengerich, 1999]. Up to 90-fold interindividual variations in CYP3A4 expression levels have been observed in Caucasian liver microsomes [Hustert *et al.*, 2001]. Furthermore, there are 40-60 fold variations in the metabolism of CYP3A substrates *in vivo* [Shimada *et al.*, 1994; Thummel and Wilkinson, 1998]. These interindividual differences are likely to influence pharmacokinetics, drug-drug interactions, efficacy, and adverse effects of drugs. Thus, it is clinically important to predict CYP3A4 activity in the liver or other tissues, such as the intestine.

CYP3A4 Polymorphisms

It has been suggested that approximately 85% of the interindividual variability in hepatic CYP3A4 activity is due to genetic factors [Ozdemir *et al.*, 2000]. Thus, several research groups have focused on the identification of *CYP3A4* variations (Lamba *et al.*, 2002). To date, 40 *CYP3A4* alleles (or haplotypes), including 20 subtypes, have been published on the Human Cytochrome P450 Allele Nomenclature Committee homepage (http://www.cypalleles.ki.se/cyp3a4.htm, as of July 11, 2006) [Lee and Goldstein, 2005; Krishna and Shekar, 2005 for review]. The distribution of *CYP3A4* alleles among different ethnic populations is summarized in Table 5 and Table 6.

An A to G mutation at –392 in the 5'-flanking region is designated as *CYP3A4*1B*. This allele is found at 0.53 to 0.87 frequencies in Africans, 0.04 to 0.10 in Caucasians, 0.06 to 0.09 in Hispanics, and 0.09 in Saudi, but is absent in other Asians (Table 5). The functional significance of this allele has been controversial. It has been reported that *CYP3A4*1B* caused a reduction in nuclear protein binding to

Table 5. Allelic Frequencies of *CYP3A4*1B* (-392A>G) in Different Ethnic Populations

Population	Allele Frequency	Number of Subjects	Reference
Caucasians			
Caucasian-American[¶]	0.096 0.036 0.090	94 273 132	Rebbeck *et al.* 1998 Ball *et al.* 1999 Walker *et al.* 1998
Finnish[¶]	0.042	59	Sata *et al.* 2000
Scottish[¶]	0.054	101	Tayeb *et al.* 2000
Dutch[¶]	0.053	199	van Schaik *et al.* 2000
Portuguese[¶]	0.040	100	Cavaco *et al.* 2003
Africans			
African-American[¶]	0.546 0.530 0.667	186 70 75	Ball *et al.* 1999 Walker *et al.* 1998 Sata *et al.* 2000
Ghanaian[¶]	0.690	100	Tayeb *et al.* 2000
Senegalese[¶]	0.780	178	Zeigler-Johnson *et al.* 2002
Nigerian[¶]	0.866	82	Kittles *et al.* 2002
Asians			
Japanese	ND ND	150 416	Naoe *et al.* 2000 Fukushima-Uesaka *et al.* 2004
Japanese-American	ND	77	Ball *et al.* 1999
Chinese-American	ND	78	Ball *et al.* 1999
Chinese	ND	96	Chowbay *et al.* 2003
Taiwanese	ND ND	130 59	Walker *et al.* 1998 Sata *et al.* 2000
Malay	ND	92	Chowbay *et al.* 2003
Indian	ND	87	Chowbay *et al.* 2003
Saudi[¶]	0.089	101	Tayeb *et al.* 2000
Hispanics			
Hispanic-American[¶]	0.093	188	Ball *et al.* 1999
Mexican[¶]	0.058	69	Reyes-Hernandez *et al.* 2004

ND: not detected.

[¶]Significant differences ($P<0.01$, chi-square test or Fisher's exact test) in allele frequencies between the Japanese population and each ethnic population. When plural studies were undertaken for each ethnic population, combined data were used for comparison. The multiple comparison was corrected by Bonferroni's method.

Table 6. Distribution of Nonsynonymous *CYP3A4* Alleles among Different Populations

Allele	Nucleotide Change	Amino Acid Change	Population	Allele Frequency	Number of Subjects	Functional Effect	Reference
*2	664T>C	S222P	Finnish	0.027	55	Altered activity depending on the substrates (*in vitro*)	Sata *et al.* 2000
			Portuguese	0.045	100		Cavaco *et al.* 2003