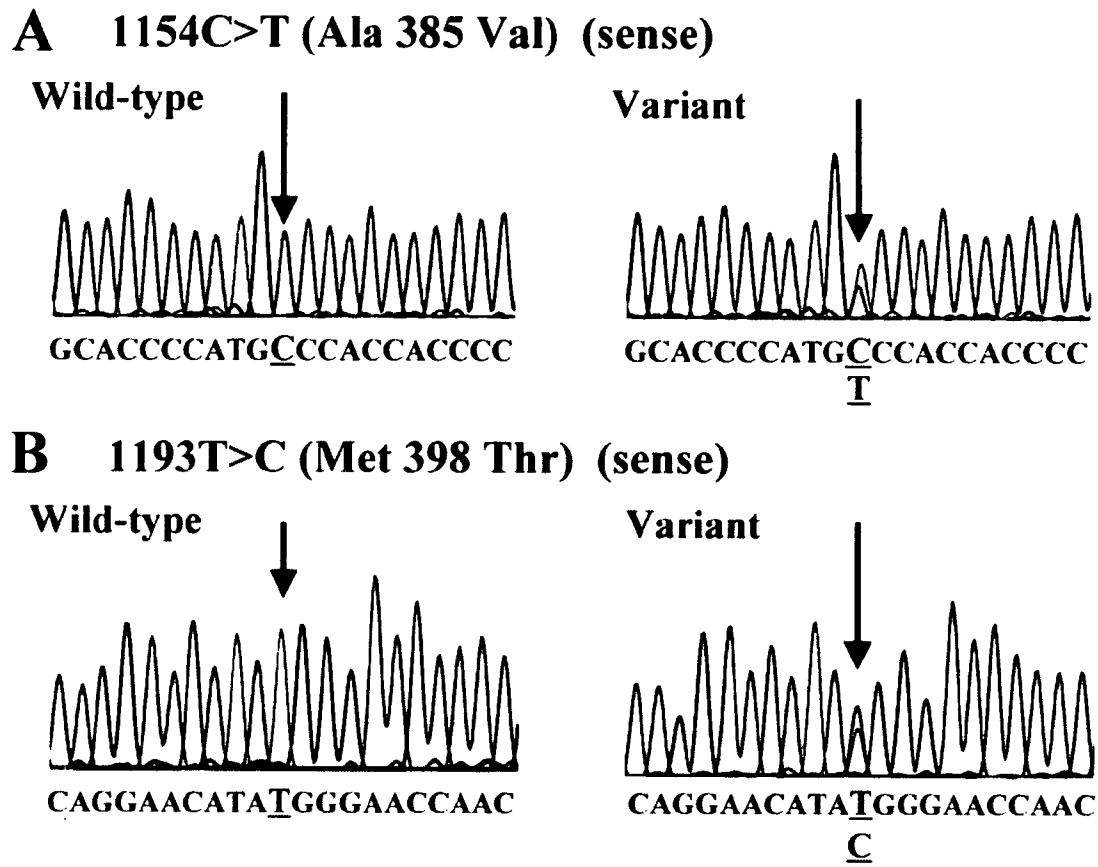

Fig. 1. Two novel nonsynonymous variations of human *HNF4A*.
(A) MPJ6_H4A_024 (wild-type, 1154C/C; variant, 1154C/T). (B) MPJ6_H4A_025 (wild-type, 1193T/T; variant, 1193T/C). Arrows indicate the positions of the nucleotide changes.

haplotype bearing 1154C>T (A385V) and 1193T>C (M398T) were confirmed as follows: the PCR fragment amplified by a high fidelity DNA polymerase KOD-Plus- (TOYOBO, Tokyo, Japan) with the primer pairs (5′-CACCGCACCTTGTTCCTTTCAACT-3′ as a forward primer and 5′-TGCCCTTTATTCCCTACCCT-3′ as a reverse primer) from the genomic DNA was cloned into pcDNA3.1-TOPO vector (Invitrogen, Carlsbad, CA, USA); and the cloned inserts (9 clones, *ca.*, 500 ng) were directly sequenced on both strands as described in the above section.

Results and Discussion

The P1 promoter regions (up to 300 bases upstream of the translational start site), all 10 exons (exon 1A and exons 2–10) and their flanking introns of *HNF4A* were sequenced in 74 Japanese type II diabetic patients. Genbank accession number NT_011362.9 was utilized for the reference sequence. The cDNA and amino acid numberings were based on isoform 2 of HNF4α (accession numbers NM_000457 and NP_000448, respectively, based on Drewes *et al.*[1]). Thirty-nine genetic variations, including 16 novel ones [1 in the promoter region, 2 in the coding exons, 5 in the 3′-untranslated region (3′-UTR), and 8 in the introns], were detected (see **Table 2**). All of the detected variations were found in Hardy-Weinberg equilibrium ($p>0.05$), except for IVS5−136T>C ($p=0.02$) because of a slightly low occurrence of homozygote compared with that of expectation.

Both of the two novel nonsynonymous variations, 1154C>T (A385V) and 1193T>C (M398T), were simultaneously found in a same patient as heterozygotes at 0.007 frequencies (**Fig. 1**), and their linkage was confirmed by cloning of the genomic DNA fragment and its sequencing analysis (data not shown). These variations are located in the F-domain of HNF4α, which has been shown to modulate the transactivation potential of AF-2.[9] According to PolyPhen, a prediction tool for the possible impact of amino acid substitutions (http://genetics.bwh.harvard.edu/pph/), it is possible that M398T (but not A385V) damages the function or structure of this protein. Moreover, a mutation located four amino acids downstream of

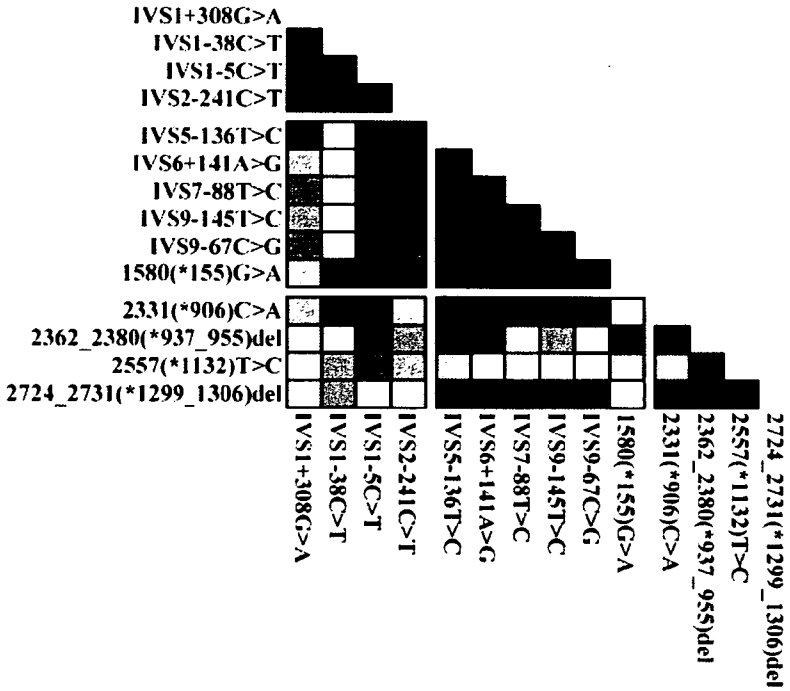

Fig. 2. Linkage disequilibrium (LD) analysis of *HNF4A* by |D'| statistics.
Pairwise LD (|D'|) between 14 common SNPs (>0.05 in their allele frequencies) are shown by a 10-graded gray color. The denser color represents the higher linkage.

M398 (V402I mutation, V393I in the original paper) results in reduced transactivation activity.[15] Further functional analysis should be pursued for these variations.

HNF4α is also known as maturity-onset diabetes of young (MODY) gene, MODY-1. Many genetic variations have been found in MODY patients, and several in the Japanese. For example, R136W (R127W in the original paper), one of the mutations found in early (<25-years-old)-onset type II diabetic patients, was shown to decrease the DNA binding ability and transcriptional activity.[16,17] As for the newly identified nonsynonymous variations 1154C>T (A385V) and 1193T>C (M398T), the association with MODY is unlikely since the variations were found in a patient who was diagnosed with type II diabetes around 60 years of age.

The known nonsynonymous variation, 416C>T (T139I, T130I in the original paper), was shown to be associated with reduced transcriptional activity and type II diabetes in Japanese and Danish white subjects.[18,19] This variation was detected at a 0.007 frequency in our diabetic patients, which is lower than the reported frequency in Japanese diabetic patients (0.035 in 423 patients) but similar to that in the nondiabetic subjects (0.008 in 354 subjects).[18] The reason for this discrepancy is currently unknown, but this might be due to the small number of patients analyzed in the present study and relatively low frequency of this variation. The patient with this T139I heterozygous variation showed a good response to glimepiride. Her HbA1c value decreased from 8.3 to 6.9 by treatment with glimepiride (1 mg/day) for four months. It is possible that the reduced transcriptional activity of HNF4α from this variation might lead to reduced CYP2C9 levels, resulting in increased glimepiride bioavailability. The frequency of intronic variation IVS1−38C>T (0.196) was slightly lower than those in the previous report on Japanese early-onset type II diabetic patients (0.28) and nondiabetic subjects (0.24).[16] The frequency of IVS1−5C>T (0.223) was slightly higher than those (0.14 and 0.15) described in the previous reports for Japanese and Chinese nondiabetic subjects, respectively.[16,20,21] Instead, our frequency for IVS1−5C>T was similar to those of Japanese and Chinese type-II diabetic patients (0.216 and 0.24, respectively).[20,21]

Using the detected variations, linkage disequilibrium (LD) was first analyzed using r^2 values (data not shown). Although found only in two subjects, perfect linkages were observed among IVS3−204C>G, IVS4+140C>G, IVS4−197A>C, IVS4−96C>G, IVS4−52G>A, IVS6+196G>A, IVS9−151A>C, and 2331C>T ($r^2=1$). Because the novel nonsynonymous variations 1154C>T (A385V) and 1193T>C (M398T)

Table 1. Primer sequences used for the analysis of *HNF4A*

		Amplified or sequenced region	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Amplified length (bp)
1st PCR	Mix 1	Exon 1	GTCAAGGGTCAAATGAGTGC	CCTTGCCGTCTCTCTGAACC	902
		Exon 4	AGCCGAGTCTTCACTGTCTT	GAAGGTGAAGACTCTGCTTG	626
		Exon 6	CACACAGAATGTTGCTTACA	TCGTGCTCTGACTTCAATGC	633
		Exon 8	TCTTTTTCTGCCTGTGTCTA	ACTGAGGCACAGACAGGTTA	571
	Mix 2	Exon 2 to 3	TACAGATGTGAAACTGAAGC	CTCTTCTCAGCCATTAGCCA	1,733
		Exon 5	CAGACTCCTTGGGGCTCTAA	CACCTATGAGATGGCAGTAA	654
		Exon 7	GATGCTTTGGTGCCTATCAG	CTTGACTTGCCTCATCTGTT	617
		Exon 9	GCACCTTGTTCCTTTCAACT	TGCCCTTTATTCCCTACCCT	529
		Exon 10	GACCTTCCAGACCTCATAAA	GGGTCTAATGCTTCCAGAAT	2,307
2nd PCR		Exon 1	GTCAAGGGTCAAATGAGTGC	CCTTGCCGTCTCTCTGAACC	902
		Exon 2	TACAGATGTGAAACTGAAGC	AAGACTTAGTATTGTGCCTG	606
		Exon 3	TACTCCAGAGGTCAAGGTTC	CTCTTCTCAGCCATTAGCCA	522
		Exon 4	AGCCGAGTCTTCACTGTCTT	GAAGGTGAAGACTCTGCTTG	626
		Exon 5	CAGACTCCTTGGGGCTCTAA	CACCTATGAGATGGCAGTAA	654
		Exon 6	CACACAGAATGTTGCTTACA	TCGTGCTCTGACTTCAATGC	633
		Exon 7	GATGCTTTGGTGCCTATCAG	CTTGACTTGCCTCATCTGTT	617
		Exon 8	TCTTTTTCTGCCTGTGTCTA	ACTGAGGCACAGACAGGTTA	571
		Exon 9	GCACCTTGTTCCTTTCAACT	TGCCCTTTATTCCCTACCCT	529
		Exon 10	GACCTTCCAGACCTCATAAA	TGGAGGAGAAAGGCGTCTTC	1,043
			GAAACGATTCCCCCAGTCAT	GGGTCTAATGCTTCCAGAAT	1,372
Sequencing		Exon 1	CTGAACATCGGTGAGTTAGG	CCTTGCCGTCTCTCTGAACC	
		Exon 2	TACAGATGTGAAACTGAAGC	AAGACTTAGTATTGTGCCTG	
		Exon 3	TACTCCAGAGGTCAAGGTTC	CTCTTCTCAGCCATTAGCCA	
		Exon 4	AGCCGAGTCTTCACTGTCTT	GAAGGTGAAGACTCTGCTTG	
		Exon 5	CAGACTCCTTGGGGCTCTAA	CACCTATGAGATGGCAGTAA	
		Exon 6	CACACAGAATGTTGCTTACA	TCGTGCTCTGACTTCAATGC	
		Exon 7	GATGCTTTGGTGCCTATCAG	CTTGACTTGCCTCATCTGTT	
		Exon 8	TCTTTTTCTGCCTGTGTCTA	ACTGAGGCACAGACAGGTTA	
		Exon 9	GCACCTTGTTCCTTTCAACT	TGCCCTTTATTCCCTACCCT	
		Exon 10	GACCTTCCAGACCTCATAAA	TATCCAGAGCAGGGCGTCAA	
			AGAGCAGGAATGGGAAGGAT	TGGAGGAGAAAGGCGTCTTC	
			GAAACGATTCCCCCAGTCAT	AAGACAGTGCCTGGGAGTAA	
			CACATCAGAGTGACATCCAG	GGGTCTAATGCTTCCAGAAT	
			CCTAGAGATTGTTTTTGTTT		

were found in the same patient, they were statistically estimated to be linked with each other, and this was confirmed by cloning and sequencing analysis as described above. By the same reason, the pairs, −208G>C and IVS5+173_176delTTAG, and IVS3−54delC and IVS8−106A>G, might be linked closely. Relatively strong LD ($r^2 \geqq 0.5$) was observed between IVS1−5C>T and IVS2−241C>T, and among IVS5−136T>C, IVS6+141A>G, IVS7−88T>C, IVS9−145T>C, and IVS9−67C>G. The r^2 values were below 0.5 for the other pairs of variations.

We found that haplotypes without block partitioning were too diverse. Thus, block partitioning for haplotype analysis was performed based on |D′| values using the 14 common variations detected with a frequency greater than 0.05 (**Fig. 2**). All the |D′| values were 1.0 between six pairs of the first four variations from IVS1+308G>A to IVS2−241C>T. Among the next six variations from IVS5−136T>C to 1580(*155)G>A, 14 out of 15 pairs (93%) had |D′| values over 0.9. If the 2331(*906)C>A was included, the percentage of the pairs over 0.9 was reduced to 76% (16/21). Among the last 4 variations [from 2331(*906)C>A to 2724_2731(*1299_1306)delTCCTCCCT], 4 out of 6 pairs (67%) had |D′| values over 0.9. If the 1580(*155)G>A was included, the percentage was reduced to 50% (5/10). Thus, the *HNF4A* haplotypes were analyzed as three blocks divided between IVS2−241C>T and IVS5−136T>C and between 1580(*155)G>A and 2331(*906)C>A. The boundary of block 1 and block 2 for the minor variations was tentatively assigned between IVS3+50C>T and IVS3−204C>G because of the perfect linkage ($r^2=1$) of IVS3−204C>G with IVS6+196G>A and a long distance between exons 3 and 4 (6.2 kb). The boundary of block 2 and block 3 for the minor variations was also tentatively assigned between 1817(*392)T>G and 1852(*427)G>T because of a moderate linkage ($r^2=0.36$) between 1852(*427)G>T and 2724_2731(*1299_1306)delTCCTCCCT. The block partitions were

Table 2. Summary of *HNF4A* variations detected in this study

SNP ID			Reference	Location	Position		Nucleotide change and flanking sequences (5′ to 3′)	Amino acid change	Number of subjects			Frequency
This Study	JSNP	dbSNP (NCBI)			NT_011362.9	From the translational initiation site or from the end of the nearest exon[b]			Wild-type	Hetero-zygote	Homo-zygote	
MPJ6_H4A_001[a]				Promoter	8082720	−208	CGGCGGGGGACC G/C ATTAACCATTAA		73	1	0	0.007
MPJ6_H4A_002[a]				Intron 1	8083273	IVS1+231	CCAGCAAAAGTC G/A ATCCCGGCTATT		71	3	0	0.020
MPJ6_H4A_003	IMS-JST006533	rs2071197		Intron 1	8083350	IVS1+308	TGGCACAGTGAC G/A TGATGGTGAGCT		18	45	11	0.453
MPJ6_H4A_004	IMS-JST006534	rs2071198		Intron 1	8083399	IVS1+357	ATTCAGCCCAGC A/T CGGCCCCTTCGT		70	4	0	0.027
MPJ6_H4A_005		rs736824	17	Intron 1	8087575	IVS1−38	CCTGACATTCTG C/T TCTTCCTGAAGC		48	23	3	0.196
MPJ6_H4A_006		rs745975	17, 21, 22	Intron 1	8087608	IVS1−5	CTTCTCTCCTGG C/T GCAGACACGTCC		42	31	1	0.223
MPJ6_H4A_007		rs6093976		Intron 2	8088695	IVS2−241	CAACCAGCAGAG C/T GACCCAGGACCA		48	25	1	0.182
MPJ6_H4A_008[a]				Intron 2	8088853	IVS2−83	AGGTTGGGGGGT C/T AACTGGATAGCC		73	1	0	0.007
MPJ6_H4A_009[a]				Intron 3	8089080	IVS3+50	CACCTGCACCCA C/T AGCTCCCCGACA		73	1	0	0.007
MPJ6_H4A_010		rs13041396		Intron 3	8095045	IVS3−204	CTTATAGCCTCT C/G CATTGTGTTGGG		72	2	0	0.014
MPJ6_H4A_011[a]				Intron 3	8095195	IVS3−54	CACAGACACCCC C/- ACCCCCTACTCC		73	1	0	0.007
MPJ6_H4A_012		rs1800961	21, 22	Exon 4	8095279	416	ACCGGATCAGCA C/T TCGAAGGTCAAG	T139I	73	1	0	0.007
MPJ6_H4A_013		rs11574738		Intron 4	8095495	IVS4+140	GGGACACTGAGT C/G CGGTTTCACATG		72	2	0	0.014
MPJ6_H4A_014		rs11574739		Intron 4	8095865	IVS4−197	ACAGGTGAAGGC A/C CAGAGGGAGCCC		72	2	0	0.014
MPJ6_H4A_015		rs3212194		Intron 4	8095966	IVS4−96	CCAGCCCCCTCC C/G CACATCTGATTC		72	2	0	0.014
MPJ6_H4A_016		rs3212195		Intron 4	8096010	IVS4−52	AGGGGACAGAGA G/A TGCGGGAGGGCC		72	2	0	0.014
MPJ6_H4A_017[a]				Intron 5	8096390_3	IVS5+173_176	ATATTAACTCAG TTAG/- CTCCTCCAACAA		73	1	0	0.007
MPJ6_H4A_018[a]				Intron 5	8099799_800	IVS5−181_−180	AGCTCTGAGCAC AT/- GTTCTTTCCCCT		73	1	0	0.007
MPJ6_H4A_019	IMS-JST162873	rs3212200		Intron 5	8099844	IVS5−136	GGGTAGTAGTTT T/C ATGATGCCCATT		37	36	1	0.257
MPJ6_H4A_020		rs6103731		Intron 6	8100208	IVS6+141	CAGGCTTGCATT A/G GAGGGCTCCAGG		33	34	7	0.324
MPJ6_H4A_021		rs11086925		Intron 6	8100263	IVS6+196	ATGCAAGGAAAT G/A TGGATGCAAGTC		72	2	0	0.014
MPJ6_H4A_022	IMS-JST069233	rs2273618		Intron 7	8105485	IVS7−88	CTCCTTGTGTGA T/C ACAAGTCAGGGG		23	42	9	0.405
MPJ6_H4A_023[a]				Intron 8	8109784	IVS8−106	CAGGCACTGCCA A/G TATTGGATGGGC		73	1	0	0.007
MPJ6_H4A_024[a]				Exon 9	8109914	1154	ATGCACCCCATG C/T CCACCACCCCCT	A385V	73	1	0	0.007
MPJ6_H4A_025[a]				Exon 9	8109953	1193	TGCAGGAACATA T/C GGGAACCAACGT	M398T	73	1	0	0.007
MPJ6_H4A_026[a]				Intron 9	8110927	IVS9−151	ACCTTAGGGATT A/C TCTGGTTTAATT		72	2	0	0.014
MPJ6_H4A_027	IMS-JST098122	rs3746574		Intron 9	8110933	IVS9−145	GGGATTATCTGG T/C TTAATTAATTCT		33	37	4	0.304
MPJ6_H4A_028	IMS-JST098123	rs3746575		Intron 9	8111011	IVS9−67	AACTTTCCCGGG C/G CTCTTCATTTAC		24	42	8	0.392
MPJ6_H4A_029	IMS-JST098124	rs3746576		Exon 10	8111106	1311	ACAGCCCTCACC G/T CCAGGTGGCTCA	P437P	73	1	0	0.007
MPJ6_H4A_030[a]				3′-UTR	8111375	1580 (*155)[c]	CATGGCCTAAGG G/A CCACATCCCACT		66	8	0	0.054
MPJ6_H4A_031		rs11086926		3′-UTR	8111612	1817 (*392)[c]	TCTCCTAGCCCC T/G GTCATGGTGTCC		69	5	0	0.034
MPJ6_H4A_032[a]				3′-UTR	8111647	1852 (*427)[c]	CTGTGAGGCTGG G/T TCCAATTGTGGC		67	7	0	0.047
MPJ6_H4A_033[a]				3′-UTR	8111975	2180 (*755)[c]	GAGAAAACAAAC C/T CAGGTTGGCGAC		73	1	0	0.007
MPJ6_H4A_034[a]				3′-UTR	8111985	2190 (*765)[c]	ACCCAGGTTGGC G/A ACTGCAACAGGA		73	1	0	0.007
MPJ6_H4A_035		rs3212210		3′-UTR	8112126	2331 (*906)[c]	GAGCCAAGGCCT C/A GTGGTAGTAAGA		25	40	9	0.392
MPJ6_H4A_036		rs3212210		3′-UTR	8112126	2331 (*906)[c]	GAGCCAAGGCCT C/T GTGGTAGTAAGA		72	2	0	0.014
MPJ6_H4A_037[a]				3′-UTR	8112157_75	2362_2380 (*937_955)[c]	CAAGAATTGAGG AAGAATGGTGTGGGAGAGG/- GATGATGAAGAG		35	32	7	0.311
MPJ6_H4A_038		rs6130615		3′-UTR	8112352	2557 (*1132)[c]	GATGATATAATG T/C GGGTGAGAGTAA		28	31	15	0.412
MPJ6_H4A_039	IMS-JST114749	rs3834658		3′-UTR	8112519_26	2724_2731 (*1299_1306)[c]	TTAATCCTCCCT TCCTCCCT/- ATTAACCTAGAG		57	16	1	0.122

[a] Novel variations detected in this study.

[b] Exon-intron boundary and amino acid numbering were based on the isoform 2.

[c] Numbered from the termination codon TGA.

Table 3. Haplotypes of *HNF4A* (Block 1)

Nucleotide change[a]			-208G>C	IVS1+ 231G>A	IVS1+ 308G>A	IVS1+ 357A>T	IVS1- 38C>T	IVS1- 5C>T	IVS2- 241C>T	IVS2- 83C>T	IVS3+50 C>T	Number	Frequency
Amino acid change													
Haplotypes[b, c, d]	*1	*1a										63	0.426
		*1b										27	0.182
		*1c										24	0.162
		*1d										18	0.122
		*1e										6	0.041
		*1f										4	0.027
		*1g										3	0.020
		*1h										1	0.007
		*1i ?										1	0.007
		*1j ?										1	0.007

[a] A of the translational start codon of *HNF4A* is numbered 1. NT_011362.9 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 haplotypes inferred in only one subject are described with haplotype names and a question mark.

Table 4. Haplotypes of *HNF4A* (Block 2)

Nucleotide change[a]			IVS3 -204 C>G	IVS3 -54 delC	416 C>T	IVS4 +140 C>G	IVS4 -197 A>C	IVS4 -96 C>G	IVS4 -52 G>A	IVS5 +173 _176 del TTAG	IVS5 -181 _-180 del AT	IVS5 -136 T>C	IVS6 +141 A>G	IVS6 +196 G>A	IVS7 -88 T>C	IVS8 -106 A>G	1154 C>T	1193 T>C	IVS9 -151 A>C	IVS9 -145 T>C	IVS9 -67 C>G	1311 G>T	1580 (*155)[b] G>A	1817 (*392)[b] T>G	Number	Frequency
Amino acid change					T139I												A385V	M398T				P437P				
Haplotypes[c, d, e]	*1	*1a																							85	0.574
		*1b																							35	0.236
		*1c																							8	0.054
		*1d																							4	0.027
		*1e																							2	0.014
		*1f																							2	0.014
		*1g																							2	0.014
		*1h																							2	0.014
		*1i																							1	0.007
		*1j																							1	0.007
		*1k ?																							1	0.007
		*1l ?																							1	0.007
		*1m ?																							1	0.007
		*1n ?																							1	0.007
	*2	*2a ?																							1	0.007
	*3	*3a																							1	0.007

[a] A of the translational start codon of *HNF4A* is numbered 1. NT_011362.9 was used as the reference sequence.
[b] Numbered from the termination codon TAG.
[c] Major allele, white; minor allele, gray.
[d] The haplotypes are described as numbers plus small alphabetical letters.
[e] The haplotypes inferred in only one subject are described with haplotype names and a question mark.

Table 5. Haplotypes of *HNF4A* (Block 3)

Nucleotide change[a]			1852 (*427)[b] G>T	2180 (*755)[b] C>T	2190 (*765)[b] G>A	2331 (*906)[b] C>A	2331 (*906)[b] C>T	2362_2380 (*937_955)[b] del AAGAATGGTG TGGGAGAGG	2557 (*1132)[b] T>C	2724_2731 (*1299_1306)[b] del TCCTCCCT	Number	Frequency
Amino acid change												
Haplotypes[c,d,e]	*1	*1a									39	0.264
		*1b									36	0.243
		*1c									26	0.176
		*1d									12	0.081
		*1e									10	0.068
		*1f									7	0.047
		*1g									7	0.047
		*1h									6	0.041
		*1i									2	0.014
		*1j									1	0.007
		*1k ?									1	0.007
		*1l ?									1	0.007

[a] A of the translational start codon of *HNF4A* is numbered 1. NT_011362.9 was used as the reference sequence.
[b] Numbered from the termination codon TAG.
[c] Major allele, white; minor allele, gray.
[d] The haplotypes are described as numbers plus small alphabetical letters.
[e] The haplotypes inferred in only one subject are described with haplotype names and a question mark.

similar to those inferred in Chinese reported previously, except for the regions that we did not analyze.[22]

Haplotype analysis was then performed (**Tables 3 to 5**). The haplotypes assigned in this study are shown as numbers plus small alphabetical letters. As for block 1 spanning 6.4 kb from the 5′ promoter region to intron 3, seven haplotypes were first unambiguously assigned by homozygous variations at all sites (**1a* to **1d*) or a heterozygous variation at only one site (**1e*, **1f*, and **1h*). Separately, the diplotype configurations (a combination of haplotypes) for all 74 patients were estimated with over 0.99 certainty by LDSUPPORT software. The additionally inferred haplotypes were three **1* subtypes (**1g*, **1i*, and **1j*). In our separate experiment, the **1g* haplotype was unambiguously identified by the presence of a **1g* homozygote in a cell line MEG-01 (data not shown). The determined/inferred haplotypes were summarized in **Table 3**. The most frequent haplotype was **1a* (frequency: 0.426), followed by **1b* (0.182), **1c* (0.162), and **1d* (0.122). The frequencies of the other haplotypes were less than 0.1.

Block 2 spans 16.6 kb from intron 3 to the 3′-UTR. Five haplotypes were unambiguously assigned by homozygous variations at all sites (**1a*) or a heterozygous variation at only one site (**1b*, **1d*, **1i*, and **1j*). From analysis with the software, the diplotype configurations for all 74 patients were estimated with over 0.99 certainty, except for 1 patient inferred to be **1d*/**1h*. The additionally inferred haplotypes were nine **1* subtypes (**1c*, **1e* to **1h*, and **1k* to **1n*), **2a* (with 416C>T, T139I), and **3a* (with 1154C>T, A385V; 1193T>C, M398T; confirmed by cloning and sequencing analysis). The summary of the determined/inferred haplotypes was shown in **Table 4**. The most frequent haplotype was **1a* (frequency: 0.574), followed by **1b* (0.236). The frequencies of the other haplotypes including **2a* and **3a* were less than 0.1.

Regarding block 3 including eight 3′-UTR variations, nine haplotypes were unambiguously assigned by homozygous variations at all sites (**1a* to **1c* and **1e*) or a heterozygous variation at only one site (**1d*, **1f*, and **1h* to **1j*). From analysis with the software, the diplotype configurations for all 74 patients were estimated with over 0.99 certainty, except for 8 patients (inferred as **1a*/**1c* for 6 patients and **1b*/**1d* for 2 patients). The additionally inferred haplotypes by the software were three **1* subtypes (**1g*, **1k*, and **1l*). The determined/inferred haplotypes were summarized in **Table 5**. The haplotypes with more than 0.1 frequency were **1a* (0.264), **1b* (0.243), and **1c* (0.176).

In conclusion, 39 genetic variations, including 16 novel ones, were detected in *HNF4A* in the Japanese patients. Using the detected variations, 10, 16, and 12 haplotypes were determined and/or inferred for block 1, 2, and 3, respectively. Our results on *HNF4A* variations and haplotypes would be useful for pharmacogenetic studies in Japanese.

Acknowledgements: We appreciate great help from Drs. Kei Fujimoto and Kanae Yasuda at Hiranuma Clinic. We also thank Ms. Chie Sudo for her secretarial assistance.

Hiromi Fukushima-Uesaka and Yoshiro Saito contributed equally to this work.

This study was supported in part by the Program for the Promotion of Fundamental Studies in Health Sciences and by the Health and Labor Sciences Research Grants (Research on Advanced Medical Technology) from the Ministry of Health, Labor and Welfare of Japan.

References

1) Drewes, T., Senkel, S., Holewa, B. and Ryffel, G. U.: Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol. Cell. Biol.* **16**: 925-931 (1996).
2) Hansen, S. K., Parrizas, M., Jensen, M. L., Pruhova, S., Ek, J., Boj, S. F., Johansen, A., Maestro, M. A., Rivera, F., Eiberg, H., Andel, M., Lebl, J., Pedersen, O., Ferrer, J. and Hansen T.: Genetic evidence that HNF-1alpha-dependent transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function. *J. Clin. Invest.* **110**: 827-833 (2002).
3) Thomas, H., Jaschkowitz, K., Bulman, M., Frayling, T. M., Mitchell, S. M., Roosen, S., Lingott-Frieg, A., Tack, C. J., Ellard, S., Ryffel, G. U. and Hattersley, A. T.: A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum. Mol. Genet.* **10**: 2089-2097 (2001).
4) Kritis, A. A., Argyrokastritis, A., Moschonas, N. K., Power, S., Katrakili, N., Zannis, V. I., Cereghini, S. and Talianidis, I.: Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4. *Gene* **173**: 275-280 (1996).
5) Eeckhoute, J., Moerman, E., Bouckenooghe, T., Lukoviak, B., Pattou, F., Formstecher, P., Kerr-Conte, J., Vandewalle, B. and Laine, B.: Hepatocyte nuclear factor 4 alpha isoforms originated from the P1 promoter are expressed in human pancreatic beta-cells and exhibit stronger transcriptional potentials than P2 promoter-driven isoforms. *Endocrinology* **144**: 1686-1694 (2003).
6) Hadzopoulou-Cladaras, M., Kistanova, E., Evagelopoulou, C., Zeng, S., Cladaras, C. and Ladias, J. A.: Functional domains of the nuclear receptor hepatocyte nuclear factor 4. *J. Biol. Chem.* **272**: 539-550 (1997).
7) Iordanidou, P., Aggelidou, E., Demetriades, C. and Hadzopoulou-Cladaras, M.: Distinct amino acid residues may be involved in coactivator and ligand interactions in hepatocyte nuclear factor-4alpha. *J. Biol.*

Chem. **280**: 21810-21819 (2005).
8) Ihara, A., Yamagata, K., Nammo, T., Miura, A., Yuan, M., Tanaka, T., Sladek, F. M., Matsuzawa, Y., Miyagawa, J. and Shimomura, I.: Functional characterization of the HNF4alpha isoform (HNF4alpha8) expressed in pancreatic beta-cells. *Biochem. Biophys. Res. Commun.* **329**: 984-990 (2005).
9) Chen, Y., Kissling, G., Negishi, M. and Goldstein, J. A.: The nuclear receptors constitutive androstane receptor and pregnane X receptor cross-talk with hepatic nuclear factor 4alpha to synergistically activate the human CYP2C9 promoter. *J. Pharmacol. Exp. Ther.* **314**: 1125-1133 (2005).
10) Ferguson, S. S., Chen, Y., LeCluyse, E. L., Negishi, M. and Goldstein, J. A.: Human CYP2C8 is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor, glucocorticoid receptor, and hepatic nuclear factor 4alpha. *Mol. Pharmacol.* **68**: 747-757 (2005).
11) Pitarque, M., Rodriguez-Antona, C., Oscarson, M. and Ingelman-Sundberg, M.: Transcriptional regulation of the human CYP2A6 gene. *J. Pharmacol. Exp. Ther.* **313**: 814-822 (2005).
12) Tirona, R. G., Lee, W., Leake, B. F., Lan, L. B., Cline, C. B., Lamba, V., Parviz, F., Duncan, S. A., Inoue, Y., Gonzalez, F. J., Schuetz, E. G. and Kim, R. B.: The orphan nuclear receptor HNF4alpha determines PXR- and CAR-mediated xenobiotic induction of CYP3A4. *Nat. Med.* **9**: 220-224 (2003).
13) Barbier, O., Girard, H., Inoue, Y., Duez, H., Villeneuve, L., Kamiya, A., Fruchart, J. C., Guillemette, C., Gonzalez, F. J. and Staels, B.: Hepatic expression of the UGT1A9 gene is governed by hepatocyte nuclear factor 4alpha. *Mol. Pharmacol.* **67**: 241-249 (2005).
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) Hani, E. H., Suaud, L., Boutin, P., Chevre, J. C., Durand, E., Philippi, A., Demenais, F., Vionnet, N., Furuta, H., Velho, G., Bell, G. I., Laine, B. and Froguel, P.: A missense mutation in hepatocyte nuclear factor-4 alpha, resulting in a reduced transactivation activity, in human late-onset non-insulin-dependent diabetes mellitus. *J. Clin. Invest.* **101**: 521-526 (1998).
16) Furuta, H., Iwasaki, N., Oda, N., Hinokio, Y., Horikawa, Y., Yamagata, K., Yano, N., Sugahiro, J., Ogata, M., Ohgawara, H., Omori, Y., Iwamoto, Y. and Bell G. I.: Organization and partial sequence of the hepatocyte nuclear factor-4 alpha/MODY1 gene and identification of a missense mutation, R127W, in a Japanese family with MODY. *Diabetes* **46**: 1652-1657 (1997).
17) Yang, Q., Yamagata, K., Yamamoto, K., Cao, Y., Miyagawa, J., Fukamizu, A., Hanafusa, T. and Matsuzawa, Y.: R127W-HNF-4alpha is a loss of function mutation but not a rare polymorphism and causes Type II diabetes in a Japanese family with MODY1. *Diabetologia* **43**: 520-524 (2000).
18) Zhu, Q., Yamagata, K., Miura, A., Shihara, N., Horikawa, Y., Takeda, J., Miyagawa, J. and Matsuzawa, Y.: T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**: 567-573 (2003).
19) Ek, J., Rose, C. S., Jensen, D. P., Glumer, C., Borch-Johnsen, K., Jorgensen, T., Pedersen, O. and Hansen, T.: The functional Thr130Ile and Val255Met polymorphisms of the hepatocyte nuclear factor-4alpha (HNF4A): gene associations with type 2 diabetes or altered beta-cell function among Danes. *J. Clin. Endocrinol. Metab.* **90**: 3054-3059 (2005).
20) Sakurai, K., Seki, N., Fujii, R., Yagui, K., Tokuyama, Y., Shimada, F., Makino, H., Suzuki, Y., Hashimoto, N., Saito, Y., Egashira, T., Matsui, K. and Kanatsuka, A.: Mutations in the hepatocyte nuclear factor-4alpha gene in Japanese with non-insulin-dependent diabetes: a nucleotide substitution in the polypyrimidine tract of intron 1b. *Horm. Metab. Res.* **32**: 316-320 (2000).
21) Rissanen, J., Wang, H., Miettinen, R., Karkkainen, P., Kekalainen, P., Mykkanen, L., Kuusisto, J., Karhapaa, P., Niskanen, L., Uusitupa, M. and Laakso, M.: Variants in the hepatocyte nuclear factor-1alpha and -4alpha genes in Finnish and Chinese subjects with late-onset type 2 diabetes. *Diabetes Care* **23**: 1533-1538 (2000).
22) Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K. A. and Cox, D. R.: Whole-genome patterns of common DNA variation in three human populations. *Science* **307**: 1072-1079 (2005).

Drug Metab. Pharmacokinet. 22 (1): *SNP1* (48)-*SNP13* (60) (2007).

SNP Communication

Genetic Variations and Haplotype Structures of the ABC Transporter Gene ABCC1 in a Japanese Population

Hiromi FUKUSHIMA-UESAKA[1], Yoshiro SAITO[1,2,*], Masahiro TOHKIN[1,3], Keiko MAEKAWA[1,2], Ryuichi HASEGAWA[3], Manabu KAWAMOTO[4], Naoyuki KAMATANI[4], Kazuko SUZUKI[5], Tatsuo YANAGAWA[5], Hiroshi KAJIO[6], Nobuaki KUZUYA[6], Kazuki YASUDA[7] and Jun-ichi SAWADA[1,2]

[1]*Project Team for Pharmacogenetics,* [2]*Division of Biochemistry and Immunochemistry,* [3]*Division of Medicinal Safety Science, National Institute of Health Sciences, Tokyo, Japan* [4]*Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan,* [5]*Nerima General Hospital, Tokyo, Japan* [6]*Division of Endocrine and Metabolic Diseases, the Hospital,* [7]*Department of Metabolic Disorder, Research Institute, International Medical Center of Japan, Tokyo, Japan*

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

Summary: Multidrug resistance-related protein 1 (MRP1), an ATP-binding cassette transporter encoded by the *ABCC1* gene, is expressed in many tissues, and functions as an efflux transporter for glutathione-, glucuronate- and sulfate-conjugates as well as unconjugated substrates. In this study, the 31 exons and their flanking introns of *ABCC1* were comprehensively screened for genetic variations in 153 Japanese subjects to elucidate the linkage disequilibrium (LD) profiles and haplotype structures of *ABCC1* that is necessary for pharmacogenetic studies of the substrate drugs. Eighty-six genetic variations including 31 novel ones were found: 1 in the 5′-flanking region, 1 in the 5′-untranslated region (UTR), 20 in the coding exons (9 synonymous and 11 nonsynonymous variations), 4 in the 3′-UTR, and 60 in the introns. Of these, eight novel nonsynonymous variations, 726G>T (Trp242Cys), 1199T>C (Ile400Thr), 1967G>C (Ser656Thr), 2530G>A (Gly844Ser), 3490G>A (Val1164Ile), 3550G>A (Glu1184Lys), 3901C>T (Arg1301Cys), and 4502A>G (Asp1501Gly), were detected with an allele frequency of 0.003. Based on the LD profiles, the analyzed regions of the gene were divided into five LD blocks (Blocks -1 and 1 to 4). The multiallelic repeat polymorphism in the 5′-UTR was defined as Block -1. For Blocks 1, 2, 3 and 4, 32, 23, 23 and 13 haplotypes were inferred, and 9, 7, 7 and 6 haplotypes commonly found on ≥10 chromosomes accounted for ≥91% of the inferred haplotypes in each block. Haplotype-tagging single nucleotide polymorphisms for each block were identified to capture the common haplotypes. This study would provide fundamental and useful information for the pharmacogenetic studies of MRP1-dependently effluxed drugs in Japanese.

Key words: ***ABCC1*; genetic variation; amino acid change; haplotype; haplotype tagging SNP**

As of August 8, 2006, the novel variations reported here are not found in the database of Japanese Single Nucleotide Polymorphisms (http://snp.ims.u-tokyo.ac.jp/), dbSNP in the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/SNP/), or PharmGKB Database (http://www.pharmgkb.org/).

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

Introduction

The multidrug resistance-related protein 1 (MRP1) encoded by the ATP-binding cassette transporter C1 gene (*ABCC1*) belongs to a superfamily of ABC transporters.[1,2] MRP1 was originally identified as the overexpressed transporter cloned from a doxorubicin-selected lung cancer cell line H69AR.[3] The *ABCC1* gene, encod-

Received; August 8, 2006, Accepted; October 5, 2006

**To whom correspondence should be addressed*: Yoshiro SAITO, Ph.D., *Division of Biochemistry and Immunochemistry, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Tel. +81-3-5717-3831, Fax. +81-3-5717-3832, E-mail: yoshiro@nihs.go.jp*

ing a 1531-amino acid, 190 kDa membrane protein, consists of 31 exons and spans over 200 kb at chromosome 16p13.1.[1,4] The protein has 17 transmembrane helices (TM) in three membrane spanning domains (consisting of 5, 6 and 6 TMs) and 2 large cytoplasmic domains between TM11 and TM12, and downstream of TM17.[1,2] These two cytoplasmic domains contain nucleotide binding domains where binding and hydrolysis of ATP occurs to facilitate substrate transport. Two sequence motifs in the nucleotide binding domain are well conserved among all ABC transporters and called Walker A and Walker B. The Walker A motif is involved in the binding of the β-phosphate of ATP, while the Walker B motif interacts with Mg^{2+} ion. In addition, a third conserved motif known as the ABC signature sequence (LSGGQ), located between the Walker A and Walker B motifs, has been identified as a possible binding site for the γ-phosphate of ATP.[2]

MRP1 protein is expressed in many tissues throughout the body at relatively high levels in the heart, adrenal gland, lung, and skeletal muscle, and at medium levels in the liver and kidney.[5] In most tissues, MRP1 is localized to the basolateral membranes of cells for the efflux of its substrates into the blood and thus might play a protective role against toxic substances and metabolites.

In vitro studies have indicated that a number of anticancer drugs are good substrates for MRP1, such as doxorubicin, vincristine and methotrexate. Therefore, MRP1 is thought to be involved in drug resistance in cancer patients.[1,2] In addition, an anti-diabetic drug glibenclamide, but not tolbutamide, is known to inhibit transporting activity of MRP1.[6] MRP1 is also a primarily active transporter of glutathione-, glucuronate- and sulfate-conjugates such as the inflammatory mediator leukotriene C_4 (LTC_4).[1,2] Many unconjugated substrates are also known to be transported concurrently with reduced glutathione. Besides, glutathione disulfide is also transported by MRP1. Thus, MRP1 is thought to be involved in the tissue distribution and elimination of drugs and organic anions, and possibly in redox homeostasis.[1,2]

Genetic polymorphisms in the metabolizing enzymes and transporters are known to influence drug metabolism and disposition. As for the *ABCC1* gene, several polymorphisms with functional significance have been found. In Caucasian populations, 1299G>T (Arg433Ser), 1898G>A (Arg633Gln), 2012G>T (Gly671Val), and 4535C>T (Ser1512Leu) have been reported.[7-9] In addition, Arg433Ser decreases the transport activity for LTC_4 and estrone sulfate, but not for estradiol 17β-glucuronide, *in vitro*.[10] Ito *et al.* found 16 genetic polymorphisms, including 4 nonsynonymous and 8 synonymous ones, in 48 Japanese subjects.[11] An *in vitro* functional study showed that one of the nonsynonymous variations, 2168G>A (Arg723Gln), leads to reduced transport activity for LTC_4, estradiol 17β-glucuronide and methotrexate.[12] However, no haplotype analysis has been reported for the Japanese population. Haplotype is the linked combinations of genetic polymorphisms on the same chromosome and has been shown to sometimes render higher associations with clinical parameters such as drug responses and adverse effects than individual polymorphisms.[13] This information may also be useful for identification of the real functionally-relevant polymorphisms from the linked polymorphisms.

In this study, we searched for genetic variations in *ABCC1* by resequencing all the 31 exons and their surrounding introns of 153 Japanese subjects. The detected variations were then used to perform linkage disequilibrium (LD) and haplotype analyses to identify the haplotype-tagging single nucleotide polymorphisms (htSNPs) that are sufficient to capture common haplotypes in Japanese subjects.

Materials and Methods

Human genomic DNA samples: One hundred fifty-three Japanese subjects participating in this study consisted of 86 diabetic patients administered glimepiride and 67 healthy volunteers. Genomic DNA was extracted from blood leukocytes of the diabetic patients, or from Epstein-Barr virus transformed B lymphocytes derived from the healthy volunteers. The ethical review boards of the International Medical Center of Japan, Nerima General Hospital, Tokyo Women's Medical University and National Institute of Health Sciences approved this study. Written informed consent was obtained from all subjects.

PCR conditions for DNA sequencing: First, two sets of multiplex PCR were performed to amplify all 31 exons of *ABCC1* from 100 ng of genomic DNA using 1.25 units of Z-Taq (Takara Bio Inc., Shiga, Japan) with 0.20 μM each of the mixed primers (Mix 1 and Mix 2) designed in the intronic regions as listed in **Table 1** ("1st PCR"). Mix 1 contained the primers for exons 1, 2 to 5, 8 to 12, 13, 14, and 20 to 23, and Mix 2 for exons 6, 7, 15 to 19, 24 to 26, and 27 to 31. The first PCR conditions were 30 cycles of 98°C for 5 sec, 55°C for 5 sec, and 72°C for 190 sec. Next, each exon, except for exons 1, 4, 22 and 23, was amplified separately in the 2nd PCR using the 1st PCR product as a template by Ex-Taq (0.625 units, Takara Bio Inc.) with the primer sets listed in "2nd PCR" in **Table 1**. Because of high GC contents, exons 1, 4, 22 and 23 were amplified using 0.25 units of LA-Taq (Takara Bio Inc.) with GC buffer II (exon 1) or GC buffer I (exons 4, 22 and 23) using 0.5 μM of the primers listed in **Table 1**. The second 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

Table 1. Primers used for sequencing *ABCC1*

		Amplified or sequenced region	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Amplified length (bp)
1st PCR	Mix 1	Exon 1	GAGGAGAGGAGAAAAGAAAGGCATC	TTAGTAGAGACGGGGGTCAATCCAT	4,788
		Exon 2 to 5	GTGGAGGTTTATTCTTGGGCAGGTA	CCTCCTCCTCCATACTGATGCCCAC	11,757
		Exon 8 to 12	CTCCTGAGTTCAAGCGATTCTCCTT	CCTATGAGACACTGACAATCCACAC	14,059
		Exon 13 and 14	CTCTTGACATCGGAGCCTGGAAAAT	GGGGATGGGCTGGGAAACGATAAAT	5,445
		Exon 20 to 23	CTTGGTGTCCTTATCAGTGTCCTTC	GGCAGGGGAATCACTTGAACTGGGA	13,579
	Mix 2	Exon 6 and 7	TAAAAAGCACGCCCAGCCTTGAATG	TTTCCTTACAATGCTCAGTCACAGA	6,789
		Exon 15 to 19	GAAGAGTAAAAGAAGACAGAGGTGC	GCTGAACTGCTTTATGAGGGAGGAA	15,848
		Exon 24 to 26	AGCCTTGTTGTTCCCATTTTACAGA	ACTTAGTGAAACCCCCATCTCTACT	5,242
		Exon 27 to 31	ACCTCCACTTGCTCTTTTGAAATAC	CTCTCTGGACTAACATAAAGGGATT	13,210
2nd PCR		Exon 1	CGTTATTTTCCCCTGGTGAC	GGTTGTTTTTACAGACGGGA	838
		Exon 2	GTGTTTTTCATAGAGGCAGC	ACTAAGCGGCAGAGCAAAGA	800
		Exon 3	GCAGGCTGATTACAGACCAT	AGCCTGGATGAAAGCGAGAT	534
		Exon 4	GAGTAGTTTTGATGTAGTCTATGGC	GCTGGGGCTGATTTAGTGAGGATTT	719
		Exon 5	GAGATGGAGTTTTGCTCTGTCACAC	CAGTATAATACAACAGGGAGTGCCG	876
		Exon 6	TCTGGAAGGAAGAGTGAGCA	AGCACAGGCAGATTCAAGAC	312
		Exon 7	AGAGGGGAGCAGCATCAGCA	TTCTCTGCCCAGTTGCTTTA	445
		Exon 8	AGCCCGTGTTTGTAACCACT	GAGACCAGCCTGACCAACAA	580
		Exon 9	TTTGGGTGACTTGCTGGTGA	TACAATGCCTGATGCGTGCT	605
		Exon 10	TGAGTTCAAGGTTGGGAGGC	CGTCCATTGTAGGGTAGTTA	650
		Exon 11	CAGGAATGAAACCACAGGCT	CGGGGCATCAGCATTTGTTA	419
		Exon 12	GTGAAACCCCATCTCTATTG	ACCTGGGCAACATAGTGACC	462
		Exon 13	CGGTCGTTGATTTATCCAGT	CAAGTGCGAGTTCTCTGTCA	563
		Exon 14	GTGGGACCTTCAGAAATAAG	AGTGTGAGACAGGACAGAAG	428
		Exon 15	CAACCCCATCTTACAAGGAC	TTCTCACGACAGCCTAAGCA	507
		Exon 16	GCCTTCAGTGTTTAGTACAG	TGCTGGGAGACAGATGGAAA	502
		Exon 17	TGGCATCTGTTGTCCTTTGT	AAGTGAGACCTGAGCCACAC	537
		Exon 18	CCTGGTCTCAAGCAGTCCTT	GGACTTACCCCTATAAAGAC	496
		Exon 19	AGACCTGAGTTTTGCCCACC	CTCATCAGGTCCAAGGTCAT	603
		Exon 20	CTCAGTGGATGGAGCCTTCT	AAGAGTGCCACATTCCTTCT	531
		Exon 21	AACACTCCGTCTCTTATGCC	TCAAATCCCAGTTCTGCCGC	464
		Exon 22	CGACTTTGTTTACTGCTGAC	AAAGCACTCAAACACCCACT	566
		Exon 23	GAGACGGAGTTTCACTGTGTTGGC	GACAGGTGGAAAAGTAAAAC	779
		Exon 24	TCATTGGTGGTGCTATTCCT	AGTGCCGCTGTCTGCTCTTA	556
		Exon 25	CTGCGGAGTTACTTGAGTTA	TGTCAAATCCGTCTCCTGCT	422
		Exon 26	TAGTGACTGATGGGGTTTCG	CAGCATCCCACAGTCTCGTA	541
		Exon 27	AGGGGGTCATTTGGGGAATA	TCATTTGGTCTTCAGGCTGT	542
		Exon 28	AGCCGAGTCATTCCTTTTGG	AAAAGAACGATGAAGTAGGG	510
		Exon 29	GTTCAAGTGATTCTCCTGCCTCAGT	CCTGGATTGAGACCCGTTTTACAGA	703
		Exon 30	ACACAGATGTTGGGAGTGGA	AGGGATAAGGACAGTGTTGA	436
		Exon 31	CTGACCCGAAGCAGTGACTT	CGGATGCCAAGGGAGAGAAT	1,429
Sequencing*		Exon 1	AAAGTGGTCGCAGGGTGTGT	GTCACCCAAGTTTCCCCCAT	
		Exon 2	GGAGCCTTGTCTGTTTCTTC	AAGGAACTTAGGGTCAACTA	
		Exon 3	GCAGGCTGATTACAGACCAT	AAAAAAAAAAGGCTACATTT	
		Exon 4	AGCCTGGGTGACAAGAGTGA	GGGCTGATTTAGTGAGGATT	
		Exon 5	GGATTACAGTTGCCCACCAC	GCCAAGTGAGAAACCTACAG	
		Exon 12	AAGTTTATGAGAAAAATAGC	ACCTGGGCAACATAGTGACC	
		Exon 22	GTTTACTGCTGACTTTGTTG	ACTCAAACACCCACTCTACA	
		Exon 23	TTATTATTATTAGAAGTTGGGAGTC	AAAACTTAGGAAAAAACTGC	
		Exon 26	AAAGGAAAGTCAAGTACGCC	CAGCATCCCACAGTCTCGTA	
		Exon 29	TACAGGCGTGAACCACCGTA	ACAGATACAGAAACTGAGGC	
		Exon 31	GACTTGCCCAGGTCAGTTGT	CCCCAAGGAAATGAAGCGTT	
			TCTTTGAGATGCTTCTGGCT	GTGGGAACAGTAATAACAGC	
			GCTGTTATTACTGTTCCCAC	TAACATCTAAAAACAAGGAA	

*Only primers which were not the same as those used for the 2nd PCR were shown.

2 min, and then a final extension for 7 min at 72°C. The PCR products were then treated with a PCR Product Pre-Sequencing Kit (USB Co., Cleveland, OH, USA) and directly sequenced on both strands using an ABI BigDye Terminator Cycle Sequencing Kit ver. 3.1 (Applied Biosystems, Foster City, CA, USA) with the sequencing primers listed in **Table 1** (Sequencing). Excess dye was removed by a DyeEX96 kit (Qiagen, Hilden, Germany) and the eluates were analyzed on an ABI Prism 3730 DNA Analyzer (Applied Biosystems).

All novel SNPs were confirmed by repeated sequencing analyses of PCR products generated by a new genomic DNA amplification. Under the conditions used, up to 400 bases upstream of the translational start codon, all of the exons and their flanking introns were successfully sequenced for all the subjects. The genomic and cDNA sequences of *ABCC1* obtained from GenBank (NT 010393.15 and NM 004996.2, respectively) were used as the reference sequences. The nucleotide positions based on the cDNA sequence were numbered from the adenine of the translational initiation site or the nearest exons.

Linkage disequilibrium (LD) and haplotype analyses: Hardy-Weinberg equilibrium and LD analyses were performed by SNPAlyze software ver. 3.1 (Dynacom Co., Yokohama, Japan), and pairwise LDs between variations were 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 configurations (a combination of haplotypes) were inferred by an expectation-maximization based program LDSUPPORT software, which determines the posterior probability distribution of the diplotype for each subject based on the estimated haplotype frequencies.[14] Haplotypes without any amino acid change were designated as **1*, and the nonsynonymous SNP-bearing haplotypes were numerically numbered. Subtypes were named with small alphabetical letters in the order of their frequencies. The ambiguous haplotypes inferred in only one subjects are grouped as "*others*" (**1* group) in **Tables 3 to 6**. The PHASE program (ver. 2.1) was also used for inferring haplotypes of the block with a microsatellite marker detected.[15,16]

Results and Discussion

***ABCC1* variations found in a Japanese population:** We found 86 genetic variations, including 31 novel ones, from 153 Japanese subjects (**Table 2**). Of them, 1 was located in the 5'-flanking region, 1 in the 5'-untranslated region (UTR), 20 in the coding exons (9 synonymous and 11 nonsynonymous variations), 4 in the 3'-UTR, and 60 in the introns. Since we did not find statistically significant difference in the allelic distributions between diabetic patients and healthy volunteers ($p > 0.05$ by χ^2 test or Fisher's exact test) only except for 1VS28-266C>G ($p = 0.027$), we analyzed all the variations as one group. As for 1VS28-266C>G, the frequency of the minor allele in diabetic patients (0.267) was somewhat lower than that of the healthy volunteers (0.388). All the observed allele frequencies were in Hardy-Weinberg equilibrium ($p > 0.05$).

Eight novel nonsynonymous variations, 726G>T (Trp242Cys), 1199T>C (Ile400Thr), 1967G>C (Ser656Thr), 2530G>A (Gly844Ser), 3490G>A (Val1164Ile), 3550G>A (Glu1184Lys), 3901C>T (Arg1301Cys), and 4502A>G (Asp1501Gly), were found heterozygously in different subjects with an allele frequency of 0.003. All substituted amino acids were located in the cytoplasmic regions of the MRP1 protein.[1,2,12] Two of the changes occurred in the loop between TM5 and TM6 (Trp242Cys) and between TM7 and TM8 (Ile400Thr). Ser656Thr and Gly844Ser were found 22 residues upstream of the Walker A motif and 52 residues downstream of the Walker B motif in the nucleotide binding domain 1 in the loop between TM11 and TM12, respectively. Both Val1164Ile and Glu1184Lys resided in the loop between TM15 and TM16. Arg1301Cys was located 26 residues upstream of the Walker A motif, while and Asp1501Gly was 47 residues downstream of the Walker B motif, in the nucleotide binding domain 2 in the C-terminal. Using PolyPhen program (http://www.bork.embl-heiderberg.de/PolyPhen) to predict the functional effects of amino acid substitutions, two substitutions, Trp242Cys and Gly844Ser, were expected to alter the protein function based on the PSIC (position specific independent count) profile score differences derived from multiple alignments. The loop between TM5 and TM 6 where Trp242 resides is known to be important for interaction with glutathione.[17] Furthermore, Trp242 is located near the regions important for LTC_4 binding (residues 260–274)[18] and LTC_4 transporting activity of MRP1 proteins (amino acids 223–232).[19] As for Ile400Thr, Lys396 mutation to Glu or Ile, located 4 residues upstream, was shown to cause a reduced transport activity.[20] The functional significance of these 8 novel variations should be clarified in the future studies.

We also detected three known nonsynonymous variations, 218C>T (Thr73Ile), 2168G>A (Arg723Gln), and 3173G>A (Arg1058Gln) at frequencies of 0.007, 0.065 and 0.003, respectively. These frequencies were similar to those found in the earlier reports for Japanese[11] and Chinese.[21] One of the variations, Arg723Gln, leads to reduced transport activities for LTC_4, estradiol 17β-glucuronide and methotrexate.[12] We did not detect three previously reported variations: 2012G>T (Gly671Val; found with approximately 0.03 frequency in Caucasians), 3140G>C (Cys1047Ser; 0.05 in African-Americans), and 4535C>T (Ser1512Leu; 0.03 in Caucasians).[8,9,12] These SNPs might be ethnic-specific.

A known microsatellite marker, GCC repeats in the 5'-UTR, was also detected: 9 repeats (1 chromosome), 10 (1), 11 (49), 12 (20), 13 (154), 14 (52), 15 (12), 16 (11), 17 (3), 18 (1), 21 (1) and 23 (1). This polymorphism was first reported from an Italian group ranging from 7 to 14 repeats, but their transcriptional activities did not change between 7 and 14 repeats in an *in vitro* reporter

Table 2. Summary of *ABCC1* variations detected in this study

SNP ID		Pharm GKB[b]	Reference	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_AC1001[a]				5′-flanking	7356447	−241	GAGACGCGCGAGG>ATGAGCGGGCGCC		0.020	0.017	0.022
MPJ6_AC1002-013			22	5′-UTR	7356549_7356569	−139_−119	CCGGCTCCCTGC/(GCC)9-23/AGCGCTAGCGCC			(see the text)	
MPJ6_AC1014[a]				Intron 1	7414381	IVS1-371	TAGGCTGGTCTCG>AAACTCCCAGCCT		0.003	0.006	0.000
MPJ6_AC1015[a]				Intron 1	7414742	IVS1-10	TCGCCTGTGTTTG>TTGTTCGCAGGAC		0.003	0.006	0.000
MPJ6_AC1016[a]				Exon 2	7414766	63	CTGGAATGTCACG>ATGGAATACCAGC	T21T	0.003	0.006	0.000
MPJ6_AC1017			11	Exon 2	7414921	218	CTCTCAACAAAAC>TCAAAACTGTAAG	T73I	0.007	0.000	0.015
MPJ6_AC1018	rs4148335		25	Intron 3	7421139	IVS3-288	ATCCCAGCACCTT>GGGGAGGCCAAGG		0.069	0.081	0.052
MPJ6_AC1019	rs4148336		25	Intron 3	7421231	IVS3-196	TAAAAATACAAAA>CATTAGCTAGGCA		0.069	0.081	0.052
MPJ6_AC1020	rs4148337		25	Intron 3	7421361	IVS3-66	CTCCAGCCTGGGT>CGACAAGAGTGAA		0.333	0.355	0.306
MPJ6_AC1021[a]				Intron 4	7421606	IVS4+42	TTCAGTGGACCCG>AGAGGGAGAGATG		0.003	0.000	0.007
MPJ6_AC1022	rs34085651			Intron 4	7423180	IVS4-252	CAGCCTCCGCCTC>TCTGGGTTCAAGT		0.069	0.081	0.052
MPJ6_AC1023	rs185005	=	25	Intron 4	7423323	IVS4-109	AACTCTTGACCTG>CAGGTGTTCTGCC		0.320	0.297	0.351
MPJ6_AC1024	rs246215	=	25	Intron 4	7423332	IVS4-100	CCTCAGGTGTTCC>TGCCTGTCTCGGC		0.320	0.297	0.351
MPJ6_AC1025[a]				Intron 5	7423677	IVS5+120	CCTCAACCCCTGA>TTACAGGGAATAT		0.013	0.006	0.022
MPJ6_AC1026[a]				Intron 5	7423818	IVS5+261	ACTTGGCATGGTG>AATTTTGGAAAAT		0.003	0.000	0.007
MPJ6_AC1027	rs3837750			Intron 5	7423921_7423922	IVS5+364_+365	GATTAGGCCTAT/delAA/TCCTACAGGGCA		0.333	0.355	0.306
MPJ6_AC1028[a]				Intron 5	7439983	IVS5-62	AAGCTCTGACCTG>AGATGAAAAGTCA		0.003	0.006	0.000
MPJ6_AC1029[a]				Exon 7	7443456	726	CAGTGACCTCTGG>TTCCTTAAACAAG	W242C	0.003	0.006	0.000
MPJ6_AC1030	rs903880	=	9	Intron 7	7443593	IVS7+54	CTCCTTTCCACTC>ACTGTGGCCTCAA		0.059	0.070	0.045
MPJ6_AC1031	rs246232	=	21	Intron 7	7443603	IVS7+64	CTCCTGTGGCCTC>GAATCCAGGATGG		0.418	0.436	0.396
MPJ6_AC1032[a]				Intron 7	7443608	IVS7+69	GTGGCCTCAATCC>TAGGATGGGGCCC		0.003	0.000	0.007
MPJ6_AC1033	rs246221	=	11	Exon 8	7451401	825	GCCGGTGAAGGTT>CGTGTACTCCTCC	V275V	0.366	0.366	0.366
MPJ6_AC1034	rs8187851	=		Exon 9	7452778	1047	TTGCAGGTTGCTC>TATCAAGTTCGTG	L349L	0.003	0.006	0.000
MPJ6_AC1035	rs35587	=	7	Exon 9	7452793	1062	CAAGTTCGTGAAT>CGACACGAAGGCC	N354N	0.366	0.366	0.366
MPJ6_AC1036[a]				Exon 9	7452930	1199	AGACCGCTGTCAT>CTGGGGCTGTCTA	I400T	0.003	0.000	0.007
MPJ6_AC1037	rs35588	=	11	Intron 9	7452957	IVS9+8	GGAAGGTAGGGGA>GCGCTGTGCCATT		0.363	0.360	0.366
MPJ6_AC1038	rs35591		25	Intron 9	7454889	IVS9-189	AGGCACTGAGCAC>GCGCGGATAAGAA		0.363	0.360	0.366
MPJ6_AC1039[a]				Intron 9	7454890	IVS9-188	GGCACTGAGCACC>GGCGGATAAGAAT		0.003	0.006	0.000
MPJ6_AC1040	rs35592		25	Intron 9	7454902	IVS9-176	CGCGGATAAGAAT>CGTGGGCTTTGAG		0.366	0.366	0.366
MPJ6_AC1041	rs4148343		25	Intron 10	7455437	IVS10+198	CAGGTTGATGTTC>ATCCTTGGTGGCA		0.114	0.105	0.127
MPJ6_AC1042[a]				Intron 10	7459543	IVS10-117	TAAAAGTAACACA>GCCTTGCCCTGAA		0.042	0.041	0.045
MPJ6_AC1043	rs3743526	=	25	Intron 11	7459874	IVS11+122	CTGCCAGTTGGAC>GTCACTTGGGGAG		0.114	0.105	0.127
MPJ6_AC1044	rs35595		25	Intron 11	7462950	IVS11-78	TAAAGTTTATGAG>AAAAAATAGCTGG		0.209	0.215	0.201
MPJ6_AC1045	rs3765129	=	11	Intron 11	7462980	IVS11-48	TTGAGTGATGGGC>TTGATCCCAGGGT		0.114	0.105	0.127
MPJ6_AC1046[a]				Intron 12	7463262	IVS12+31	TTCCTGGCCCTCA>CTTGTTTGATGTT		0.007	0.012	0.000
MPJ6_AC1047	rs17265551		23	Intron 12	7463287	IVS12+56	TTATTTTCTCTGC>TGTACCTGAATAT		0.069	0.081	0.052
MPJ6_AC1048	rs4148348	=	23	Intron 12	7475007	IVS12-85	GGGAGGGCCCAAG>ACGCGTCTCCAGG		0.124	0.128	0.119
MPJ6_AC1049	rs35604	=	25	Intron 12	7475055	IVS12-37	TGACTCTCACTCA>GGGGCACAGCAGT		0.245	0.227	0.269
MPJ6_AC1050	rs35605	=	11	Exon 13	7475098	1684	TTGCAGGTGGCCC>TTGTGCACATTTG	L562L	0.245	0.227	0.269
MPJ6_AC1051	rs35606	=	25	Intron 13	7475343	IVS13+105	CACACTCCCGGTC>TGGGCTCCATGAG		0.150	0.134	0.172
MPJ6_AC1052			21	Intron 14	7478780	IVS14+115	ACTCTGCCCAGCC>TGCTTGTCTGCGA		0.020	0.023	0.015
MPJ6_AC1053[a]				Exon 15	7483316	1967	CCTGGGCCAGGAG>CCGACCCTCCCAC	S656T	0.003	0.006	0.000
MPJ6_AC1054[a]				Intron 15	7486189	IVS15-99	TTCTGTCTTTCTC>GTTTCTCTACTTG		0.003	0.000	0.007
MPJ6_AC1055	rs8187863		7	Exon 16	7486300	2001	CATCACCTTCTCC>TATCCCCGAAGGT	S667S	0.003	0.006	0.000
MPJ6_AC1056	rs2301666	=	11	Exon 16	7486306	2007	CTTCTCCATCCCC>TGAAGGTGCTTTG	P669P	0.042	0.047	0.037
MPJ6_AC1057	rs2644983	=		Intron 16	7486595	IVS16+181	TCTGTTCTGTTCC>TGTGCTGTCCTAG		0.278	0.262	0.299

cont.

Table 2.

SNP ID				Position			Amino acid change	Frequency		
This Study	dbSNP (NCBI)	Pharm GKB[b]	Reference	Location	NT_010393.15	From the translational initiation site or from the end of the nearest exon[c]	Nucleotide change		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.

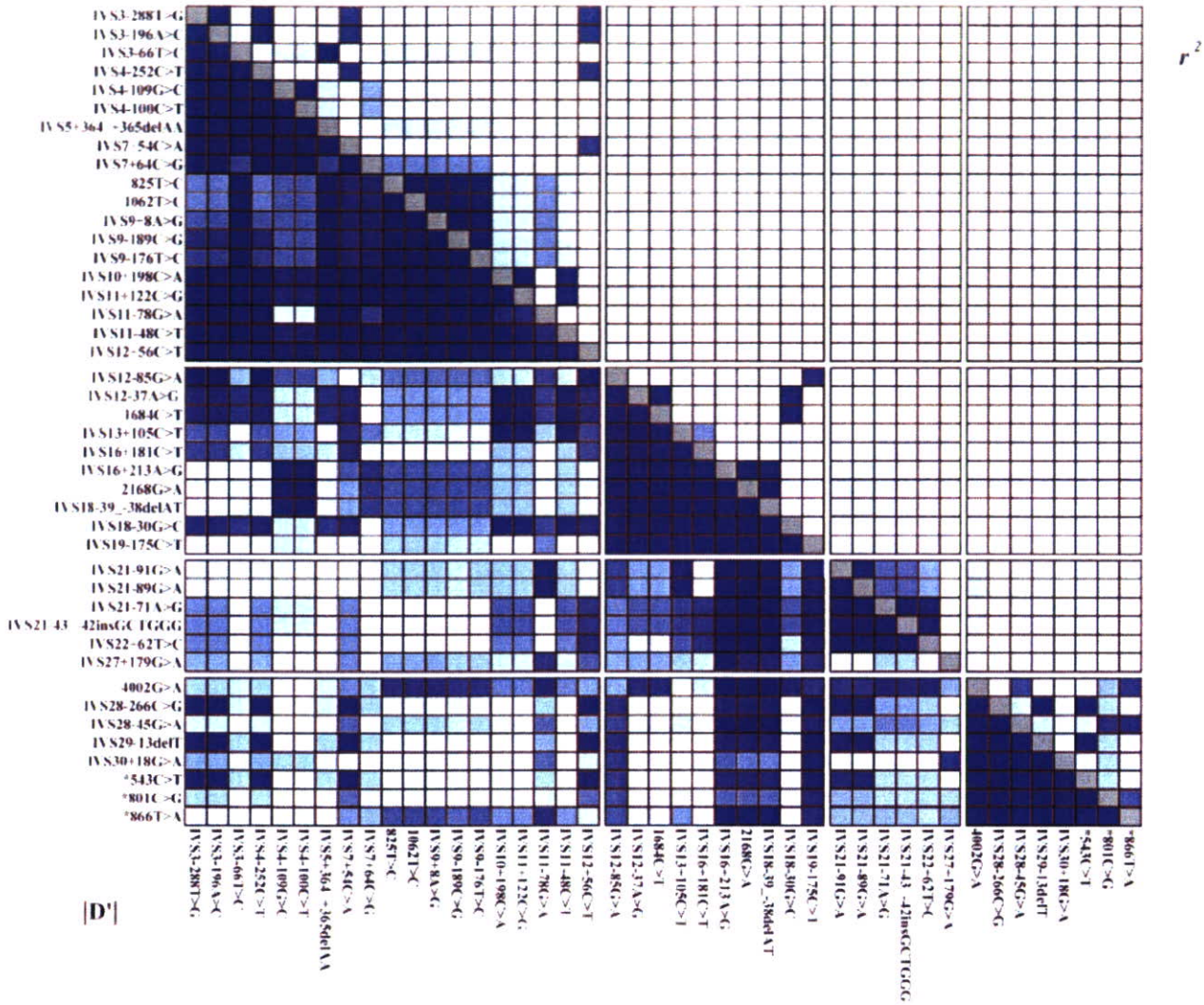

Fig. 1. Linkage disequilibrium (LD) analysis of *ABCC1*. Pairwise LD is expressed as r^2 (upper right) and |D'| (lower left) values (from 0 to 1) by 10-graded blue colors. Denser color represents closer linkage.

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[b]			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 C>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[a,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																	20	0.065	0.065
	*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 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]			1VS21 +11 C>G	1VS21 -91 G>A	1VS21 -89 G>A	1VS21 -71 A>G	1VS21 -43_-42ins GCTGGG	1VS22 +62 T>C	1VS22 -43 C>T	3173 G>A	1VS23 -131 G>C	3490 G>A	3550 G>A	1VS25 +114 C>T	1VS26 -30 C>G	1VS26 -24_-14 del TCTCTC CCTTC	3901 C>T	1VS27 +98 C>T	1VS27 +179 G>A			
Amino acid change										R1058Q		V1164I	E1184K				R1301C					
Haplotypes[b,c]	*1	*1a																		110	0.359	
		*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	0.987
		*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																		1	0.003	0.003
	*3	*3a																		1	0.003	0.003
	*4	*4a[e]																		1	0.003	0.003
	*5	*5a[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 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	
		*1b													85	0.278	
		*1c													58	0.190	
		*1d													26	0.085	
		*1e													18	0.059	0.997
		*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