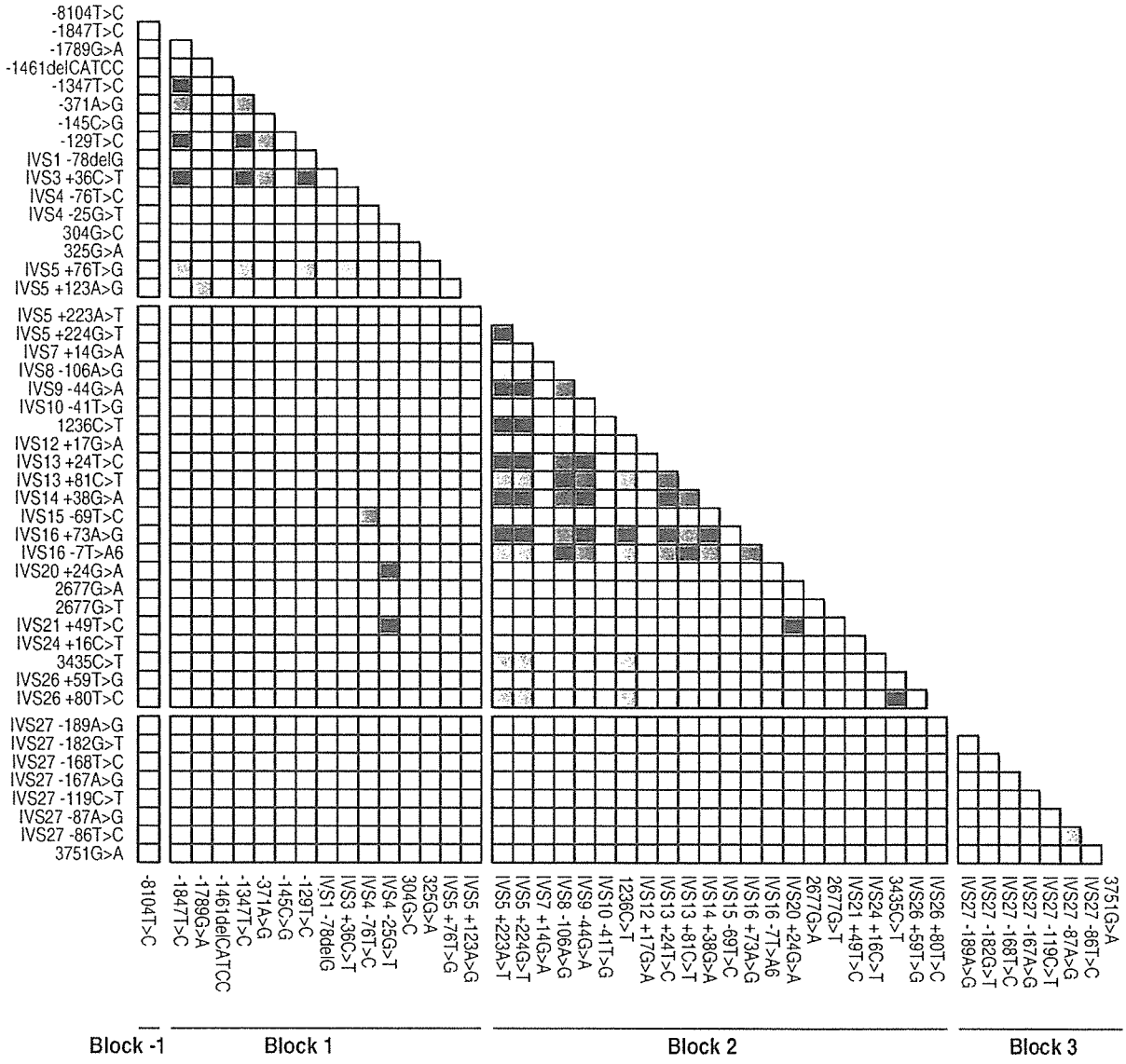


Figure 1 Linkage disequilibrium (LD) analysis of the *ABCB1* gene. Pairwise LD (r^2 values) of the polymorphisms detected in 5 or more chromosomes is shown as a 10-graded blue colour.

upstream of the transcriptional start site was assigned as Block −1 as described previously.

Haplotype Analysis

We estimated the diplotype configurations (haplotype combinations) of all 4 blocks using LDSUPPORT software. Diplotype configurations were obtained at probabilities over 0.9 for 100%, 92%, 95%, and 98% of the subjects for Blocks −1, 1, 2, and 3, respectively.

In Block −1, one novel haplotype, *1d, was identified. Thus Block 1 contained four *1 haplotypes (Fig. 2). The most common haplotype was *1a with a frequency of 0.988.

For Block 1 five haplotype groups consisting of 39 haplotypes were newly assigned. Of the 35 haplotypes in the *1 group, 10 haplotypes were ambiguous and were included as "Others" in Fig. 3. Haplotype groups *2 to *5 were defined by the nonsynonymous SNPs 325G>A(E109K) (*2), 304G>C(G102R) (*3),

49T>C(F17L) (*4) and 144G>T(K48N) (*5). The most frequent haplotype was *1a at a frequency of 0.541, followed by *1b (−1789G>A and IVS5 + 123A>G), *1c (IVS1 − 78delG), and *1d (IVS4 − 25G>T) at frequencies of 0.098, 0.079, and 0.041, respectively. The nonsynonymous *2 and *3 groups occurred at frequencies of 0.017 and 0.005, respectively.

Site		5'-Flanking				
Position		-8128 _8121	-8104	-7970		
Nucleotide change		del CTAA CCAA	T>C	C>T		
Amino acid change					N	Frequency
*1	*1a				1053	0.988
	*1b				2	0.002
	*1c				9	0.008
	*1d				2	0.002

Figure 2 *ABCB1* haplotypes in Block −1 for 533 Japanese subjects. The haplotype nomenclature followed the definitions used in our previous study (Sai *et al.* 2003). Newly identified haplotypes were consecutively named as shown in boldface.
N: number of chromosomes analyzed.

In Block 2 15 haplotype groups consisting of 61 haplotypes were inferred, including 38 newly-defined haplotypes. Of the 61 haplotypes 24 were detected in only one patient or ambiguously inferred. Ambiguous haplotypes within each group (groups *1, *8, and *10) were indicated as "Others" or "?" in Fig. 4. The most frequent haplotype was the *2 group at a frequency of 0.386, which harboured 1236C>T (exon 12), 2677G>T(A893S) (exon 21) and 3435C>T (exon 26). Groups *1, *10 [2677G>A(A893T)] and *8 (1236C>T) were found at frequencies of 0.216, 0.174 and 0.141, respectively. Other minor haplotype groups were *6 (3435C>T), *9 [1236C>T and 2677G>T(A893S)], *4 (1236C>T and 3435C>T) and *11 [1236C>T and 2677G>A(A893S)] at frequencies of 0.034, 0.020, 0.016, and 0.005, respectively. All these frequencies were comparable with our previous findings (Sai *et al.* 2003). Novel haplotype groups bearing amino acid substitutions were assigned as *12 [1804G>A (D602N)], *13 [2719G>A (V907I)], *14 [1342G>A (E448K)], *15 [2956A>G (M986V)], *16 [3043A>G (T1015A)], and *17 [2359C>T(R787W)],

Site														Ex. 1(5'-UTR)		Int. 1	Ex. 2	Int. 3	Ex. 4	Int. 4		Ex. 5		Int. 5			
Position		-2340	-2319	-1847	-1789	-1466	-1461 _-1457	-1347	-1279	-1081	-1065	-986	-371	-145	-129	IVS1 -78	49	IVS3 +36	144	IVS4 -76	IVS4 -25	304	325	IVS5 +76	IVS5 +123		
Nucleotide change		T>C	G>T	T>C	G>A	A>G	delCA TCC	T>C	A>G	T>C	T>C	A>G	A>G	C>G	T>C	del G	T>C	C>T	G>T	T>C	G>T	G>C	G>A	T>G	A>G		
Amino acid change																	F17L		K48N			G102R	E109K			N	Frequency
*1	*1a																										0.541
	*1b																										0.098
	*1c																										0.079
	*1d																										0.041
	*1e																										0.029
	*1f																										0.028
	*1g																										0.028
	*1h																										0.022
	*1i																										0.017
	*1j																										0.016
	*1k																										0.015
	*1L																										0.012
	*1m																										0.010
	*1n																										0.010
	*1o																										0.004
	*1p																										0.003
	*1q																										0.002
	*1r																										0.002
	*1s																										0.002
	*1t																										0.002
	*1u																										0.002
	*1v																										0.001
	*1w																										0.001
	*1x																										0.001
	*1y																										0.001
	Others																										0.009
*2	*2a																						2				0.017
*3	*3a																					3					0.005
*4	*4a?																4										0.001
*5	*5a?																		5								0.001

Figure 3 *ABCB1* haplotypes in Block 1 for 533 Japanese subjects. Block 1 haplotypes were newly defined due to the change of the Block 1 border. Rare and ambiguous haplotypes (n = 1) are shown with "?". Haplotypes assigned in only one patient or ambiguously inferred are shown as "Others". Sites for nonsynonymous substitutions are indicated by their group-name numbers.
N: number of chromosomes analyzed.

Site		Int. 5	Ex. 6	Int. 6	Int. 7	Int. 8	Int. 9	Int. 10	Ex. 12		Int. 12	Int. 13		Int. 14	Ex. 15	Int. 15		Int. 16			Int. 18		Ex. 19	Int. 19	Int. 20		Ex. 21		Int. 21		Ex. 22	Ex. 24		Int. 24	Ex. 26			
	Position	IVS5 +223	IVS5 +224	447	IVS6 -109	IVS7 +14	IVS8 -106	IVS9 -44	IVS10 -41	1238	1342	IVS12 +17	IVS13 +24	IVS13 +81	IVS14 +38	1804	IVS15 -95	IVS15 -89	IVS16 +52	IVS16 +73	IVS16 -78	IVS18 +87	IVS18 -35	2359	IVS19 -88	IVS20 +24	IVS20 -153	2677	2677	IVS21 +49	IVS21 -73_-76	2719	2956	3043	IVS24 +18	3435	IVS26 +59	IVS26 +80	N	Frequency
	Nucleotide change	A>T	G>T	A>G	G>T	G>A	A>G	G>A	T>G	C>T	G>A	G>A	T>C	C>T	G>A	A>G	T>C	A>C	A>G	T>A	C>T	G>C	C>T	T>C	G>A	A>G	G>A	G>T	T>C	del GTCT	G>A	A>G	A>G	C>T	C>T	T>G	T>C		
	Amino acid change			K149K					G412G	E448K						D602N							R787W				A593T	A593S			V907I	M966V	T1015A		I1145I				
*1	*1e																																					117	0.110
	*1f																																					47	0.044
	*1g																																					27	0.025
	*1L																																					10	0.009
	*1h																																					6	0.006
	*1k																																					6	0.006
	*1m																																					3	0.003
	*1n																																					3	0.003
	*1i																																					2	0.002
	*1o																																					2	0.002
	*1p																																					2	0.002
	Others																																					5	0.005
*2	*2d								2																			2							2			392	0.368
	*2e								2																			2							2			7	0.007
	*2f								2																			2							2			4	0.004
	*2g								2																			2							2			3	0.003
	*2h								2																			2							2			3	0.003
	*2i								2																			2							2			1	0.001
	*2j								2																			2							2			1	0.001
	*2k								2																			2							2			1	0.001
*4	*4a								4																									4			16	0.015
	*4c?								4																									4			1	0.001
*6	*6a																																	6			20	0.020
	*6d?																																	6			15	0.013
*8	*8a								8																												1	0.001
	*8b								8																												88	0.082
	*8e								8																												17	0.016
	*8h								8																												10	0.009
	*8d								8																												10	0.009
	*8i								8																												3	0.003
	*8k								8																												3	0.003
	*8f								8																												2	0.002
	Others								8																												1	0.001
*9	*9b								9																			9										6	0.006
	*9a								9																			9										14	0.013
*10	*10a																											10										7	0.007
	*10e																											10										180	0.169
	Others																																					1	0.001
*11	*11a								11																			11										5	0.005
	*11c?								11																			11										4	0.004
*12	*12a															12																						1	0.001
*13	*13a																														13							2	0.002
*14	*14a?									14																												1	0.001
*15	*15a?																															16						1	0.001
*16	*16a?																																16					1	0.001
*17	*17a?																							17														1	0.001
*18	*18a?																											18							18			1	0.001

Figure 4 *ABCB1* haplotypes in Block 2 for 533 Japanese subjects. Haplotype nomenclature followed the definitions used in our previous study (Sai *et al.* 2003) except for deletion of *8c. Newly identified haplotypes were consecutively named as shown in boldface. Rare and ambiguous haplotypes (n = 1) are shown with ";". Haplotypes assigned in only one patient or ambiguously inferred are shown as "Others". Marker sites for the *2 to *18 groups are indicated by their group-name numbers. N: number of chromosomes analyzed.

with frequencies that ranged from 0.002 for *12 and *13, to 0.001 for *14 to *17. Another new haplotype was defined as *18 based on the simultaneous presence of 2677G>A (A893T) and 3435C>T, with a frequency of 0.001. It was also noted that *1f in Block 2 was completely linked with *1d (IVS4 − 25G>T) in Block 1.

In Block 3 three haplotype groups consisting of 21 haplotypes were inferred, including four new haplotypes. Of the 21 haplotypes three were ambiguously inferred and included in "Others" in Fig. 5. The most frequent haplotype was *1a with a frequency of 0.753, followed by *1b (0.176). As observed previously, the rare haplotype groups *2 [3751G>A (V1251I)] and *3 [3587T>G (I1196S)] were observed at frequencies of 0.014 and 0.001, respectively.

We also analyzed the diplotype combinations for all 4 blocks (i.e. the whole gene) for all 533 subjects. The combination patterns were highly diverse with a total of 353 diplotype combinations observed. The frequencies for the majority of diplotypes were less than 0.01. The 10 major combinations are listed in Table 3; all combinations were made up of the major haplotypes in each block.

Network Analysis and Nucleotide Diversity

We performed a network analysis of the haplotypes in each block to obtain cladograms based on the sites and numbers of mutational events. For Block − 1 the rare haplotypes, *1b to *1d, appeared to be derived from the major haplotype *1a (Fig. 6a). For Block 1 most of the minor haplotypes were connected to one of the major haplotypes *1a or *1b. However, *1e, *1g, *1h, and *1p were shown to be distant from the above haplotypes (Fig. 6b). Haplotype groups including *1b and the closely related haplotypes *1f, *1i, *1k, *1j, *1m, *1L, and *1q were characterized by the presence of − 1789G>A. Of these haplotypes the *1k and *1j subgroups were characterized by the additional SNP − 371A>G, while the subgroups *1m,*1L and *1q contained the SNP − 145C>G. The separate subgroup that consisted of *1e, *1g, *1h and *1p contained

Site		Int. 26	Ex. 27	Int. 27										Ex. 28			
Position		IVS26 -78	3587	IVS27 +63	IVS27 -189	IVS27 -182	IVS27 -172	IVS27 -168	IVS27 -167	IVS27 -119	IVS27 -87	IVS27 -86	IVS27 -80	3747	3751		
Nucleotide change		C>G	T>G	C>G	A>G	G>T	G>A	T>C	A>G	C>T	A>G	T>C	ins C	C>G	G>A		
Amino acid change			I1196S											G1249G	V1251I	N	Frequency
*1	*1a															803	0.753
	*1b															188	0.176
	*1c															27	0.025
	*1d															5	0.005
	*1f															5	0.005
	*1h															4	0.004
	*1j															4	0.004
	*1e															3	0.003
	*1L															2	0.002
	*1g															1	0.001
	*1i															1	0.001
	*1k															1	0.001
	*1m															1	0.001
	*1o															1	0.001
	*1r															1	0.001
	Others															3	0.003
*2	*2a														2	9	0.008
	*2b														2	6	0.006
*3	*3a		3													1	0.001

Figure 5 *ABCB1* haplotypes in Block 3 for 533 Japanese subjects. Haplotype nomenclature followed the definitions used in our previous study (Sai *et al.* 2003). Newly identified haplotypes were consecutively named as shown in boldface. Haplotypes assigned in only one patient or ambiguously inferred are shown as "Others". Sites for nonsynonymous substitutions are indicated by their group-name numbers.
N: number of chromosomes analyzed.

Table 3 Diversity of block diplotype combinations across the 4 blocks

Block −1	Block 1	Block 2	Block 3	Number of subjects
*1a/*1a	*1a/*1a	*2d/*2d	*1a/*1a	24
*1a/*1a	*1a/*1a	*2d/*2d	*1b/*1a	13
*1a/*1a	*1c/*1a	*10a/*2d	*1a/*1a	12
*1a/*1a	*1a/*1a	*2d/*1e	*1b/*1a	11
*1a/*1a	*1e/*1a	*10a/*2d	*1a/*1a	9
*1a/*1a	*1a/*1a	*2d/*2d	*1c/*1a	9
*1a/*1a	*1b/*1a	*8a/*2d	*1a/*1a	9
*1a/*1a	*1c/*1a	*10a/*1e	*1b/*1a	6
*1a/*1a	*1g/*1a	*10a/*2d	*1a/*1a	6
*1a/*1a	*1d/*1c	*10a/*1f	*1a/*1a	5
*1a/*1a	*2a/*1a	*10a/*2d	*1a/*1a	5
*1a/*1a	*1c/*1a	*10a/*2d	*1b/*1a	5

A total of 353 diplotype-combinations across the 4 blocks were detected in 533 subjects.
The number of subjects for the other combinations was less than 5.

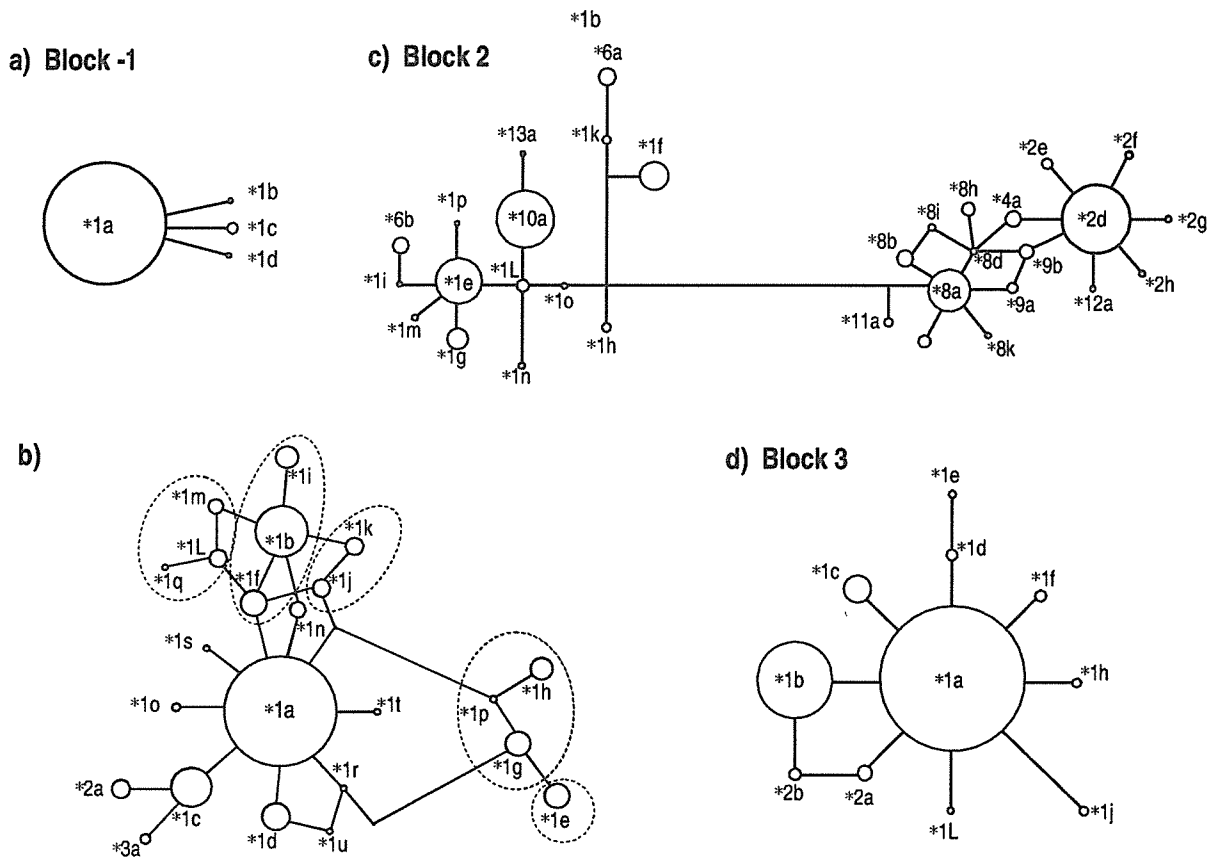


Figure 6 Network analysis of *ABCB1* haplotypes of Block −1 (a), Block 1 (b). Block 2 (c) and Block 3 (d). For each block, the circle area represents the approximate haplotype frequency, and line length between the circles is proportional to the number of mutations. Haplotypes inferred in only one chromosome were omitted from this analysis. The classification by Takane *et al.* (2004) is indicated with dashed lines.

the three linked SNPs, −1847T>C, −1347T>C and −129T>C. This network profile supported the previous classification of promoter region haplotypes by Takane *et al.* (2004), as indicated by the dashed lines. However, our current study revealed the presence of additional subtypes. Detailed comparisons between our Block 1 haplotypes and previously described promoter region haplotypes are described in the next section.

The Block 2 cladogram showed that there were four major haplotypes, *2d, *10a, *1e and *8a, and that most of the minor haplotypes appeared to be derived from *1e, *2d or *8a (Fig. 6c). Network analysis showed that the *2 and *8 haplotypes, which share the SNP 1236C>T, were distant from *1e and *10a, and that *10a and the adjacent *13a that both have 2677G>A without the common polymorphisms 1236C>T, 2677G>T and 3435C>T, are relatively closely related to the *1 group. The *8 group bearing 1236C>T was highly diverse and included many haplotypes. The *4 (1236C>T and 3435C>T) and *9 (1236C>T and 2677G>T) groups were related to *8 and *2. This network profile supported the previous classification of Block 2 haplotypes based on common polymorphisms. For the *6 group, containing 3435C>T, *6a and *6b were distantly related to *1 haplotypes, which suggested that different mutational and/or recombinational events were responsible for these haplotypes.

For Block 3, *1b and the other minor *1 haplotypes were related to the major *1a haplotype, while the *2 group (*2a and *2b) with nonsynonymous SNPs appeared to be derived from either *1a or *1b (Fig. 6d).

Comparison of Block 1 Haplotypes with Reported Promoter Haplotypes

To compare our Block 1 haplotype structures with the previously reported promoter region haplotypes (Taniguchi *et al.* 2003: Takane *et al.* 2004), we classified Block 1 haplotypes into 6 subgroups (A, B, E, G, J and L) based on the marker sites in the reported haplotypes and the network analysis performed in our present study. The summary of this comparison is shown in Fig. 7.

Haplotypes that did not harbour any of the previous markers were classified into subgroup A, in which *1a was the major haplotype. The B subgroup, which included *1b, was defined as haplotypes that contained −1789G>A. Haplotypes that contained additional SNPs −371A>G or −145C>G were classified into subgroups J (including *1j) or L (including *1L), respectively. Subgroup G was defined as those haplotypes that contained the three linked variations (−1847T>C, −1347T>C and −129T>C) in which *1g was the major haplotype, and subgroup E (including *1e) was defined as those haplotypes with the three linked variations plus -1461_-1457delCATCC.

As indicated by the cladograms in the previous section, our study revealed that subgroup A, previously classified as wild-type, could be further classified into six types: the major *1a type without any marker variation and five other types with either IVS1 −78delG (*1c), IVS4 − 25G>T (*1d), 325G>A(E109K) (*2a), IVS5 + 123A>G (*1n), or 304 G>C(G102R) (*3a). Each of the B, J, and L subgroups that shared −1789G>A were further divided into two types based on the presence of IVS5 + 123A>G. Subgroup G, with the three reported marker SNPs, was also linked to IVS3 + 36C>T, and this subgroup was further characterized by the presence of IVS4 − 25G>T (*1h and *4a?) or IVS5 + 76T>G (*1g). Subgroup E was linked with both IVS3 + 36C>T and IVS5 + 76T>G.

Ethnic Differences

It is well known that there are differences in the frequencies of functionally important haplotypes involving common SNPs (1236C>T, 2677G>T/A and 3435C>T) and promoter region SNPs between different ethnic groups (Kim *et al.* 2001; Kroetz *et al.* 2003; Tang *et al.* 2002, 2004; Takane *et al.* 2004). To characterize these haplotypes in the Japanese population, we compared the frequencies of Block 1 and Block 2 haplotypes that harbour common SNPs with representative reported data from different ethnic groups (Kroetz *et al.* 2003; Takane *et al.* 2004) (Tables 4 and 5). Block 1 haplotype frequencies were generally consistent with those from previous reports for Japanese (Takane *et al.* 2004), except that our study did not detect the reported H7 haplotype that contained −1154T>C alone (Table 4). It has also been suggested that there is much more haplotypic variation in Japanese than in Caucasian populations (Takane *et al.* 2004), and our study supported this.

As for Block 2 haplotypes, the *1 and *2 groups were the common major haplotypes in all the ethnic groups.

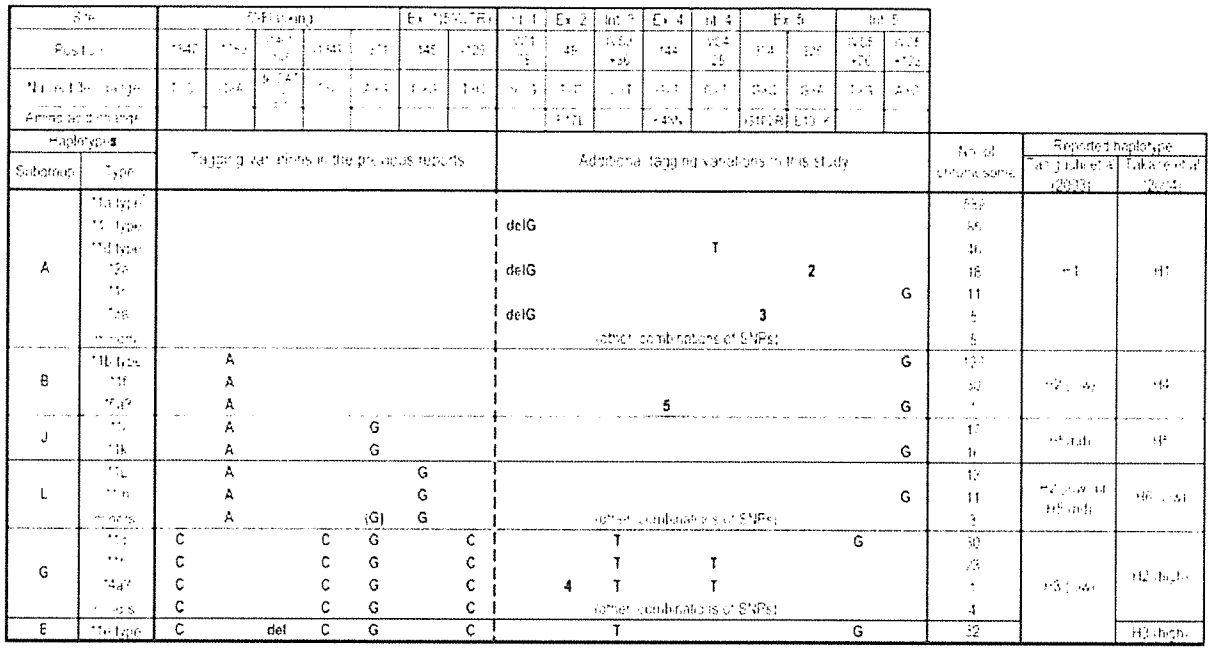


Figure 7 New classification of Block 1 haplotypes and comparison with reported promoter region haplotypes. Genetic variations (allele frequency >0.01) and nonsynonymous variations in Block 1 were sorted according to marker variation, and classified into 6 subgroups (A, B, J, L, G and E).
[a]The positions in other reports were adjusted to the nucleotide numbers used in this study.
[b]The *1a type includes *1a, *1o, *1s, *1t, *1v, *1w, *1x, *1y.
[c]The *1c type includes the *1c haplotype and an ambiguously defined *1 haplotype.
[d]The *1d type includes the *1d haplotype and two ambiguously defined *1 haplotypes.
[e]"Minors" include the *1u and *1r haplotypes and one ambiguously defined *1 haplotype.
[f]The *1b type includes the *1b and *1i haplotypes and three ambiguously defined *1 haplotypes.
[g]"Minors" include the *1q haplotype and one ambiguously defined *1 haplotype.
[h]"Minors" include the *1p haplotype and one ambiguously defined *1 haplotype.
[i]The *1e type includes the *1e haplotype and one ambiguously defined *1 haplotype.
[j]Altered promoter activity in the reporter gene assay is shown in parenthesis.
nd; not determined.

However, the frequency of the *2 group was much lower than that of the *1 group in Africans. The frequencies of *4 and *8 were higher in Japanese than in Caucasians, and the frequency of the *6 group was higher in Caucasians than in other ethnic groups. The most prominent characteristic of the Japanese population was the high frequency of *10 compared with the other ethnic groups. The variations that characterized *11 to *18 were only detected in our study, probably due to the relatively large number of subjects used. The haplotype distribution in Japanese was similar to that described for Asians, but with slight differences in the frequencies of *6, *8, *9, and *10 reported for a mixed Asian population (Kroetz *et al.* 2003).

Tagging SNPs for *ABCB1* Genotyping

For genotyping *ABCB1* in association studies it would be critical to select SNPs for the major haplotypes, including functional ones in Blocks 1 and 2. Table 6 shows the major tagging SNPs for genotyping which are applicable to Japanese and also to other ethnic populations. Genotyping with these SNPs can assign the diplotypes of Blocks 1 and 2 in more than 95% of Japanese. The nonsynonymous SNPs in Blocks 1 and 2, and the additional tagging variations in Block 1 obtained in our study (Fig. 7), could be included in the list for evaluation of their functional significance.

Table 4 Ethnic differences in *ABCB1* Block 1 haplotypes

Marker site[a]	This study: Subgroup (see Fig. 7)	This study: Japanese (n = 1066)	Reported data (Takane *et al.* 2004): Group	Reported data (Takane *et al.* 2004): Japanese (n = 188)	Reported data (Takane *et al.* 2004): Caucasian (n = 192)
	A	0.712	H1	0.665	0.964
−1789G>A	B	0.147	H4	0.191	nd
−1789G>A, −371A>G	J	0.031	H5	0.027	nd
−1789G>A, −145C>G	L	0.025	H6	0.032	nd
−1847T>C[b]	G	0.054	H2	0.043	nd
−1461delCATCC, −371A>G, −1847T>C[b]	E	0.030	H3	0.037	nd
−1154T>C		nd	H7	0.005	nd
−1753delGA		nd	H8	nd	0.010
−1347T>C, −129T>C		nd	H9	nd	0.016
−1085A>G		nd	H10	nd	0.010

[a]Each reported position was adjusted to the nucleotide numbers used in this study.
[b]This SNP is linked to −1347T>C and −129T>C.
n; 2 × number of subjects.
nd; not detected.

Table 5 Ethnic differences in the *ABCB1* Block 2 haplotypes

Group	This study: Japanese (n = 1066)	Reported data (Kroetz *et al.* 2003)[a]: Asian (n = 60)	Caucasian (n = 200)	African (n = 200)
*1	0.216	0.216	0.370	0.721
*2	0.386	0.365	0.410	0.075
*3	nd	nd	0.010	0.010
*4	0.016	0.016	0.005	0.090
*6	0.034	0.016	0.120	0.035
*7	nd	nd	0.015	0.005
*8	0.141	0.216	0.010	0.040
*9	0.020	0.082	0.025	0.010
*10	0.174	0.066	0.025	0.005
*11	0.005	nd	nd	nd
*12	0.002	nd	nd	nd
*13	0.002	nd	nd	nd
*14	0.001	nd	nd	nd
*15	0.001	nd	nd	nd
*16	0.001	nd	nd	nd
*17	0.001	nd	nd	nd
*18	0.001	nd	0.01	nd

[a]Reported haplotypes were re-assigned according to our haplotype nomenclature.
n = 2 × number of subjects.
nd; not detected.

Discussion

Extensive studies of *ABCB1* haplotypes and their functional significance have been conducted, mostly focused on the common SNPs of 1236C>T, 2677G>T/A, and 3435C>T. However, recent association studies on promoter region haplotypes have indicated the importance of haplotypes within this region (Taniguchi *et al.* 2003; Takane *et al.* 2004). The results of functional or P-gp expression analyses based on these polymorphisms/haplotypes have not always been consistent, possibly due to the small number of subjects used, different ethnic backgrounds, or insufficient haplotyping over a limited region. In the present study, we have conducted a re-assignment of Block 1 haplotypes by extending the region sequenced to the distal promoter, and

Table 6 Major tagging SNPs of *ABCB1* for genotype-phenotyhpe association studeis

i) Block 1 haplotypes (subgroups)					
Position	−1847	−1789	−1461_−1457	−371	−145
Nucleotide change	T > C	G > A	delCATCC	A > G	C > G
	a	a	a	a	a
A					
B		A			
J		A		G	
L		A			G
G	C			G	
E	C		del	G	

ii) Block 2 haplotypes				
Position	1236		2677	3435
Nucleotide change	C > T	G > A	G > T	C > T
Amino acid change		A893T	A893S	
		a		
*1				
*2	T		T	T
*4	T			T
*6				T
*8	T			
*9	T		T	
*10		A		
*11	T	A		
*18		A		T

[a]Specific for Asian populatioins.

added novel haplotypes in other blocks after assessing a large number of subjects.

LD analysis revealed that one of the marker SNPs in the promoter region, −1789G>A, was moderately linked to IVS5 + 123A>G, previously classified into Block 2. Therefore, we shifted the border between Block 1 and Block 2 and re-analyzed the Block 1 haplotypes. Two promoter haplotype classes associated with functional changes have been reported previously (Taniguchi *et al.* 2003; Takane *et al.* 2004). One class included the −1789G>A SNP, and the other included the three linked SNPs of −1847T>C, −1347T>C and −129T>C. In our analysis these SNPs were included in our Block 1 region.

The haplotype containing −1789G>A was reported to be associated with reduced P-gp expression levels in the colon and liver, and reduced promoter activity was shown in a reporter gene assay (Taniguchi *et al.* 2003) (see Fig. 7). However, another study found that a haplotype containing −1789G>A without −145C>G (subgroups B and J in our present study) showed no change in the reporter assay, while another haplotype that contained −1789G>A together with −145C>G (subgroup L) showed reduced promoter activity (Takane *et al.* 2004). Data on the functional effects of haplotypes harbouring the three linked SNPs (G and E subgroups) are also contradictory. While one study showed an association with reduced colon and liver P-gp expression levels in patients and reduced promoter activity in a reporter gene assay (Taniguchi *et al.* 2003), another study reported an association with increased P-gp expression levels in the placenta and liver, and with increased promoter activity in a reporter gene assay (Takane *et al.* 2004). By expanding Block 1 into intron 5 we identified additional types within previously reported wild-type sequences (corresponding to subgroup A in this study) and other variant haplotypes (subgroups B, E, G, J, and L) (Fig. 7). In total our data revealed 11 tagging variations in Block 1: −1789G>A, −1461_−1457CATCdel, −371A>G, −145C>G, −129T>C, IVS1 − 78delG, IVS4 − 25G>T, 304G>C (G102R), 325G>A (E109K), IVS + 76T>G and IVS5 + 123A>G. Thus, if some of these markers are of functional importance it is possible that

our subdivisions (types within A and other subgroups) might explain the discrepancies in P-gp expression levels in the previously reported studies. In fact, our preliminary observation has suggested possible influences of some of the tagging variations in Block 1 on pharmacokinetic parameters of paclitaxel (data not shown). However, this hypothesis requires further clarification in large scale clinical studies.

Several novel haplotypes were added to the other 3 blocks (1, 38, and 4 new haplotypes in Blocks −1, 2, and 3, respectively). We identified a new haplotype *1d in Block −1, but this variant haplotype was very rare and the functional significance of uncommon Block −1 haplotypes remains unknown. We added 7 new groups to Block 2 haplotypes (*12 to *18), but their frequencies were also very low (0.002 and less). We also confirmed the previous finding that, in order of frequency, the major groups were *2d, *10a, *1e, and *8a. In our previous study we estimated the relative P-gp activity of the different haplotypes according to the renal clearance of irinotecan and its metabolites in Japanese cancer patients. While we found a significant association between *2, which contained the three common SNPs, and reduced renal clearance levels, associations with the *6, *8, and *10 groups that contained only one of the common markers remained unclear. For the *4, *9, and *11 groups, which harbour two marker SNPs in Block 2, functional evaluation was impossible due to the small number of subjects. Previously we showed that *1f may have been associated with reduced P-gp activity. The current study revealed that *1f in Block 2 was completely linked with the newly defined *1d in Block 1, which contained IVS4 −25G>T. A further association study is needed to clarify the effects of the linked *1d (Block 1) and *1f (Block 2) haplotypes. Regarding Block 3 we added several minor *1 haplotypes and confirmed the previous findings that *1a and *1b were the major haplotypes. We previously observed a trend for an association between *1b and higher P-gp activity. Taking into consideration the haplotype-combinations across the blocks this trend also needs to be confirmed in a larger number of subjects.

It is well recognized that there are large ethnic differences in the frequencies of functionally important haplotypes, including 1236C>T, 2677G>T, and 3435>T (corresponding to the *2 group in Block 2), and promoter region SNPs (corresponding to the variant Block 1 subgroups). Comparison of our data with the results from other ethnic groups indicated the existence of unique haplotype profiles in the Japanese population. As suggested by the previous report on the promoter region (Takane *et al.* 2004), Japanese samples exhibited large variations in Block 1 haplotypes. This suggested that not only *2 in Block 2 but also certain Block 1 haplotypes may be functionally important in the Japanese ethnic group. For Block 2 we confirmed our previous findings that the major groups were *1 and *2, and that *2d was the most frequent haplotype. While both groups were detected as the major types in other Asian and Caucasian populations, *1 was considerably more frequent than *2 in Africans (Kroetz *et al.* 2003). Another recent study found that the two major haplotypes were common to 5 ethnic groups (Tang *et al.* 2004). That study also revealed that the Chinese and Malay haplotype profiles were very similar, and that while some similarities were also observed between Caucasian and Indian populations, Africans differed from all other non-African populations. Furthermore, their study suggested that positive selection for 2677T-3435T had occurred in Chinese and Malays, and for 3435C in Africans. As pointed out previously, frequent occurrence of *10 (2677G>A) was unique to Japanese compared with Caucasians and African populations. Our study revealed higher frequencies of *10 (2677G>A) and *6 (3435C>T) and lower frequencies of *8 (1236C>T) and *9 (1236C>T and 2677G>T) than reported for Asian populations in Kroetz *et al.* (2003) (Table 5). This difference might be due to the mixed Asian population used in the report, as differences in the frequencies of 2677G>A between the Chinese, Malay, and Indian populations have been noted (Tang *et al.* 2004). The finding that the high frequency of 2677G>A is shared among Japanese, Koreans (Yi *et al.* 2004) and Chinese (Tang *et al.* 2004) suggests a close evolutionary relationship between these three populations.

A whole-genome haplotype database for three populations is now available at the Perlegen website (www.perlegen.com), which provides a good tool for investigation of the structures of human genetic variation within and between different populations (Hinds *et al.* 2005). For the *ABCB1* gene, however, we could not directly compare their data with ours because their

SNPs are mostly intronic and did not overlap with our SNP markers (<20%).

For genotype-phenotype association studies on the *ABCB1* gene, genotyping of the major functional key SNPs in Blocks 1 and 2 (Table 6) would be useful. Further studies on the clinical significance of the haplotypes described in the present study and elucidation of the haplotype-combinations across blocks, will be required to achieve the goal of personalized drug therapy.

Conclusions

We re-established *ABCB1* haplotypes in the Japanese population based on novel polymorphisms found in a large number of subjects, expanding the promoter region. Our current data added more detailed information on functionally-important haplotypes in Blocks 1 and 2 in the Japanese population, and identified differences in haplotype profiles between ethnic groups. The information provided in this study will be of use in further studies investigating the relationship between genetic markers and functional changes.

Acknowledgements

This study was supported in part by the Program for the Promotion of Fundamental Studies in Health Sciences and in part by the Program for the Promotion of Studies in Health Sciences of the Ministry of Health, Labor and Welfare of Japan.

References

Chowbay, B., Cumaraswamy, S., Cheung, Y. B., Zhou, Q. & Lee, E. J. (2003) Genetic polymorphisms in MDR1 and CYP3A4 genes in Asians and the influence of MDR1 haplotypes on cyclosporin disposition in heart transplant recipients. *Pharmacogenetics* **13**, 89–95.

Cordon-Cardo, C., O'Brien, J. P., Casals, D., Rittman-Grauer, L., Biedler, J. L., Melamed, M. R. & Bertino, J. R. (1989) Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites. *Proc Natl Acad Sci U S A* **86**, 695–698.

Debry, P., Nash, E. A., Neklason, D. W. & Metherall, J. E. (1997) Role of multidrug resistance P-glycoproteins in cholesterol esterification. *J Biol Chem* **272**, 1026–1031.

Drach, J., Gsur, A., Hamilton, G., Zhao, S., Angerler, J., Fiegl, M., Zojer, N., Raderer, M., Haberl, I., Andreeff, M. & Huber, H. (1996) Involvement of P-glycoprotein in the transmembrane transport of interleukin-2 (IL-2), IL-4, and interferon-gamma in normal human T lymphocytes. *Blood* **88**, 1747–1754.

Fojo, A. T., Ueda, K., Slamon, D. J., Poplack, D. G., Gottesman, M. M. & Pastan, I. (1987) Expression of a multidrug-resistance gene in human tumors and tissues. *Proc Natl Acad Sci USA* **84**, 265–269.

Furuno, T., Landi, M. T., Ceroni, M., Caporaso, N., Bernucci, I., Napp, G., Martignoni, E., Schaeffeler, E., Eichelbaum, M., Schwab, M. & Zanger, U. M. (2002) Expression polymorphism of the blood-brain barrier component P-glycoprotein (MDR1) in relation to Parkinson's disease. *Pharmacogenetics* **12**, 529–534.

Geick, A., Eichelbaum, M. & Burk, O. (2001) Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* **276**, 14581–14587.

Goto, M., Masuda, S., Saito, H., Uemoto, S., Kiuchi, T., Tanaka, K. & Inui, K. (2002) C3435T polymorphism in the MDR1 gene affects the enterocyte expression level of CYP3A4 rather than Pgp in recipients of living-donor liver transplantation. *Pharmacogenetics* **12**, 451–457.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A. & Cox, D. R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079.

Hoffmann, U. & Kroemer, H. K. (2004) The ABC transporters MDR1 and MRP2: multiple functions in disposition of xenobiotics and drug resistance. *Drug Metab Rev* **36**, 669–701.

Hoffmeyer, S., Burk, O., von Richter, O., Arnold, H. P., Brockmoller, J., Johne, A., Cascorbi, I., Gerloff, T., Roots, I., Eichelbaum, M. & Brinkmann, U. (2000) Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci USA* **97**, 3473–3478.

Ieiri, I., Takane, H. & Otsubo, K. (2004) The MDR1 (ABCB1) gene polymorphism and its clinical implications. *Clin Pharmacokinet* **43**, 553–576.

Itoda, M., Saito, Y., Komamura, K., Ueno, K., Kamakura, S., Ozawa, S. & Sawada, J. (2002) Twelve Novel Single Nucleotide Polymorphisms in ABCB1/MDR1 among Japanese Patients with Ventricular Tachycardia who were Administered Amiodarone. *Drug Metab Pharmacokinet* **17**, 566–571.

Johne, A., Kopk, K., Gerloff, T., Mai, I., Rietbrock, S., Meisel, C., Hoffmeyer, S., Kerb, R., Fromm, M. F., Brinkmann, U., Eichelbaum, M., Brockmoller, J., Cascorbi, I. & Roots, I. (2002) Modulation of steady-state kinetics of digoxin by haplotypes of the P-glycoprotein MDR1 gene. *Clin Pharmacol Ther* **72**, 584–594.

Kim, R. B., Leake, B. F., Choo, E. F., Dresser, G. K., Kubba, S. V., Schwarz, U. I., Taylor, A., Xie, H. G., McKinsey, J., Zhou, S., Lan, L. B., Schuetz, J. D., Schuetz, E. G. & Wilkinson, G. R.. (2001) Identification of functionally

variant MDR1 alleles among European Americans and African Americans. *Clin Pharmacol Ther* **70**, 189–199.

Kim, R. B. (2002) MDR1 single nucleotide polymorphisms: multiplicity of haplotypes and functional consequences. *Pharmacogenetics* **12**, 425–427.

Kimchi-Sarfaty, C., Gribar, J. J. & Gottesman, M. M. (2002) Functional characterization of coding polymorphisms in the human MDR1 gene using a vaccinia virus expression system. *Mol Pharmacol* **62**, 1–6.

Kitamura, Y., Moriguchi, M., Kaneko, H., Morisaki, H., Morisaki, T., Toyama, K. & Kamatani, N. (2002) 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.

Kurata, Y., Ieiri, I., Kimura, M., Morita, T., Irie, S., Urae, A., Ohdo, S., Ohtani, H., Sawada, Y., Higuchi, S. & Otsubo, K. (2002) Role of human MDR1 gene polymorphism in bioavailability and interaction of digoxin, a substrate of P-glycoprotein. *Clin Pharmacol Ther* **72**, 209–219.

Kroetz, D. L., Pauli-Magnus, C., Hodges, L. M., Huang, C. C., Kawamoto, M., Johns, S. J., Stryke, D., Ferrin, T.E., DeYoung, J., Taylor, T., Carlson, E. J., Herskowitz, I., Giacomin, i K. M. & Clark, A. G. (2003) Pharmacogenetics of Membrane Transporters Investigators. Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. *Pharmacogenetics* **13**, 481–494.

Liu, Y. Y., Han, T. Y., Giuliano, A. E. & Cabot, M. C. (2001) Ceramide glycosylation potentiates cellular multidrug resistance. *FASEB J* **15**, 719–730.

Morita, N., Yasumori, T. & Nakayama, K. (2003) Human MDR1 polymorphism: G2677T/A and C3435T have no effect on MDR1 transport activities. *Biochem Pharmacol* **65**, 1843–1852

Riordan, J. R., Deuchars, K., Kartner, N., Alon, N., Trent, J. & Ling, V. (1985) Amplification of P-glycoprotein genes in multidrug-resistant mammalian cell lines. *Nature* **316**, 817–819.

Sai, K., Kaniwa, N., Itoda, M., Saito, Y., Hasegawa, R., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Shirao, K., Minami, H., Ohtsu, A., Yoshida, T. Saijo, N., Kitamura, Y., Kamatani, N., Ozawa, S. & Sawada, J. (2003) Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan. *Pharmacogenetics* **13**, 741–757.

Schwab, M., Schaeffeler, E., Marx, C., Fromm, M. F., Kaskas, B., Metzler, J., Stange, E., Herfarth, H., Schoelmerich, J., Gregor, M., Walker, S., Cascorbi, I., Roots, I., Brinkmann, U., Zanger, U. M. & Eichelbaum, M. (2003) Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. *Gastroenterology* **124**, 26–33.

Siegsmund, M., Brinkmann, U., Schaffeler, E., Weirich, G., Schwab, M., Eichelbaum, M., Fritz, P., Burk, O., Decker, J., Alken, P., Rothenpieler, U., Kerb, R. & Hoffmeyer, S. (2002) Association of the P-glycoprotein transporter MDR1(C3435T) polymorphism with the susceptibility to renal epithelial tumors. *J Am Soc Nephrol* **13**, 1847–1854.

Takane, H., Kobayashi, D., Hirota, T., Kigawa, J., Terakawa, N., Otsubo, K. & Ieiri, I. (2004) Haplotype-oriented genetic analysis and functional assessment of promoter variants in the MDR1 (ABCB1) gene. *J Pharmacol Exp Ther* **311**, 1179–1187.

Tanabe, M., Ieiri, I., Nagata, N., Inoue, K., Ito, S., Kanamori, Y., Takahashi, M., Kurata, Y., Kigawa, J., Higuchi, S., Terakawa, N. & Otsubo, K. (2001) Expression of P-glycoprotein in human placenta: relation to genetic polymorphism of the multidrug resistance (MDR)-1 gene. *J Pharmacol Exp Ther* **297**, 1137–1143.

Tang, K., Ngoi, S. M., Gwee, P. C., Chua, J. M., Lee, E. J., Chong, S. S. & Lee, C. G. (2002) Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug transporter gene locus in three ethnic Asian populations. *Pharmacogenetics* **12**, 437–450.

Tang, K., Wong, L. P., Lee, E. J., Chong, S. S. & Lee, C. G. (2004) Genomic evidence for recent positive selection at the human MDR1 gene locus. *Hum Mol Genet* **13**, 783–797.

Taniguchi, S., Mochida, Y., Uchiumi, T., Tahira, T., Hayashi, K., Takagi, K., Shimada, M., Maehara, Y., Kuwano, H., Kono, S., Nakano, H., Kuwano, M. & Wada, M. (2003) Genetic polymorphism at the 5′ regulatory region of multidrug resistance 1 (MDR1) and its association with interindividual variation of expression level in the colon. *Mol Cancer Ther* **2**, 1351–1359.

van Helvoort, A., Smith, A. J., Sprong, H., Fritzsche, I., Schinkel, A. H., Borst, P. & van Meer, G. (1996) MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. *Cell* **87**, 507–517.

Wong, M., Evans, S., Rivory, L. P., Hoskins, J. M., Mann, G. J., Farlow, D., Clarke, C. L., Balleine, R. L. & Gurney, H. (2005) Hepatic technetium Tc 99m-labeled sestamibi elimination rate and ABCB1 (MDR1) genotype as indicators of ABCB1 (P-glycoprotein) activity in patients with cancer. *Clin Pharmacol Ther* **77**, 33–42.

Yi, S. Y., Hong, K. S., Lim, H. S., Chung, J. Y., Oh, D. S., Kim, J. R., Jung, H. R., Cho, J. Y., Yu, K. S., Jang, I. J. & Shin, S. G. (2004) A variant 2677A allele of the MDR1 gene affects fexofenadine disposition. *Clin Pharmacol Ther* **76**, 418–427.

Received: 20 April 2005
Accepted: 21 October 2005

Drug Metab. Pharmacokinet. 21 (2): 109-121 (2006).

Regular Article

Genetic Variation and Haplotype Structure of the ABC Transporter Gene ABCG2 in a Japanese Population

Keiko MAEKAWA[1,2], Masaya ITODA[1], Kimie SAI[1,3], Yoshiro SAITO[1,2], Nahoko KANIWA[1,4], Kuniaki SHIRAO[6], Tetsuya HAMAGUCHI[6], Hideo KUNITOH[7], Noboru YAMAMOTO[7], Tomohide TAMURA[7], Hironobu MINAMI[8], Kaoru KUBOTA[9], Atsushi OHTSU[10], Teruhiko YOSHIDA[11], Nagahiro SAIJO[12], Naoyuki KAMATANI[13], Shogo OZAWA[1,5] and Jun-ichi SAWADA[1,2]

[1]*Project team for Pharmacogenetics,* [2]*Division of Biochemistry and Immunochemistry,*
[3]*Division of Xenobiotic Metabolism and Disposition,* [4]*Division of Medicinal Safety Science,*
[5]*Division of Pharmacology, National Institute of Health Sciences, Tokyo, Japan*
[6]*Gastrointestinal Oncology Division,* [7]*Thoracic Oncology Division, National Cancer Center Hospital,*
[11]*Genetics Division, National Cancer Center Research Institute, National Cancer Center, Tokyo, Japan*
[8]*Division of Oncology/Hematology,* [9]*Thoracic Oncology Division,*
[10]*Division of GI Oncology/Digestive Endoscopy,* [12]*Deputy Director,*
National Cancer Center Hospital East, Kashiwa, Japan
[13]*Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science,*
Tokyo Women's Medical University, Tokyo, Japan

Full text of this paper is available at http://www.jstage.jst.go.jp/browse/dmpk

Summary: The ATP-binding cassette transporter, ABCG2, which is expressed at high levels in the intestine and liver, functions as an efflux transporter for many drugs, including clinically used anticancer agents such as topotecan and the active metabolite of irinotecan (SN-38). In this study, to elucidate the linkage disequilibrium (LD) profiles and haplotype structures of *ABCG2*, we have comprehensively searched for genetic variations in the putative promoter region, all the exons, and their flanking introns of *ABCG2* from 177 Japanese cancer patients treated with irinotecan. Forty-three genetic variations, including 11 novel ones, were found: 5 in the 5′-flanking region, 13 in the coding exons, and 25 in the introns. In addition to 9 previously reported nonsynonymous single nucleotide polymorphisms (SNPs), 2 novel nonsynonymous SNPs, 38C>T (Ser13Leu) and 1060G>A (Gly354Arg), were found with minor allele frequencies of 0.3%. Based on the LD profiles between the SNPs and the estimated past recombination events, the region analyzed was divided into three blocks (Block −1, 1, and 2), each of which spans at least 0.2 kb, 46 kb, and 13 kb and contains 2, 24, and 17 variations, respectively. The two, eight, and five common haplotypes detected in 10 or more patients accounted for most (>90%) of the haplotypes inferred in Block −1, Block 1, and Block 2, respectively. The SNP and haplotype distributions in Japanese were different from those reported previously in Caucasians. This study provides fundamental information for the pharmacogenetic studies investigating the relationship between the genetic variations in *ABCG2* and pharmacokinetic/pharmacodynamic parameters.

Key words: ***ABCG2*; nonsynonymous SNP; haplotype; haplotype-tagging SNP**

On August 8, 2005, the novel variations described in this paper were not found in the Japanese Single Nucleotide Polymorphisms (JSNP) (http://snp.ims.u-tokyo.ac.jp/), dbSNP in the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/SNP/) or PharmGKB (http://www.pharmgkb.org/do/) databases.

Introduction

The ATP-binding cassette transporter G2, ABCG2, also named breast cancer resistance protein (BCRP), placenta-specific ATP-binding cassette transporter

Received; August 15, 2005, Accepted; December 1, 2005
To whom correspondence should be addressed: Keiko MAEKAWA, Ph.D., *Division of Biochemistry and Immunochemistry, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Tel.* +81-3-3700-9453, *Fax* +81-3-5717-3832, *E-mail: maekawa@nihs.go.jp*

(ABCP) or mitoxantrone resistance protein (MXP), belongs to a superfamily (*white*) of ABC half transporters. ABCG2 was originally identified as a multi-drug resistance transporter because it was cloned from cancer cell lines selected with high concentrations of mitoxantrone[1] or verapamil and doxorubicin.[2] The *ABCG2* gene spanning over 66 kb, is located at 4q22 and consists of 16 exons, which encode a 72.6 kDa membrane protein of 655 amino acids.[2-4] Unlike many other drug-resistant transporters, such as ABCB1 (multidrug resistance protein 1 or P-glycoprotein) and ABCC1 (multidrug resistance-associated protein 1), ABCG2 has only one ATP binding region and six transmembrane domains, requiring dimerization or oligomerization to transport substrates.[5,6]

In normal human tissues, ABCG2 is expressed at high levels in placental trophoblast cells, the epithelium of the small intestine and colon, liver canalicular membranes, and the venous and capillary endothelium.[7] These distribution patterns suggest that ABCG2 may play a protective role against toxic substances and metabolites by extruding them across the apical membrane. *In vitro* studies have also indicated that a number of anticancer drugs are good substrates for ABCG2: e.g. topotecan, an irinotecan metabolite, 7-ethyl-10-hydroxycamptothecin (SN-38), and its glucuronide conjugate, SN-38G.[8-10] Indeed, inhibition of the murine ABCG2 homologue, Bcrp 1, increases the bioavailability of topotecan when orally administered to *mdr1a/1b*-deficient mice.[11] In a clinical study, coadministration of topotecan with GF120918, a dual inhibitor for ABCG2 and P-glycoprotein, was shown to markedly increase the bioavailability and systemic exposure of topotecan.[12]

The cloning of ABCG2 from drug-selected cell lines revealed that acquired amino acid substitutions at residue 482 (Arg482Gly and Arg482Thr) of ABCG2 resulted in marked alterations in substrate recognition and transport ability.[13] Thereafter, naturally occurring genetic variations in *ABCG2* have been extensively examined in various ethnic populations[14-21] because they were expected to explain interindividual differences in oral bioavailability and clearance of ABCG2 substrate drugs.[22] Two nonsynonymous polymorphisms, 34G>A (Val12Met) and 421C>A (Gln141Lys), were detected at relatively high frequencies in most ethnic groups including Caucasians, Asians, and Africans.[14,16,18,21,23] Both polymorphisms were reported to be associated with reduced protein expression *in vitro* and/or the increased sensitivity of the expressed cells toward several anticancer drugs although conflicting data were also reported.[16,24-26] The expression of ABCG2 protein in placenta was significantly lower in homozygotes with the 421A alleles than in those with the 421C alleles, while 34G>A (Val12Met) did not affect ABCG2 protein expression.[23] However, in intestinal samples, no association was found between the ABCG2 protein levels and the 421C>A (Gln141Lys) genotype.[18] A pharmacokinetic study showed that 421A (Gln141Lys) was unlikely to influence the *in vivo* disposition of irinotecan in European Caucasian cancer patients.[27] On the other hand, diflomotecan pharmacokinetics were significantly affected by the 421A genotype.[28] To explain these inconsistencies, the elucidation of the haplotype structure of *ABCG2* would be helpful; however, only limited information is available for the linkage disequilibrium (LD) profile and haplotype structure of this gene.[20,21] Also, to facilitate future pharmacogenetic studies on *ABCG2* genetic variations, haplotype analysis using its high-density SNPs found in a large number of samples is warranted.

In this study, we searched for genetic variations in *ABCG2* by sequencing 5′ regulatory regions and all the exons with their surrounding introns from 177 Japanese cancer patients administered irinotecan. Then, LD and haplotype analysis was performed using the detected variations, and haplotype-tagging single nucleotide polymorphisms (htSNPs) were identified to discriminate the common haplotypes in Japanese.

Materials and Methods

Human genomic DNA samples: All of the 177 patients participating in this study were administered irinotecan for the treatment of various cancers (mainly lung, stomach and colon cancers) at the National Cancer Center. Previously, we examined genetic variations in all *ABCG2* exons with their surrounding introns for 60 Japanese subjects treated with irinotecan.[17] In this study, we sequenced the additional 5′ regulatory regions in *ABCG2* for these patients and incorporated all the polymorphism data obtained from them into the haplotype analysis. Other 117 Japanese cancer patients were newly recruited. Genomic DNA was extracted from blood leukocytes from all the subjects and used as template in the polymerase chain reaction (PCR). The ethical review boards of the National Cancer Center and National Institute of Health Sciences approved this study. Written informed consent was obtained from all subjects.

PCR conditions for DNA sequencing: First, multiplex PCR was performed to amplify all 16 exons of *ABCG2* from 600 ng of genomic DNA utilizing 1.25 units of Ex-Taq (Takara Bio Inc., Shiga, Japan) with 0.30 μM each of the 32 primers designed in the intronic regions (listed in **Table 1**). The first round of PCR conditions were 94°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 1 min, and 72°C for 2 min, and then a final extension for 7 min at 72°C. Next, each PCR fragment was separately amplified in a second PCR using the first round PCR product as a template by

Table 1. Primer sequences used for the analysis of the *ABCG2* gene

Amplified and sequenced region		Forward primer		Reverse primer		PCR product (bp)
		Sequence (5′ to 3′)	Position[a]	Sequences (5′ to 3′)	Position[a]	
PCR	5′-flanking	TTAAGGGTCTTGAAACTGAC	13575766	CACTCTCAGCGAAACTGGTT	13574208	1559
	Exon 1	CTGTGCCCACTCAAAAGGTT	13574540	GAAACTGCGAAAGGCTAAAA	13574110	431
	Exon 2	GGATGTTCTTATCACAATGG	13556027	CAAATGAAAGCATGTGTCTG	13555509	519
	Exon 3	TGGTTTGTGCTTGTGTTCTA	13548617	GCTCAATAAATACCTGCTCG	13548337	281
	Exon 4	GGATTCAAAGTAGCCATGAGAT	13547885	GTCACATAATCAACTGGAAGCA	13547563	323
	Exon 5	GGCTTTGCAGACATCTATGGAG	13547246	CAGGTTAATTTCCACGTTCA	13546776	471
	Exon 6	ACCAGGTATCCACTTATTTG	13537791	TGACTTTCACTCCAACAGAA	13537366	426
	Exon 7	AAGACTGTCCTAGAATCTGC	13534384	TAGCACCAAATGGAACAAAC	13533817	568
	Exon 8	ATTACATGGGAAGAAGAGAG	13531063	TTGACTGGTATCAGAAGACTG	13530689	375
	Exon 9	TAGAATGAAGGTGTTAGGGA	13529516	AGGTGGAGTGAAGATAACAA	13529040	477
	Exon 10	GCCAAGCCATTGAGTGTTTA	13523258	CTGACTCATCCTACCCTCAA	13522908	351
	Exon 11	ACCAGAACAGTTTCCCTTTT	13517307	AAAAGTACTGGTAATCCTCCG	13516972	336
	Exon 12	TCATGGGATGCTTTCTCAGG	13515499	GTGTTTCCTTATCTCATGGT	13515056	444
	Exon 13	CATGGACAGACACAACATTG	13513633	GGCAAAGAGGAAAGTTAGTA	13513133	501
	Exon 14	ACCGTAAATGACTTCAGCTA	13511673	ATTCTCATTCCTTGCTCCTA	13511235	439
	Exon 15	TTGGTGAGACAAAGACTGTG	13510655	GCAGCAGAATACTGAGGGGT	13510242	414
	Exon 16	AGGCTTGGTTCAATTTTAGG	13508342	CATGATGTCTTGGGTTCTTT	13507897	446
Sequencing	5′-flanking[b]	GGTCTTGAAACTGACAGAAA	13575761	AATGGGTGGTTTCTGGTGAA	13575176	
		ATTCAGTAAGTTTCTCCCCT	13575291	CTGACACGAACTTCCTAAGC	13574712	
		TTAGGAAGTTCGTGTCAGCG	13574729	TAACCAAAGGCTCAGGATCT	13574265	
	Exon 1	GTGCCCACTCAAAAGGTTCC	13574538	CTAAAAAACTCAGTCGTCTCGT	13574124	
	Exon 2	GGATGTTCTTATCACAATGG	13556027	CAAATGAAAGCATGTGTCTG	13555509	
	Exon 3	GGTTTGTGCTTGTGTTCTAT	13548616	GCTCAATAAATACCTGCTCG	13548337	
	Exon 4	GGATTCAAAGTAGCCATGAGAT	13547885	GTCACATAATCAACTGGAAGCA	13547563	
	Exon 5	CTTTGCAGACATCTATGGAGT	13547244	GACCATACACATTACAGGAAAC	13546816	
	Exon 6	TGTCTCTTACAGGACTGGCA	13537738	TGACTTTCACTCCAACAGAA	13537366	
	Exon 7	GACAAAGTCAGGCTGAACTA	13534231	CTACCCAAAGACCAAACAGC	13533869	
	Exon 8	TCTGTCTTCTCTAGCCTTACC	13531015	ACAGAAATTCACAAAGCCAC	13530715	
	Exon 9	CATCCAAGAAAGGGTTCACA	13529494	AAGGGTGGGTAGAAGATAAA	13529068	
	Exon 10	CAAGCCATTGAGTGTTTATC	13523256	GAATGACATTTACACTATACTTG	13522933	
	Exon 11	CCCTTTTTTTCCTGCTTAAC	13517294	AACCCCAGATGTAATCAGTC	13517011	
	Exon 12	CTGTCAGCAGAGGTCTGTAAC	13515397	TTCCTTATCTCATGGTTTGG	13515060	
	Exon 13	GGACAGACACAACATTGGAG	13513630	AAGTAAAGCAGAGCCCCATT	13513189	
	Exon 14	TGCAGAGGAGAAGAGTTTAG	13511637	TTCTGGATGGGAGACTTTCA	13511264	
	Exon 15	AGACAAAGACTGTGAATATGTT	13510649	AGCAGAATACTGAGGGGTTG	13510244	
	Exon 16	GCTTGGTTCAATTTTAGGCT	13508340	GATGGCAAGGGAACAGAAAA	13507989	

[a]The nucleotide position of the 5′ end of each primer on NT_016354.17.
[b]Three sets of overlapping primers were used for sequencing of 5′-flanking region.

Table 2. Summary of *ABCG2* SNPs detected in a Japanese population

SNP ID				Location	Position		Nucleotide change and flanking sequences (5′ to 3′)	Amino acid change	Number of subjects			Allele Frequency
This Study	JSNP	dbSNP (NCBI)	Reference		NT_016354.17 (or AC084732.1)	From the translational initiation site or from the end of nearest exon			Wild-type	Hetero-zygote	Homo-zygote	
MPJ6_AG2027[a]				5′-Flanking	19496[b]	−1412[b,c]	TCTTTTTAAAAAA/GTTATGTATAGTC		176	1	0	0.003
MPJ6_AG2028	ssj0001952	rs3219191		5′-Flanking	19705_19708[b]	−1203_−1200[b,c]	TGGCTACACTCACTCA/-CAAAGCCTGATG		103	64	10	0.237
MPJ6_AG2029[a]				5′-Flanking	20155[b]	−753[b,c]	GGCGGGAGTGTTT/CGGCTTGTCCCTG		176	1	0	0.003
MPJ6_AG2030[a]				5′-Flanking	20246[b]	−662[b,c]	CGCGGCAGGACAC/TGTGTGCGCTTTC		175	2	0	0.006
MPJ6_AG2001		rs2231136		5′-Flanking	20646[b]	−262[b,c]	CAGTGCCCGCCAC/TTCCCACTGAGAT		173	4	0	0.011
MPJ6_AG2031[a]				Intron 1	13574206	IVS1+37	CGCTGAGAGTGTG/TTGTCTGGAAAGG		176	1	0	0.003
MPJ6_AG2032[a]				Intron 1	13555999	IVS1−128	ATGGTATGGGCCA/GTTCATTGGAAAT		176	1	0	0.003
MPJ6_AG2033[a]				Intron 1	13555999	IVS1−128	ATGGTATGGGCCA/TTTCATTGGAAAT		176	1	0	0.003
MPJ6_AG2002	IMS-JST131649, ssj0001922	rs1564481		Intron 1	13555970	IVS1−99	CTGCTCATTGCCG/ACACATTTAAAAA		70	87	20	0.359
MPJ6_AG2003	ssj0001924	rs2231137		Exon 2	13555819	34	GTTTTTATCCCAG/ATGTCACAAGGAA	Val12Met	111	64	2	0.192
MPJ6_AG2034[a]				Exon 2	13555815	38	TTATCCCAGTGTC/TACAAGGAAACAC	Ser13Leu	176	1	0	0.003
MPJ6_AG2004	ssj0001925	rs4148152		Intron 2	13555614	IVS2+36	GACAGCTTTTTAA/GTTTACCTACAGT		111	64	2	0.192
MPJ6_AG2005			17)	Intron 2	13548585	IVS2−93	TTTGTAATTTCAT/CCAACCTTCATTG		167	10	0	0.028
MPJ6_AG2006	ssj0001929	rs2231138		Intron 3	13548423	IVS3+10	TTCGTGAGTATAA/GGAGAGTATAAGT		153	24	0	0.068
MPJ6_AG2007		rs5860119	17,23)	Intron 3	13548361_13548362	IVS3+71_+72	CCACTTTTTTTT-/TGTGTGCGAGCAG		174	3	0	0.008
MPJ6_AG2008		rs2231139		Exon 4	13547669	369	TAATTCAGGTTAC/TGTGGTACAAGTA	Tyr123Tyr	176	1	0	0.003
MPJ6_AG2009			16,23)	Exon 4	13547662	376	GGTTACGTGGTAC/TAAGTAAGTATTA	Gln126X	167	10	0	0.028
MPJ6_AG2010	IMS-JST080116, ssj0001930	rs2231142		Exon 5	13547028	421	AGAGAAAACTTAC/AAGTTCTCAGCAG	Gln141Lys	79	83	15	0.319
MPJ6_AG2035			21)	Exon 5	13546970	479	AAAAAAACGAACG/AGATTAACAGGGT	Arg160Gln	176	1	0	0.003
MPJ6_AG2036		rs2231144		Intron 5	13537665	IVS5−16	ATTGACAATTATA/GATATTTGTGATT		176	1	0	0.003
MPJ6_AG2011	ssj0001931	rs1871744		Intron 6	13534334	IVS6−217	AGTTGATTCTAGA/GTTGTCAATACAA		102	63	12	0.246
MPJ6_AG2012			17)	Intron 6	13534321	IVS6−204	TTGTCAATACAAC/TACACTGAAAATT		166	11	0	0.031
MPJ6_AG2013	ssj0001932	rs2231145		Intron 6	13534289	IVS6−172	TGTTTTAATAAAA/GGCATTGAATTTA		133	43	1	0.127
MPJ6_AG2014	ssj0001933	rs2231146		Intron 6	13534205	IVS6−88	GAACTAGAGCAAA/GCAATCTAAAGGC		133	43	1	0.127
MPJ6_AG2037[a]				Exon 9	13529294	1060	CATCAACTTTCCG/AGGGGTGAGAAGA	Gly354Arg	176	1	0	0.003
MPJ6_AG2015			17,23)	Exon 9	13529256	1098	AGTCTTCAAGGAG/AATCAGCTACACC	Glu366Glu	172	5	0	0.014
MPJ6_AG2016	ssj0001936	rs2231148		Intron 9	13523183	IVS9−60	TAAATGGTGTGTA/TTAAGTTTTTATC		124	46	7	0.169
MPJ6_AG2038[a]				intron10	13522946	IVS10+95	AGACAAGTATAGT/AGTAAATGTCATT		175	2	0	0.006
MPJ6_AG2017			17)	Exon 11	13517163	1291	GCTGGGGTTCTCT/CTCTTCCTGACGA	Phe431Leu	176	1	0	0.003
MPJ6_AG2039			23,24)	Exon 11	13517132	1322	AGTGTTTCAGCAG/ATGTTTCAGCCGT	Ser441Asn	176	1	0	0.003
MPJ6_AG2018	ssj0001939	rs2231153		Intron 11	13517067	IVS11+20	GGTCTTTGTTCTA/GGGAACGGGGCTG		124	49	4	0.161
MPJ6_AG2019			17)	Intron 11	13515440	IVS11−135	CATGCATAGTGGG/ATCTAGCCCTGAG		171	6	0	0.017
MPJ6_AG2020			17,23)	Exon 12	13515208	1465	CCAAGTATTATAT/CTTACCTGTATAG	Phe489Leu	174	3	0	0.008
MPJ6_AG2021	ssj0001941	rs2231156		Intron 12	13515132	IVS12+49	AGTCTTGCCTATG/TGGTGAAGTCAGT		95	71	11	0.263
MPJ6_AG2040			23)	Exon 13	13513442	1515	AAAGGCAGATGCC/-TTCTTCGTTATG	Phe506SerfsX4	176	1	0	0.003
MPJ6_AG2022	IMS-JST064085, ssj0001943	rs2231157		Intron 13	13513270	IVS13+40	AAGGAATTGTTTC/TTTTCCTTCATTT		28	85	64	0.602
MPJ6_AG2023			17)	Intron 13	13513245	IVS13+65	CTTCCTGCACATT/GCACTTGTCATGT		175	2	0	0.006
MPJ6_AG2041[a]				Intron 13	13511651	IVS13−185	ACTTCAGCTATAT/CAGAGATGTGCAC		176	1	0	0.003
MPJ6_AG2024	IMS-JST064086, ssj0001944	rs2231162		Intron 13	13511487	IVS13−21	GCAGGCCTGACTC/TTTAGTATTTGCT		124	49	4	0.161
MPJ6_AG2042			21)	Exon 14	13511391	1723	TTCAGCATTCCAC/TGATATGGATTTA	Arg575X	176	1	0	0.003
MPJ6_AG2025	ssj0001946	rs2231164		Intron 14	13510562	IVS14−46	TGGAAACTTCTTA/GAAATTTAAAACT		58	86	33	0.429
MPJ6_AG2026	IMS-JST064087, ssj0001947	rs2231165		Intron 15	13510324	IVS15+110	ACTGAATTTTTCC/TGAGCCTACGTTT		136	39	2	0.121
MPJ6_AG2043a				Intron 15	13510292	IVS15+142	CATAAAGTGAGTT/GGTGCGCGAATGC		175	2	0	0.006

[a]Novel variations detected in this study.

[b]For the 5′-flanking variations, the genomic sequence, AC084732.1, was used as the reference sequence.

[c]Intron 1 was skipped for numbering.

Table 3. Haplotypes of Block -1 in a Japanese population

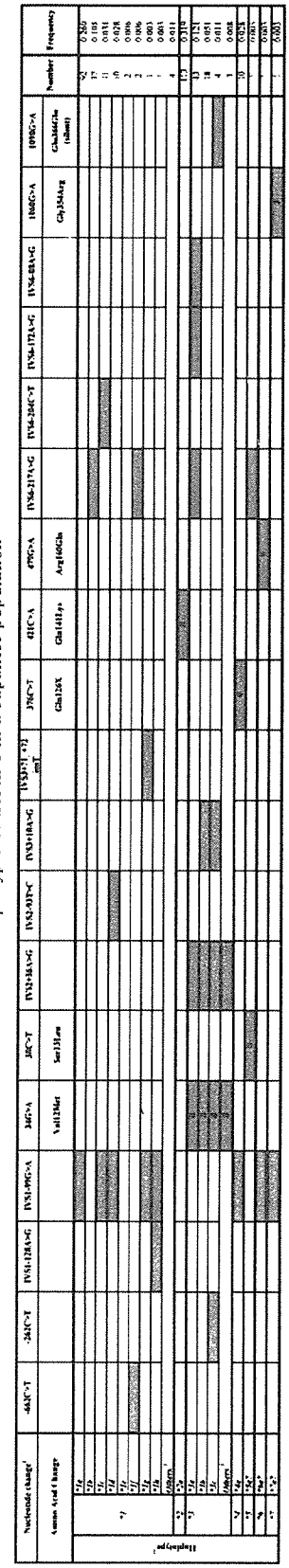
Nucleotide change[1]		-1412A>G	-1203_-1200 delCTCA	Number	Frequency
Amino Acid Change					
Haplotype[2]	*1 *1a			270	0.763
	*1b			83	0.234
	*1c?			1	0.003

[1]The position on the cDNA sequence (A of the translational start codon is 1, based on AC084732.1. Intron 1 was skipped for numbering the nucleotide positions.)
[2]The haplotype, *1c, ambiguously inferred in only one patient is described with a question mark.
White cell, wild-type; gray cell, nucleotide alteration.

Ex-Taq (0.625 units) with the same primers and conditions used in the first round PCR. Because of the high GC content in the 5′-flanking region of *ABCG2*, this region was separately amplified by a single PCR reaction from 50 ng of genomic DNA using 0.625 units of Ex-Taq and GC buffer I (Takara Bio Inc.) with 0.2 μM primers (listed in **Table 1**). The PCR conditions for amplifying the 5′-flanking region were the same as described above. The PCR products were then treated with a PCR Product Pre-Sequencing Kit (USB Co., Cleveland, OH, USA) and were directly sequenced on both strands using an ABI BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) with the sequencing primers listed in **Table 1** (Sequencing column). The excess dye was removed by a DyeEx96 kit (Qiagen, Hilden, Germany). The eluates were analyzed on an ABI Prism 3700 DNA Analyzer (Applied Biosystems). All the novel SNPs were confirmed by repeat sequencing of PCR products that were newly generated by amplification of genomic DNA. The genomic and cDNA sequences of *ABCG2* obtained from GenBank (NT_016354.17 and NM_004827.1, respectively) were used as the reference sequences for all the analyzed regions except for the 5′-flanking region. Unless otherwise indicated, the description of the nucleotide position was based on cDNA sequence, and adenine of the translational initiation site in exon 2 was numbered as +1. For the 5′-flanking variations (upstream of exon 1), the genomic sequence, AC084732.1, was used as the reference sequence, and intron 1 was skipped for numbering the nucleotide positions. For intronic polymorphisms, the position was numbered from the nearest exon.

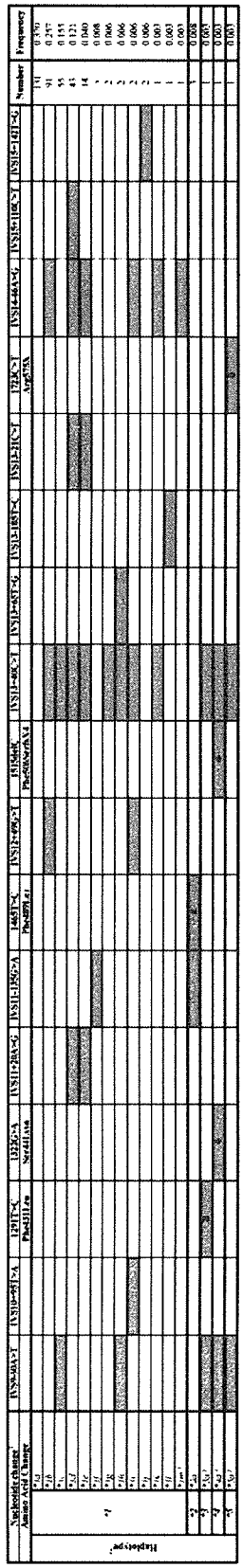
Linkage disequilibrium (LD) and haplotype analyses: Hardy-Weinberg equilibrium and LD analysis were performed by SNPAlyze software (Dynacom Co., Yokohama, Japan) and a pairwise LD between variations was obtained for the |D′| and rho square (r^2) values. Some of the haplotypes were unambiguous from subjects with homozygous variations at all sites or a heterozygous variation at only one site. The diplotype

Table 4. Haplotypes of Block 1 in a Japanese population



[1]The position on the cDNA sequence (A of the translational start codon is 1, based on NT_016354.17) or position from the nearest exon.
[2]For the 5′-flanking variations, the genomic sequence, AC084732.1, was used as the reference sequence, and intron 1 was skipped for numbering the nucleotide positions.
[3]The haplotypes inferred in only one patient are classified into "Others" (in the *1 and *3 groups) or indicated with a question mark (*5a, *6a and *7a).
[4]"Others" include haplotypes harboring −753T>C, IVS1+37G>T, IVS1−128A>T, IVS3+71_+72insT, 369C>T (Tyr123Tyr), IVS5−16A>G or 1098G>A (Glu366Glu).
White cell, wild-type; gray cell, nucleotide alteration.

Table 5. Haplotypes of Block 2 in a Japanese population



[1]The position on the cDNA sequence (A of the translational start codon is 1, based on NT_016354.17) or position from the nearest exon.
[2]The haplotypes inferred in only one patient are indicated with a question mark (**1m*, **3a*, **4a* and **5a*).
White cell, wild-type; gray cell, nucleotide alteration.

configurations were inferred by LDSUPPORT software, which determines the posterior probability distribution of the diplotype for each subject based on the estimated haplotype frequencies.[29] The haplotypes without any amino acid change were designated as **1*, and the nonsynonymous SNP-bearing haplotypes were numerically numbered. The subtypes were named by use of alphabetical small letters in the order of their frequencies. The haplotypes ambiguously inferred in only one patient were placed into an "Others" category within each group, or are shown as a haplotype name plus a question mark in **Tables 3–5**.

Network analysis of the inferred haplotypes was performed with Network 4.1.0.9 (http://www.fluxus-engineering.com.) based on the algorithms of the reduced median network.

Results

ABCG2 variations found in a Japanese population: We found 43 variations, including 11 novel ones, from 177 Japanese cancer patients treated with irinotecan (**Table 2**). Five variations were located in the 5′-flanking region, 13 (2 synonymous and 11 nonsynonymous) in the coding exons, and 25 in the introns. The observed allele frequencies were all in Hardy-Weinberg equilibrium ($p > 0.05$) except for 34G>A (Val12Met) and IVS2 +36A>G ($p = 0.028$ for both variations). The deviation from Hardy-Weinberg equilibrium for both variations resulted from a slightly smaller number of homozygotes than expected. Currently, the reason for this deviation is not clear.

Two novel nonsynonymous SNPs, 38C>T (Ser13Leu) and 1060G>A (Gly354Arg), were heterozygous in different patients at an allele frequency of 0.003. Both Ser13 and Gln354 are in the intracellular loop of the N-terminus. Of the other novel variations, three (−1412A>G, −753T>C, and −662C>T) were found in the 5′ flanking region and six (IVS1+37G>T, IVS1-128A>G, IVS1-128A>T, IVS10+95T>A, IVS13-185T>C, and IVS15+142T>G) in the introns.

Thirty-two variations detected were already reported or publicized in the dbSNP database and/or the JSNP database. Their database ID numbers are also listed in **Table 2**. For these known polymorphisms, similar allele frequencies were obtained as described in previous reports[16,23] and the JSNPs database for Japanese populations. However, marked ethnic differences in the allele frequencies were observed with −1203 −1200delCTCA, 34G>A (Val12Met), 421C>A (Gln141Lys), and some intronic variations. The deletion of four nucleotides, −1203 −1200delCTCA, is located upstream of the basal promoter and identified at a much lower frequency in our population (0.237) than in 60 Swedish individuals (0.540).[19] This deletion is within the region characterized as a positive regulatory element

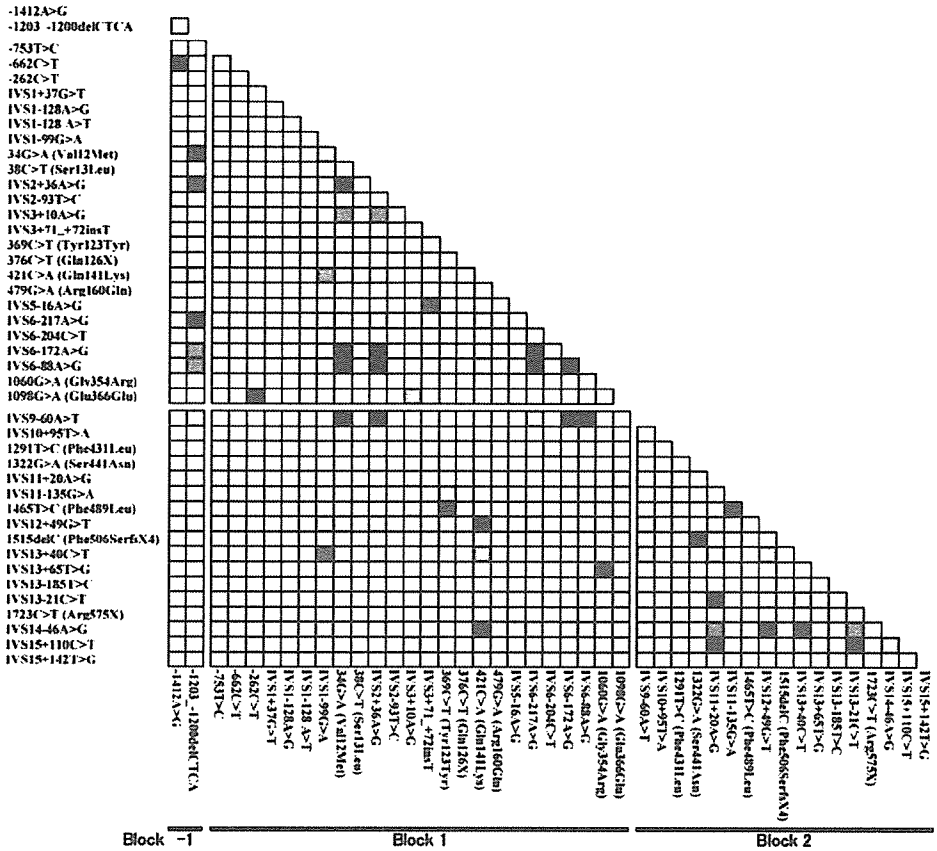


Fig. 1. Linkage disequilibrium (LD) analysis of *ABCG2* by r^2 statistics. Pairwise LD between all the detected SNPs is depicted as r^2 (from 0 to 1) by a 10-graded gray color. The denser color represents a higher linkage.

(from −1836 to −1175 bp upstream of the translational start site) of the basal promoter,[30] and was suggested to influence irinotecan pharmacokinetics.[31]

The frequencies of two well-known nonsynonymous SNPs, 34G>A (Val12Met) and 421C>A (Gln141Lys), were 0.192 and 0.319 in our study, which were comparable to those in Chinese (0.204 and 0.222–0.350, respectively).[20,27] However, the frequencies were much higher than those in Caucasians (0.02–0.065 and 0.08–0.15), African-Americans (0–0.09 and 0–0.05), and a Swedish population (0.02 and 0.1).[18,19,21,23,27] Of other relatively rare nonsynonymous SNPs, 376C>T (Gln126X), 1291T>C (Phe431Leu), 1322G>A (Ser441Asn), 1465T>C (Phe489Leu), and 1515delC (Phe506SerfsX4) were already detected in a Japanese population by Itoda *et al.*[17] and/or Kobayashi *et al.*,[23] but not found in other ethnic groups. Recently, the two nonsynonymous SNPs, 479G>A (Arg160Gln) and 1723C>T (Arg575X), have been reported by Bosch *et al.*[21] Arg160Gln and Arg575X were found as heterozygotes in single Asian and Caucasian subjects, respectively. Therefore, the distribution of 479G>A (Arg160Gln) might be restricted in Asians, while 1723C>T (Arg575X) is likely to be detected both in Caucasians and Japanese at low frequencies. Arg575X results in a truncated protein that terminates at the extracellular loop between transmembrane domains 5 and 6 and lacks transmembrane domain 6.

On the other hand, several nonsynonymous SNPs reported in other ethnic groups were not detected: 805C>T (Pro269Ser) found in Chinese at a 0.037 frequency,[20] 1858G>A (Asp620Asn) in undefined (combined) ethnicities[14] (0.011) and in a Dutch population[21] (0.005), 616A>C (Ile206Leu) in Hispanics (0.100), and 1768A>T (Asn590Tyr) in Caucasians (0.010).[18] Thus, these SNPs are either ethnic-specific or rare.

The ethnic differences in the allele frequencies were also observed with intronic variations. Bosch *et al.* searched for variation in *ABCG2* in 100 healthy Dutch volunteers that consisted mainly of Caucasians.[21] The reported allele frequencies of IVS11+20A>G (0.045), IVS12+49G>T (0.095), IVS13-21C>T (0.035), and IVS14-46A>G (0.160) in Caucasians were lower than those in Japanese (0.161, 0.263, 0.161, and 0.429, respectively). On the other hand, IVS9-60A>T was detected at a higher frequency in Caucasians (0.360)

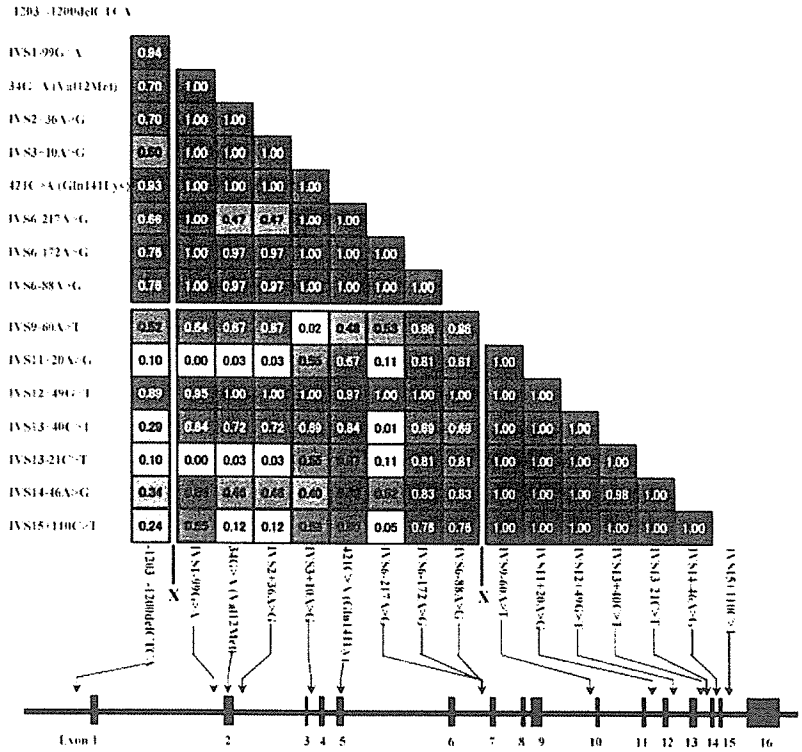


Fig. 2. Linkage disequilibrium (LD) analysis of *ABCG2* by |D′| statistics. The values of |D′| between 16 common SNPs (≥0.05 in their allele frequencies) are shown. The positions of the estimated minimum recombination events are indicated by a "X", and the location of the variations is indicated in the schematic diagram of the *ABCG2* gene.

than in Japanese (0.169), while IVS13+40C>T had a comparable frequency between Caucasians (0.535) and Japanese (0.602).

Linkage disequilibrium (LD) analysis and haplotype block partition: Using the 43 detected genetic variations, LD analysis was performed by |D′| and r^2 statistics, and the pairwise values of r^2 are depicted with 10-graded blue color in **Fig. 1**. As previously reported in various ethnic groups,[14,18] perfect LD ($r^2=1.0$) was observed between 34G>A (Val12Met) and IVS2+36A>G. In addition, two pairs of SNPs, between IVS6-172A>G and IVS6-88A>G and between IVS11+20A>G and IVS13-21C>T, were perfectly linked ($r^2=1.0$). Because the two rare nonsynonymous variations, 1515delC (F506SfsX4) and 1322G>A (Ser441Asn), were found in the same patient, they were statistically estimated to be linked with each other. Strong LDs ($r^2>0.7$) were observed between −262C>T and 1098G>A (Glu366Glu) ($r^2=0.798$), between 421C>A (Gln141Lys) and IVS12+49G>T ($r^2=0.709$), and among IVS11+20A>G, IVS13-21C>T, and IVS15+110C>T ($r^2>0.720$).

As for |D′| values, when considering 16 common SNPs (≥0.05 in their allele frequencies), only 48 pairs (40%) out of 120 pairs gave |D′|=1.0 (**Fig. 2**), indicating that recombination had occurred within this gene. To further estimate the past recombination events, the absence or presence of all four gametes between these 120 site pairs was assessed (data not shown).[32] The results obtained implied that recombination had happened in at least two sites: between −1203 −1200delCTCA and IVS1-99G>A, and between IVS6-88A>G and IVS9-60A>T (**Fig. 2**). Therefore, the *ABCG2* gene was divided into three blocks (Block −1, Block 1, and Block 2) as indicated in **Fig. 1**. Block −1 included two variations, −1412A>G and −1203 −1200delCTCA, in the 5′ regulatory region and might have its 5′ boundary upstream of the analyzed region. Block 1, spanning 46 kb, included 24 variations from −753T>C in the 5′ regulatory region to 1098G>A (Glu366Glu) in exon 9. The 5′ boundary of Block 1 was tentatively assigned because of the very low frequencies of −753T>C and −662C>T. The 3′ boundary of Block 1 was assigned by considering the strong association between −262C>T and 1098G>A (Glu366Glu) described above. Block 2, which included the remaining 17 variations, ranged from intron 9 to intron 15 (13 kb). No close associations were observed ($r^2<0.70$) between the variations across the blocks except for one pair of SNPs, 421C>A (Gln141Lys) and IVS12+49G>T,