Table 5 Hematological and non-hematological toxicity ($n=30$)

Toxicity	NCI-CTC Version 2.0, grade							
	0–1		2		3		4	
	n	%	*n*	%	*n*	%	*n*	%
Leukopenia	11	37	10	33	9	30	0	0
Neutropenia	10	33	4	13	14	47	2	7
Anemia	7	23	18	60	3	10	2	7
Thrombocytopenia	27	90	1	3	2	7	0	0
Febrile neutropenia	29	97	–		1	3	0	0
Nausea	27	90	3	10	0	0	–	
Fatigue	30	100	0	0	0	0	0	0
Neuropathy	20	67	10	33	0	0	0	0
Arthralgia	21	70	8	27	1	3	0	0
Rash	28	93	0	0	2	6	0	0
Infection	29	97	0	0	1	3	0	0
Arrhythmia	29	97	0	0	1	3	0	0
Alopecia	21	70	9	30	–		–	
AST/ALT	29	97	1	3	0	0	0	0
Hyponatremia	25	83	–		5	17	0	0

as second-line therapy as well if the patients had not been previously treated with the platinum/taxane combination chemotherapy.

Hotta at al. reported a meta-analysis based on abstracted data to compare the effect of carboplatin-based chemotherapy with that of cisplatin-based chemotherapy on overall survival, response rate, and toxicity in the first-line treatment of patients with advanced NSCLC [16]. The results indicated that combination chemotherapy consisting of cisplatin plus a third generation agent produced a significant survival benefit compared with carboplatin plus a third generation agent, although the toxicity profiles of the two modalities were quite different. Recently, Pignon et al. reported a pooled analysis from five randomized clinical trials of cisplatin-based chemotherapy in completely resected NSCLC patients [17]. Their analysis suggested that adjuvant cisplatin-based chemotherapy improved survival in patients with NSCLC. Based on the results of their meta-analysis, cisplatin-based chemotherapy should be recommended as first-line therapy for patients with advanced NSCLC. Moreover, in view of the results of our own study, we speculate that the combination of carboplatin plus paclitaxel may be suitable as second-line treatment for advanced NSCLC patients who had experienced progression after first-line cisplatin-based chemotherapy.

Care must be exercised in interpreting the favorable outcome in our study. One concern is that it was a single-institution phase II study, and therefore patient selection may have influenced the outcome. The responders to any of the prior chemotherapy regimens accounted for 50% of the 30 patients enrolled in this study, and about 80% of the patients had received only one prior chemotherapy regimen. The selection criteria, such as an ECOG PS of 0 or 1, may also have contributed to this favorable outcome. Another concern is that our study had included only five patients who were previously treated with chemotherapy using taxanes. Therefore, the efficacy of carboplatin plus paclitaxel as the secondary therapy after chemotherapy using taxanes is not clear. A further randomized study is warranted to be able to draw definitive conclusions about our results.

Conflict of interest statement

None declared.

Acknowledgement

This work was supported in part by a grant from the Bristol-Myers Squibb Company.

References

[1] Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical, trials. BMJ 1995;311:899–909.
[2] Bunn Jr PA, Kelly K. New chemotherapeutic agents prolong survival and improve quality of life in non-small cell lung cancer: a review of the literature and future directions. Clin Cancer Res 1998;4:1087–100.
[3] Huisman C, Smit EF, Giaccone G, Postmus PE. Second-line chemotherapy in relapsing or refractory non-small-cell lung cancer: a review. J Clin Oncol 2000;18:3722–30.
[4] Kelly K, Crowley J, Bunn Jr PA, et al. Randomized phase III trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non-small-cell lung cancer: a Southwest Oncology Group trial. J Clin Oncol 2001;19:3210–8.
[5] Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 2002;346:92–8.
[6] Shepherd FA, Dancey J, Ramlau R, et al. Prospective randomized trial of docetaxel versus best supportive care in patients with non-small-cell lung cancer previously

treated with platinum-based chemotherapy. J Clin Oncol 2000;18:2095–103.
[7] Fossella FV, DeVore R, Kerr RN, et al. Randomized phase III trial of docetaxel versus vinorelbine or ifosfamide in patients with advanced non-small-cell lung cancer previously treated with platinum-containing chemotherapy regimens. The TAX 320 Non-Small Cell Lung Cancer Study Group. J Clin Oncol 2000;18:2354–62.
[8] Belani CP, Lee JS, Socinski MA, et al. Randomized phase III trial comparing cisplatin-etoposide to carboplatin-paclitaxel in advanced or metastatic non-small cell lung cancer. Ann Oncol 2005;16:1069–75.
[9] Bunn Jr PA. Chemotherapy for advanced non-small-cell lung cancer: who, what, when, why? J Clin Oncol 2002;20:23–33.
[10] Hainsworth JD, Thompson DS, Greco FA. Paclitaxel by 1-hour infusion: an active drug in metastatic non-small-cell lung cancer. J Clin Oncol 1995;13:1609–14.
[11] Sculier JP, Berghmans T, Lafitte JJ, et al. A phase II study testing paclitaxel as second-line single agent treatment for patients with advanced non-small cell lung cancer failing after a first-line chemotherapy. Lung Cancer 2002;37:73–7.
[12] Hanna N, Shepherd FA, Fossella FV, et al. Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. J Clin Oncol 2004;22:1589–97.
[13] Shepherd FA, Pereira JR, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005;353:123–32.
[14] Scagliotti GV, De Marinis F, Rinaldi M, et al. Phase III randomized trial comparing three platinum-based doublets in advanced non-small-cell lung cancer. J Clin Oncol 2002;20:4285–91.
[15] Ohe Y, Ohashi Y, Kubota K, et al. Randomized phase III study of cisplatin plus irinotecan versus carboplatin plus paclitaxel, cisplatin plus gemcitabine, and cisplatin plus vinorelbine for advanced non-small-cell lung cancer: Four-Arm Cooperative Study in Japan. Ann Oncol 2007;18:317–23.
[16] Hotta K, Matsuo K, Ueoka H, et al. Meta-analysis of randomized clinical trials comparing Cisplatin to Carboplatin in patients with advanced non-small-cell lung cancer. J Clin Oncol 2004;22:3852–9.
[17] Pignon JP, Tribodet H, Scagliotti GV, et al. Lung Adjuvant Cisplatin Evaluation (LACE): a pooled analysis of five randomized clinical trials including 4584 patients. Proc Am Soc Clin Oncol 2006;24:7008.

0090-9556/07/3510-1865–1872$20.00
DRUG METABOLISM AND DISPOSITION
Copyright © 2007 by The American Society for Pharmacology and Experimental Therapeutics
DMD 35:1865–1872, 2007

Vol. 35, No. 10
15339/3257496
Printed in U.S.A.

Haplotypes and a Novel Defective Allele of *CES2* Found in a Japanese Population

Su-Ryang Kim, Kimie Sai, Toshiko Tanaka-Kagawa, Hideto Jinno, Shogo Ozawa, Nahoko Kaniwa, Yoshiro Saito, Akira Akasawa, Kenji Matsumoto, Hirohisa Saito, Naoyuki Kamatani, Kuniaki Shirao, Noboru Yamamoto, Teruhiko Yoshida, Hironobu Minami, Atsushi Ohtsu, Nagahiro Saijo, and Jun-ichi Sawada

Project Team for Pharmacogenetics (S.-R.K., K.Sa., H.J., S.O., N.Kan., Y.S., J.S.), Division of Biosignaling (K.Sa.), Division of Environmental Chemistry and Exposure Assessment (T.T.-K., H.J.), Division of Pharmacology (S.O.), Division of Medicinal Safety Sciences (N.Kan.), Division of Biochemistry and Immunochemistry (Y.S., J.S.), National Institute of Health Sciences, Tokyo, Japan; Department of Allergy and Immunology, National Research Institute for Child Health and Development (K.M., H.S.), National Children's Medical Center (A.A.), National Center for Child Health and Development, Tokyo, Japan; Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan (N.Kam.); Division of Internal Medicine (K.Sh., N.Y.), National Cancer Center Hospital, Genetics Division (T.Y.), National Cancer Center Research Institute, Tokyo, Japan; and Division of Oncology/Hematology (H.M.), Division of GI Oncology/Digestive Endoscopy (A.O.), Deputy Director (N.S.), National Cancer Center Hospital East, Chiba, Japan

Received February 23, 2007; accepted July 17, 2007

ABSTRACT:

Human carboxylesterase 2 (hCE-2) is a member of the serine esterase superfamily and is responsible for hydrolysis of a wide variety of xenobiotic and endogenous esters. hCE-2 also activates an anticancer drug, irinotecan (7-ethyl-10-[4-(1-piperidino)-1-piperidino]-carbonyloxycamptothecin, CPT-11), into its active metabolite, 7-ethyl-10-hydroxycamptothecin (SN-38). In this study, a comprehensive haplotype analysis of the *CES2* gene, which encodes hCE-2, in a Japanese population was conducted. Using 21 single nucleotide polymorphisms (SNPs), including 4 nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$, *2), 424G>A ($Val^{142}Met$, *3), 1A>T ($Met^{1}Leu$, *5), and 617G>A ($Arg^{206}His$, *6), and a SNP at the splice acceptor site of intron 8 (IVS8-2A>G, *4), 20 haplotypes were identified in 262 Japanese subjects. In 176 Japanese cancer patients who received irinotecan, associations of *CES2* haplotypes and changes in a pharmacokinetic parameter, (SN-38 + SN-38G)/CPT-11 area under the plasma concentration curve (AUC) ratio, were analyzed. No significant association was found among the major haplotypes of the **1* group lacking nonsynonymous or defective SNPs. However, patients with nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$) or 1A>T ($Met^{1}Leu$), showed substantially reduced AUC ratios. In vitro functional characterization of the SNPs was conducted and showed that the 1A>T SNP affected translational but not transcriptional efficiency. These findings are useful for further pharmacogenetic studies on CES2-activated prodrugs.

Human carboxylesterases are members of the serine esterase superfamily and are responsible for hydrolysis of a wide variety of xenobiotic and endogenous esters. They metabolize esters, thioesters, carbamates, and amides to yield soluble acids and alcohols or amines (Satoh and Hosokawa, 1998; Satoh et al., 2002). In the human liver, two major isoforms of carboxylesterase, hCE-1 and hCE-2, have been identified (Shibata et al., 1993; Schwer et al., 1997). hCE-2 is a 60-kDa monomeric enzyme with a pI value of approximately 4.9 and shares 48% amino acid sequence identity with hCE-1 (Pindel et al., 1997; Schwer et al., 1997; Takai et al., 1997). The *CES2* gene, which encodes hCE-2, is located on chromosome 16q22.1 and consists of 12 exons. Distribution of hCE-2 is relatively limited to several tissues, such as the small intestine, colon, heart, kidney, and liver, whereas hCE-1 is ubiquitously expressed (Satoh et al., 2002; Xie et al., 2002).

Although both hCE-1 and hCE-2 show broad substrate specificities, hCE-2 is relatively specific for heroin, cocaine (benzoyl ester), 6-acetylmorphine, procaine, and oxybutynin (Pindel et al., 1997; Takai et al., 1997; Satoh et al., 2002). In addition, hCE-2 is reported to play a major role in the metabolic activation of the antitumor drug irinotecan (7-ethyl-10-[4-(1-piperidino)-1-piperidino]-carbonyloxycamptothecin; CPT-11). Irinotecan is a water-soluble derivative of the plant alkaloid camptothecin and is widely used for treatment of several types of cancer. Irinotecan is converted to 7-ethyl-10-hydroxy-

This study was supported in part by the Program for the Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation and by a Health and Labour Science Research Grant from the Ministry of Health, Labour and Welfare of Japan.

Article, publication date, and citation information can be found at http://dmd.aspetjournals.org.

doi:10.1124/dmd.107.015339.

ABBREVIATIONS: hCE-1, human carboxylesterase 1; hCE-2, human carboxylesterase 2 (EC 3.1.1.1); irinotecan, 7-ethyl-10-[4-(1-piperidino)-1-piperidino]-carbonyloxycamptothecin, CPT-11; SN-38, 7-ethyl-10-hydroxycamptothecin; SN-38G, SN-38 glucuronide; SNP, single nucleotide polymorphisms; PCR, polymerase chain reaction; LD, linkage disequilibrium; 5-FU, 5-fluorouracil; MMC, mitomycin C; AUC, area under plasma concentration curve; RT, reverse transcriptase; UTR, untranslated region; ORF, open reading frame.

camptothecin (SN-38), a topoisomerase inhibitor, by carboxylesterases (Humerickhouse et al., 2000) and further conjugated by hepatic uridine diphosphate glucuronosyltransferase to form the inactive metabolite SN-38 glucuronide (SN-38G) (Iyer et al., 1998). To a lesser extent, irinotecan is also converted to 7-ethyl-10-[4-*N*-(5-aminopentanoic acid)-1-piperidino]carbonyloxycamptothecin and 7-ethyl-10-(4-amino-1-piperidino)carbonyloxycamptothecin by cytochrome P450 3A4 (Dodds et al., 1998; Santos et al., 2000). Irinotecan and its metabolites are excreted by the efflux transporters, ABCB1 (P-glycoprotein), ABCC2 (canalicular multispecific organic anion transporter), and ABCG2 (breast cancer resistance protein), via a hepatobiliary pathway (Mathijssen et al., 2001). Although irinotecan metabolism is rather complex, hCE-2 is a key enzyme that determines the plasma levels of the active metabolite SN-38.

Hepatic hCE-2 activity toward irinotecan varies 3-fold in microsomes obtained from a panel of human livers (Xu et al., 2002). The activity loosely correlates with hCE-2 protein levels, but some microsomal samples showed unanticipated deviating activities. This result might be caused by genetic polymorphisms, such as single nucleotide polymorphisms (SNPs) in the *CES2* gene. Several SNPs and haplotypes have been reported for the *CES2* gene (Charasson et al., 2004; Marsh et al., 2004; Wu et al., 2004), and large ethnic differences in *CES2* SNP frequencies are found among Europeans, Africans, and Asian-Americans (Marsh et al., 2004).

Previously, 12 exons and their flanking regions of *CES2* were sequenced from 153 Japanese subjects, who received irinotecan or steroidal drugs, and 12 novel SNPs, including the nonsynonymous SNP, 100C>T ($Arg^{34}Trp$), and the SNP at the splice acceptor site of intron 8 (IVS8-2A>G) were found (Kim et al., 2003). In vitro functional characterization of these SNPs and an additional nonsynonymous SNP, 424G>A ($Val^{142}Met$), suggested that the ^{34}Trp and ^{142}Met variants were defective, and that IVS8-2G might be a low-activity allele (Kubo et al., 2005). In the present study, the same regions were sequenced from an additional 109 subjects (a total of 262 patients), and their haplotypes/diplotypes were determined/inferred. Then, associations between the haplotypes and pharmacokinetic parameters of irinotecan and its metabolites were analyzed for 177 cancer patients who were given irinotecan. Functional characterization of novel SNPs 1A>T and 617G>A, which were found in this study, was also performed by using a transient expression system with COS-1 cells.

Materials and Methods

Chemicals. Irinotecan, SN-38, and SN-38G were kindly supplied by Yakult Honsha Co. Ltd. (Tokyo, Japan).

Patients. A total of 262 Japanese subjects analyzed in this study consisted of 85 patients with allergies who received steroidal drugs and 177 patients with cancer who received irinotecan. The ethical review boards of the National Cancer Center, National Center for Child Health and Development, and National Institute of Health Sciences approved this study. Written informed consent was obtained from all participants.

DNA Sequencing. Total genomic DNA was extracted from blood leukocytes or Epstein-Barr virus-transformed lymphocytes and used as a template in the polymerase chain reaction (PCR). Sequence data of the *CES2* gene from 72 patients and 81 cancer patients were described previously (Kim et al., 2003). In addition, the *CES2* gene was sequenced from 13 allergic patients and 96 cancer patients. Amplification and sequencing of the *CES2* gene were performed as described previously (Kim et al., 2003). Rare SNPs found in only one heterozygous subject were confirmed by sequencing PCR fragments produced by amplification with a high-fidelity DNA polymerase KOD-Plus (Toyobo, Tokyo, Japan). GenBank accession number NT_010498.15 was used as the reference sequence.

Linkage Disequilibrium and Haplotype Analyses. LD analysis was performed by the SNPAlyze software (version 5.1; Dynacom Co., Yokohama, Japan), and a pairwise two-dimensional map between SNPs was obtained for the D' and rho square (r^2) values. All allele frequencies were in Hardy-Weinberg equilibrium. Some haplotypes were unambiguously assigned in the subjects with homozygous variations at all sites or a heterozygous variation at only one site. Separately, the diplotype configurations (combinations of haplotypes) were inferred by LDSUPPORT software, which determines the posterior probability distribution of the diplotype configuration for each subject on the basis of estimated haplotype frequencies (Kitamura et al., 2002). The haplotype groups were numbered according to the allele nomenclature systems suggested by Nebert (2000). The haplotypes harboring nonsynonymous or defective alleles were assigned as haplotype groups **2* to **6*. The subgroups were described as the numbers plus small alphabetical letters.

Administration of Irinotecan and Pharmacokinetic Analysis. The demographic data and eligibility criteria for 177 cancer patients who received irinotecan in the National Cancer Center Hospitals (Tokyo and Chiba, Japan) were described elsewhere (Minami et al., 2007).

Each patient received a 90-min i.v. infusion at doses of 60 to 150 mg/m^2, which varied depending on regimens/coadministered drugs: i.e., irinotecan dosages were 100 or 150 mg/m^2 for monotherapy and combination with 5-FU, 150 mg/m^2 for combination with mitomycin C (MMC), and 60 (or 70) mg/m^2 for combination with platinum anticancer drugs. Heparinized blood was collected before administration of irinotecan and at 0 min (end of infusion), 20 min, 1 h, 2 h, 4 h, 8 h, and 24 h after infusion. Plasma concentrations of irinotecan, SN-38, and SN-38G were determined as described previously (Sai et al., 2002). The AUCs from time 0 to infinity of irinotecan and its metabolites were calculated as described (Sai et al., 2004). Associations between genotypes and pharmacokinetic parameters including the AUC ratio (SN-38 + SN-38G)/CPT-11 were evaluated in 176 patients in whom pharmacokinetic parameters were obtained.

Construction of Expression Plasmids. The coding region of CES2L (long form) cDNA starts at an additional ATG translation initiation codon located 192 nucleotides upstream of the conventional ATG codon (Wu et al., 2003) and encodes a 623-amino acid protein found in the National Center for Biotechnology Information database (NP_003860.2). The wild-type CES2L cDNA was amplified by PCR from Human Liver QUICK-Clone cDNA (Clontech, Mountain View, CA) using *CES2*-specific primers, 5′-CACCCACCTATGACTGCTCA-3′ and 5′-AGGGAGCTACAGCTCTGTGT-3′. The PCR was performed with 1 unit of the high-fidelity DNA polymerase KOD-Plus and a 0.5 μM concentration of the *CES2* specific primers. The PCR conditions were 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 60°C for 30 s, and 68°C for 3 min and then a final extension at 68°C for 5 min. The PCR products were cloned into the pcDNA3.1 vector by a directional TOPO cloning procedure (Invitrogen, Carlsbad, CA), and the sequences were confirmed in both directions. The resultant plasmid was designated pcDNA3.1/CES2L-WT. The 1A>T variation was introduced into pcDNA3.1/CES2L-WT by using a QuikChange Multi site-directed mutagenesis kit (Stratagene, La Jolla, CA) with the 5′-phosphorylated oligonucleotide, 5′-phospho-GAGACCAGCGAGCCGACCTTGCGGCTGCACAGACTTCG-3′ (the substituted nucleotide is underlined). The sequence of the variant cDNA was confirmed in both strands, and the resultant plasmid was designated pcDNA3.1/CES2L-A1T. Expression plasmids for the short-form wild-type (CES2S) and $Arg^{206}His$ variant CES2 were prepared and introduced into COS-1 cells according to the method described previously (Kubo et al., 2005).

Expression of Wild-Type and Variant CES2 Proteins in COS-1 Cells. Expression of wild-type and variant CES2 proteins in COS-1 cells was examined as described previously (Kubo et al., 2005). In brief, microsomal fractions (30 μg of protein/lane) or postmitochondrial fractions (0.4 μg of protein/lane) were separated by 8% SDS-polyacrylamide gel electrophoresis and transferred onto a nitrocellulose membrane. Immunochemical detection of each type of CES2 protein was performed using rabbit anti-human CES2 antibody raised against a peptide antigen (residues 539–555, KKALPQKIQELEEPEER) (diluted 1:1500). To verify that the samples were evenly loaded, the blot was subsequently treated with a stripping buffer and reprobed with polyclonal anti-calnexin antibody (diluted 1:2000; Stressgen Biotechnologies Corp., San Diego, CA). Visualization of these proteins was achieved with horseradish peroxidase-conjugated donkey anti-rabbit IgG (1:4000) and the Western Lightning Chemiluminescence Reagent Plus (PerkinElmer Life and Analytical Sciences, Boston, MA). Protein band densities were quantified with Diana III

and ZERO-Dscan software (Raytest, Straubenhardt, Germany). The relative expression levels are shown as the means ± S.D. of three separate transfection experiments.

Determination of CES2 mRNA by Real-Time RT-PCR. Total RNA was isolated from transfected COS-1 cells using the RNeasy Mini Kit (QIAGEN, Tokyo, Japan). After RNase-free DNase treatment of samples to minimize plasmid DNA contamination, first-strand cDNA was prepared from 1 μg of total RNA using the High-Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA) with random primers. Real-time PCR assays were performed with the ABI7500 Real Time PCR System (Applied Biosystems) using the TaqMan Gene Expression Assay for *CES2* (Hs01077945_m1; Applied Biosystems) according to the manufacturer's instructions. The relative mRNA levels were determined using calibration curves obtained from serial dilutions of the pooled wild-type CES2 cDNA. Samples without reverse transcriptase were routinely included in the RT-PCR reactions to measure possible contributions of contaminating DNA, which was usually less than 1% of the mRNA-derived amplification. Transcripts of β-actin were quantified as internal controls using TaqMan β-Actin Control Reagent (Applied Biosystems), and normalization of CES2 mRNA levels were based on β-actin concentrations.

Enzyme Assay. CPT-11 hydrolyzing activity of the postmitochondrial supernatants (microsomal fraction plus cytosol) was assayed over the substrate concentration range of 0.25 to 50 μM as described previously (Kubo et al., 2005), except that the hydrolysis product, SN-38, was determined by the high-performance liquid chromatography method of Hanioka et al. (2001).

Statistical Analysis. Statistical analysis of the differences in the AUC ratios among *CES2* diplotypes, coadministered drugs. or irinotecan dosages was performed using the Kruskal-Wallis test, Mann-Whitney test, or Spearman rank correlation test (Prism 4.0, GraphPad Software, Inc., San Diego, CA). The *t* test (Prism 4.0) was applied to the comparison of the average values of protein expression and mRNA levels between wild-type and variant CES2.

Results

***CES2* Variations Detected in a Japanese Population.** Previously, the promoter region, all 12 exons, and their flanking introns of the *CES2* gene were sequenced from 72 allergic patients and 81 cancer patients and resulted in the identification of 12 novel SNPs (Kim et al., 2003). Additionally, the same region of *CES2* was sequenced from 13 allergic patients and 96 cancer patients. A total of 21 SNPs were found in 262 Japanese subjects (Table 1). Novel SNPs found in this study were −1233T>C, 1A>T, IVS2-71C>G, IVS7 + 27G>A, and IVS9 + 78C>T, but their frequency was low (0.002, identified in a single heterozygous subject for each SNP). The SNP 1A>T is nonsynonymous (M1L) and results in a substitution of the translation initiation codon ATG to TTG in the *CES2* gene. The other novel SNPs were located in the introns or the 5′-flanking region.

The nonsynonymous SNP 424G>A (V142M) reported by our group (Kubo et al., 2005) and another nonsynonymous SNP 617G>A (R206H) published in the dbSNP (rs8192924) and JSNP (ssj0005417) databases were found at a frequency of 0.002. Recently, several noncoding SNPs in *CES2* were also reported (Kim et al., 2003; Charasson et al., 2004; Marsh et al., 2004; Wu et al., 2004). Among them, the three SNPs, −363C>G in the 5′-UTR, IVS10-108(IVS10 + 406)G>A in intron 10, and 1749(*69)A>G in the 3′-UTR of exon 12, were found at frequencies of 0.031, 0.269, and 0.239, respectively, in this study.

LD and Haplotype Analysis. Using the detected SNPs, LD analysis was performed, and the pairwise values of r^2 and D' were obtained. A perfect linkage (r^2 = 1.00) was observed between SNPs −363C>G and IVS10-87G>A. A close association (r^2 = 0.85) was found between SNPs IVS10-108G>A and 1749A>G. Other associations were much lower (r^2 < 0.1). Therefore, the entire *CES2* gene was analyzed as one LD block. The determined/inferred haplotypes are summarized in Fig. 1 and are shown as numbers plus small

TABLE 1

Summary of SNPs in the CES2 gene in a Japanese population

SNP Identification			Location	Position		Nucleotide Change and Flanking Sequences (5′ to 3′)	Amino Acid Change	Allele Frequency
This Study	NCBI (dbSNP)	JSNP		NT_010498.15	From the Translational Initiation Site or from the Nearest Exon			
MPJ6_CS2001			5′-Flanking	20582067	−1671[a]	CTGGAACAACTCG/CCTCCCCTCGGAA		0.010
MPJ6_CS2002			5′-Flanking	20582484	−1254[a]	AACCACCACCGCT/CGATCCTAGCAGG		0.002
MPJ6_CS2016[b]			5′-Flanking	20582505	−1233[a]	CAGGCGTGCGCTT/CCGCTCCAACCC		0.002
MPJ6_CS2003			Exon 1 (5′-UTR)	20582979	−759[a]	AAATGTTTGTCAA/GGTGGATAAATGA		0.006
MPJ6_CS2004	rs11075646		Exon 1 (5′-UTR)	20583375	−363[a]	CCTCCTATCGATC/GCCCCAGCGCGCT		0.031
MPJ6_CS2017[b]			Exon 1	20583738	1[a]	AGCGAGCCGACCA/TTGCGGCTGCACA	Met[1]Leu	0.002
MPJ6_CS2005			Exon 2	20586162	100[a]	GCCAGTCCCATCC/TGGACCACACACA	Arg[34]Trp	0.002
MPJ6_CS2021[b]			Intron 2	20587248	IVS2–71	GGTGGCTGGGAGC/GACCTCTGAACCC		0.002
MPJ6_CS2015			Exon 4	20588325	424[a]	TGATTTCCCCAGG/ATGATGGTGTGGA	Val[142]Met	0.002
MPJ6_CS2006			Intron 4	20588486	IVS4 + 29	GCTGGGCAACCCG/AGGCTGAGCGGGG		0.002
MPJ6_CS2007			Exon 5	20588560	579[a]	CAAGCACGCAACC/TGGCAACTGGGGC	Thr[193]Thr (silent)	0.002
MPJ6_CS2018	rs8192924	ssj0005417	Exon 5	20588598	617[a]	TGGCTGCACTACG/ACTGGTCCAGCA	Arg[206]His	0.002
MPJ6_CS2008			Exon 5	20588746	765[a]	CATGGAGAGTGGC/TGTGGCCCTCTG	Gly[255]Gly (silent)	0.002
MPJ6_CS2009			Intron 5	20589157	IVS5–69	CCTGTTCTTGGCC/TAGGGCCTTGGGC		0.017
MPJ6_CS2019[b]			Intron 7	20589775	IVS7 + 27	AAGCCCACAAGTG/ACCTGGGGAGCCC		0.002
MPJ6_CS2010			Intron 7	20589845	IVS7–25	CCCATCCCCAGCT/AACAGACTCTCTC		0.002
MPJ6_CS2011			Intron 8	20590205	IVS8–2	TCCACCTGGGGTA/GGATGTTGCCTCC	Splicing defect	0.002
MPJ6_CS2020[b]			Intron 9	20590429	IVS9 + 78	ACCTGCTGCTGTC/TCGGGTCAGCACT		0.002
MPJ6_CS2012	rs2241409	IMS-JST1013275	Intron 10	20591293	IVS10–108	GGAAGAAAAAGCG/AGAGAAGCAGGAC		0.269
MPJ6_CS2013	rs28382825		Intron 10	20591314	IVS10–87	GGACTGGGGACCG/AAGGTCTCGGGGG		0.031
MPJ6_CS2014	rs8192925	ssj0005418	Exon 12 (3′-UTR)	20592196	1749 (*69)[a]	GTGCCCACACACA/GCCCACTAAGGAG		0.239

[a] A of the conventional translation initiation codon ATG in *CES2* (GenBank Y09616) is numbered 1, and the number in the parentheses indicates the position from the termination codon TGA.

[b] Novel variations detected in this study.

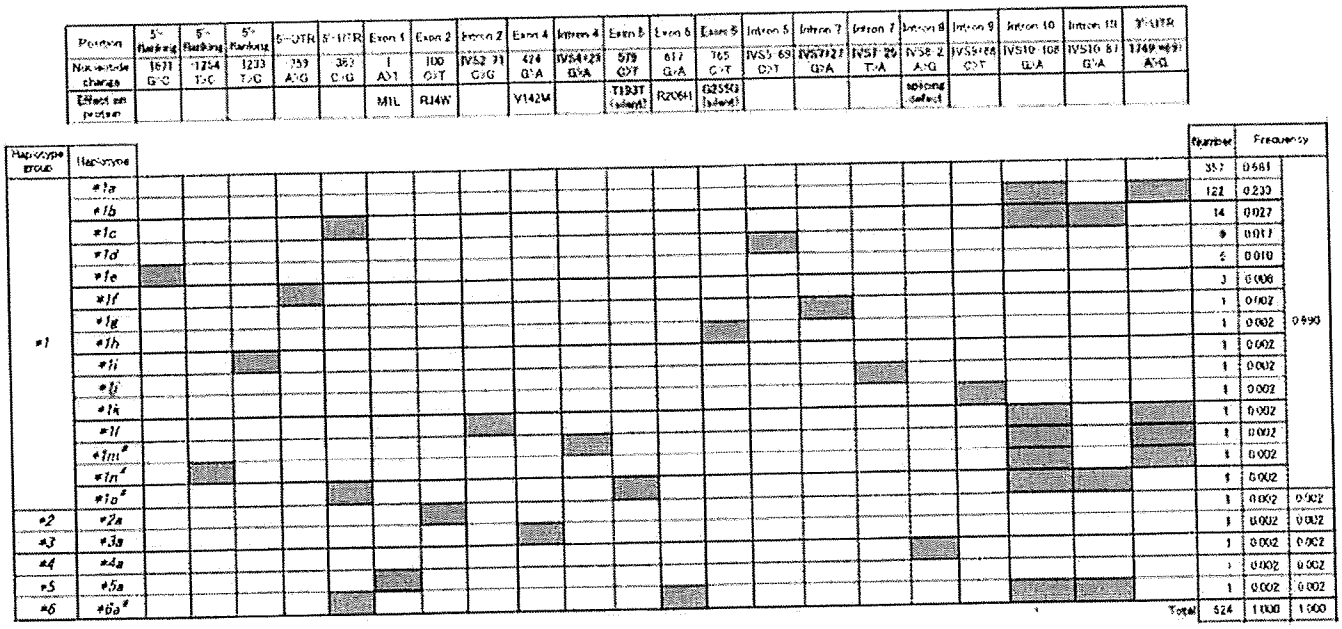

FIG. 1. Haplotypes of the *CES2* gene assigned for 262 Japanese subjects. The haplotypes assigned are described with lower case numbers and alphabetical letters. #, this haplotype was inferred in only one patient and is thus ambiguous.

alphabetical letters. Our nomenclature of haplotypes is distinct from those of previous studies (Charasson et al., 2004; Marsh et al., 2004; Wu et al., 2004). In this study, the haplotypes without amino acid changes and splicing defects were defined as the **1* group. The haplotypes harboring the nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$), 424G>A ($Val^{142}Met$), 1A>T ($Met^{1}Leu$), and 617G>A ($Arg^{206}His$), were assigned as haplotypes **2*, **3*, **5*, and **6*, respectively. In addition, the haplotype harboring a SNP at the splice acceptor site of intron 8 (IVS8-2A>G) was assigned as haplotype **4*. Several haplotypes were first unambiguously assigned by homozygous variations at all sites (**1a* and **1b*) or heterozygous variation at only one site (**1d* to **1l*, **2a*, **3a*, **4a*, and **5a*). Separately, the diplotype configurations (combinations of haplotypes) were inferred by LDSUPPORT software. The additionally inferred haplotypes were **1c* and **1m* to **1o*. The most frequent haplotype was **1a* (frequency, 0.681), followed by **1b* (0.233), **1c* (0.027), and **1d* (0.017). The frequencies of the other haplotypes were less than 0.01.

Association between *CES2* Genotypes and Irinotecan Pharmacokinetics. Next, the relationships between the *CES2* genotype and AUC ratio [(SN-38 + SN-38G)/CPT-11], a parameter of in vivo CES activity (Cecchin et al., 2005), in irinotecan-administered patients were investigated. The diplotype distribution of 176 patients, who received irinotecan and were analyzed for the AUC ratio, was similar to that of the 262 subjects. We examined preliminarily the effects of irinotecan dosage and comedication on the AUC ratio and obtained significant correlations of irinotecan dosage (Spearman $r = -0.559$, $p < 0.0001$) and comedication ($p < 0.0001$, Kruskal-Wallis test) with the AUC ratios. Because irinotecan dosages also depended on the drugs coadministered (see *Materials and Methods*), we finally stratified the patients with the coadministered drugs. As shown in Fig. 2, no significant differences in the median AUC ratios were observed among the **1* diplotypes in each group (p values in the Kruskal-Wallis test among **1a/*1a*, **1a/*1b*, and **1b/*1b* were 0.260, 0.470, 0.129, and 0.072 for irinotecan alone, with 5-FU, with MMC and with platinum, respectively.). The relatively rare haplotype **1c*, which harbors −363C>G, did not show any associations with altered AUC ratio ($p = 0.756$ for irinotecan alone and $p = 0.230$ for irinotecan with platinum, Mann-Whitney test).

To estimate the effects of nonsynonymous SNPs on the metabolism of irinotecan, the AUC ratios in the patients carrying nonsynonymous SNPs were compared with the median AUC ratio of the **1/*1* patients. Three nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$, **2*), 1A>T ($Met^{1}Leu$, **5*), and 617G>A ($Arg^{206}His$, **6*), and a SNP at the splice acceptor site of intron 8 (IVS8-2A>G, **4*) were found in 177 patients who received irinotecan. These SNPs were single heterozygotes. The AUC ratios of the patients with **2a/*1a* (0.17) and **5a/*1a* (0.10) in the monotherapy group were 60 and 36%, respectively, of the median value for the **1/*1* group (0.28) and substantially lower than the 25th percentile of the **1/*1* group (0.23) (Fig. 2). It must be noted that the **5a/*1a* patient had an extremely low AUC ratio. The AUC ratio of the **6* heterozygote who received cisplatin (0.25) was lower than the median value (0.37) but within the range for the **1/*1* group treated with platinum-containing drugs (Fig. 2). Regarding the effect of the heterozygous **4*, the AUC ratio (0.40) was not different from the median AUC ratio of the **1/*1* treated with platinum-containing drugs. To elucidate the effects of two novel amino acid substitutions, $Met^{1}Leu$ (**5*) and $Arg^{206}His$ (**6*), the functional analysis was conducted in vitro.

In Vitro Functional Analysis of the $Met^{1}Leu$ Variant. To clarify the functional significance of the novel variant $Met^{1}Leu$ (**5*), the protein expression level of CES2 carrying the nonsynonymous SNP 1A>T was examined. Wu et al. (2003) reported that transcription of CES2 mRNA was initiated from several transcriptional start sites, resulting in the expression of three *CES2* transcripts. Two longer transcripts carry a potential inframe translational initiation codon ATG at −192 that can encode an open reading frame (ORF) extending 64 residues at the amino terminus, as shown in the reference sequence in the National Center for Biotechnology Information database (NP_003860.2). Therefore, the expression of the CES2 protein from the long *CES2* ORF (CES2L), which encodes a potential 623 residue protein, was analyzed. Western analysis of membrane fraction proteins obtained from COS-1 cells

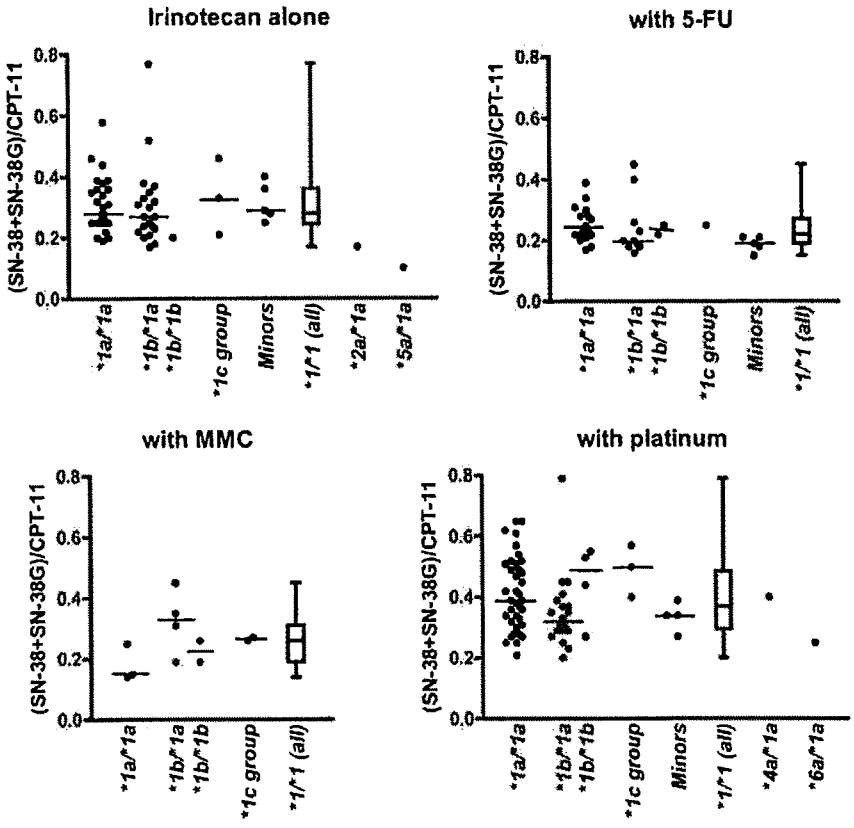

FIG. 2. Relationship between the *CES2* diplotypes and (SN-38 + SN-38G)/CPT-11 AUC ratios in Japanese cancer patients who received irinotecan. Each point represents an individual patient, and the median value in each genotype is shown with a horizontal bar. Distribution of the **1* group is shown by a box representing the 25th to 75th percentiles with a line at the median and bars representing the highest and lowest values. The **1c* group consists of **1c/*1a* and **1c/*1b*. "Minors" represents the heterozygous patients bearing minor **1* haplotypes (**1d, *1e, *1f, *1g, *1k,* and **1m*). Irinotecan alone, irinotecan monotherapy (n = 58); with 5-FU, combination therapy with 5-FU including tegafur (n = 35); with MMC, combination therapy with mitomycin C (n = 11); with platinum, combination therapy with either cisplatin (n = 62), cisplatin plus etoposide (n = 2), or carboplatin (n = 8).

transfected with the expression plasmid pcDNA3.1/CES2L-WT showed that the mobility (approximately 60 kDa) of the protein product from the CES2L cDNA was the same as that from the CES2S cDNA, which encodes a 559 residue protein (Kubo et al., 2005), and the CES2 protein in the human liver microsome (Fig. 3A). Western blot analysis of whole cell extracts also showed that CES2L yielded a single 60-kDa protein product (data not shown), indicating that translation of *CES2* was initiated from the second ATG codon of the *CES2L* ORF but not from the inframe translation initiation codon located at −192.

When the effect of the 1A>T SNP on the expression of the CES2 protein was examined by Western blotting (Fig. 3A), the relative expression levels of CES2 protein from cells transfected with plasmid pcDNA3.1/CES2L-A1T were 11.7 ± 2.4% (p = 0.0003) of the wild type. The mRNA expression levels determined by the TaqMan real-time RT-PCR assay were similar between the wild-type and variant CES2L cDNAs in COS-1 cells (Fig. 4A), indicating that the 1A>T SNP affects translational but not transcriptional efficiency. Thus, the $Met^{1}Leu$ variant was functionally deficient.

In Vitro Functional Analysis of the $Arg^{206}His$ Variant. The known nonsynonymous SNP 617G>A changes an arginine to a histidine at residue 206. Western blot analysis of the postmitochondrial supernatant (including microsomes and cytosol) fractions obtained from COS-1 cells transfected with wild-type (CES2S) and $Arg^{206}His$ variant CES2-expressing plasmids showed that the protein expression level of the $Arg^{206}His$ variant was approximately 82 ± 7% (p = 0.017) of the wild-type (Fig. 3B). No significant differences in the mRNA expression levels determined by the TaqMan real-time RT-PCR assay were observed between the wild-type and 617G>A variant CES2s (82 ± 7%, p = 0.06) (Fig. 4B). Table 2 summarizes the apparent kinetic parameters for CPT-11 hydrolysis of wild-type and $Arg^{206}His$ variant CES2.

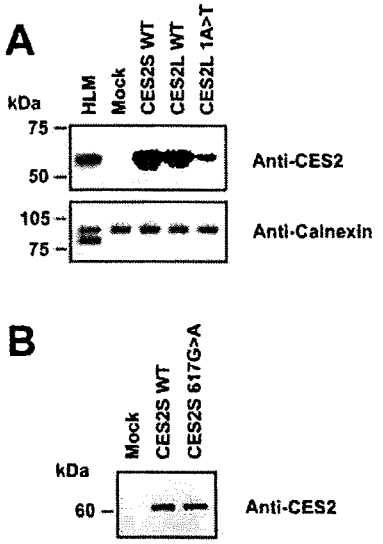

FIG. 3. Expression of CES2 protein from the wild-type and 1A>T (A) and 617G>A (B) variant *CES2* genes in COS-1 cells. Membrane fraction (A) or the postmitochondrial supernatant (B) from the cDNA-transfected cells was subjected to SDS-polyacrylamide gel electrophoresis, followed by transfer to the nitrocellulose membrane. Detection of CES2 and calnexin was performed with rabbit anti-human CES2 antiserum (A and B) and a rabbit anti-human calnexin antiserum (A) and horseradish peroxidase-conjugated donkey anti-rabbit IgG antibody as described under *Materials and Methods*. A representative result from one of three independent experiments is shown. HLM, human liver microsomes.

Although a slight difference in the K_m values was obtained with statistical significance ($p < 0.01$), the kinetic parameters (V_{max} and V_{max}/K_m) were not significantly different when normalized by protein expression levels.

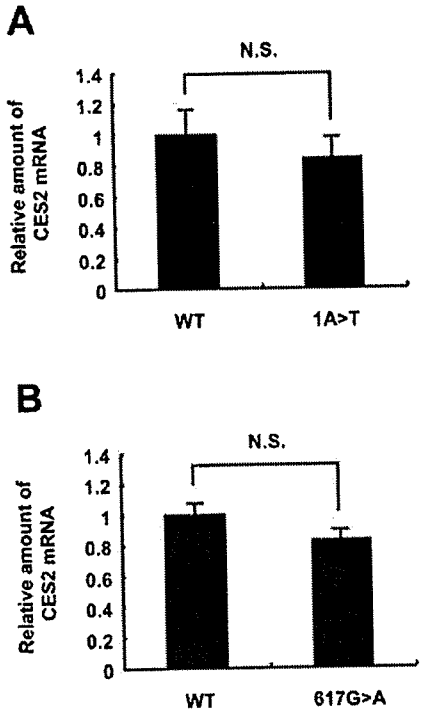

FIG. 4. Quantification of CES2 mRNA by TaqMan real-time RT-PCR in COS-1 cells transfected with wild-type (WT) and 1A>T (A) and 617G>A (B) variants. CES2 mRNA expression levels after 48 h were normalized with β-actin mRNA levels, and the mean level of the wild-type was set as 1.0. The results indicate the mean ± S.D. from three independent preparations. No significant difference in mRNA level was observed between the wild-type and variants ($p = 0.21$ an 0.06 in A and B, respectively).

Discussion

The present study provides comprehensive data on the haplotype analysis of the *CES2* gene, which encodes human carboxylesterase 2. From additional sequence analysis, a total of 21 SNPs including 4 nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$), 424G>A ($Val^{142}Met$), 1A>T ($Met^{1}Leu$), and 617G>A ($Arg^{206}His$), and a SNP at the splice acceptor site of intron 8 (IVS8-2A>G) were found in 262 Japanese subjects. Among the nonsynonymous SNPs, in vitro functional analysis of the two nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$) and 424G>A ($Val^{142}Met$), has already been performed to identify effects of these SNPs on expression levels and carboxylesterase activity. Kubo et al. (2005) showed that $Arg^{34}Trp$ and $Val^{142}Met$ variants had little carboxylesterase activity toward irinotecan, *p*-nitrophenyl acetate, and 4-methylumbelliferyl acetate, whereas expression levels of these variants were higher than those of the wild-type. An in vitro splicing assay using the *CES2* minigene carrying SNP IVS8-2A>G showed that IVS8-2A>G yielded mostly aberrantly spliced transcripts, resulting in the production of truncated CES2 proteins. These results have suggested that 100C>T ($Arg^{34}Trp$), 424G>A ($Val^{142}Met$), and IVS8-2A>G are functionally defective SNPs.

A novel SNP 1A>T found in this study changes the translation start codon ATG to TTG. Wu et al. (2003) identified three transcription start sites of *CES2*, resulting in the synthesis of three transcripts with either 78, 629, or 1187 nucleotides in the 5′-UTR. Another inframe ATG codon is present 192 nucleotides upstream of the conventional translational initiation codon, and two longer transcripts with 629 and 1187 nucleotides in the 5′-UTR can encode an ORF with 64 additional residues at the amino terminus (NP_003860.2). However, as shown in Fig. 3A, our in vitro experiment for the expression of CES2 showed that translation of CES2 mRNA started from the previously reported ATG codon but not from the inframe ATG codon at −192, when transiently expressed from the wild-type *CES2L* cDNA encoding a potential 623-amino acid CES2 protein in COS-1 cells. In vertebrate mRNAs, a purine residue in position −3 (A of the translational start codon is +1) is highly conserved and required for efficient translation (Kozak, 1991). The surrounding sequences of both ATG codons were accATGc for the functional ATG codon and cctATGa for the potential inframe ATG codon at −192. Thus, it is likely that their efficiencies of translation initiation depend on the flanking sequences of the translational start codon ATG.

When the expression levels between the wild-type and 1A>T variant were compared, the protein level of 1A>T was drastically reduced without changes in the mRNA levels, suggesting that the reduced protein level of the 1A>T variant might have been caused by its reduced translation initiation. It has been reported that alterations of the translational start codon ATG to TTG diminish or reduce the translation of growth hormone receptor (Quinteiro et al., 2002), protoporphyrinogen oxidase (Frank et al., 1999), low-density lipoprotein receptor (Langenhoven et al., 1996), and mitochondrial acetoacetyl-CoA thiorase (Fukao et al., 2003). Thus, it is likely that the 1A>T variation is a low-activity variation.

The functional effect of the known nonsynonymous SNP 617G>A ($Arg^{206}His$) was also investigated. The Arg^{206} residue is located in the α-helix within the catalytic domain and conserved among human carboxylesterases (Bencharit et al., 2002). However, no significant differences were found between the intrinsic enzyme activities of the wild-type and $Arg^{206}His$ variant for irinotecan hydrolysis.

In this study, 20 haplotypes of the *CES2* gene were identified. The most frequent haplotype was **1a* (frequency, 0.681), followed by **1b* (0.233), **1c* (0.027), and **1d* (0.017). Haplotype **1b* includes the polymorphisms IVS10-108G>A and 1749A>G, and haplotype **1c* harbors −363C>G, IVS10-108G>A, and IVS10-87G>A. The haplotype corresponding to **1b* in this study was found in Caucasians with a frequency of 0.086 (haplotypes 3 and 7 in Wu et al., 2004). Our **1c* corresponds to haplotypes 2 and 12 in Wu et al. (2004) and genotypes **1* and **6* in Charasson et al. (2004). Among the SNPs consisting of haplotype **1b* and **1c*, the three SNPs, −363C>G in the 5′-UTR, IVS10-108(IVS10 + 406)G>A in intron 10, and

TABLE 2

Kinetic parameters of CPT-11 hydrolysis by wild-type and $Arg^{206}His$ variant CES2 expressed in COS-1 cells

Results are expressed as the mean ± S.D. from four independent transfection experiments.

CES2	Apparent K_m	V_{max}	V_{max}/K_m	Normalized V_{max}[a]	Normalized V_{max}/K_m[a]
	μM	*pmol/min/mg protein*	*nl/min/mg protein*	*pmol/min/mg protein*	*nl/min/mg protein*
Wild-type	0.46 ± 0.01	3.45 ± 0.29	7.43 ± 0.54	3.46 ± 0.23	7.45 ± 0.50
$Arg^{206}His$	0.51 ± 0.02‡	2.81 ± 0.22†	5.53 ± 0.52‡	3.44 ± 0.16	6.77 ± 0.46

[a] V_{max} values were normalized by the relative protein expression level of the $Arg^{206}His$ variant (0.82 ± 0.07).
† Significantly different from that of the wild-type at $P < 0.05$.
‡ $P < 0.01$.

1749(*69)A>G in the 3′-UTR of exon 12, were previously reported, and their frequencies varied among several ethnic groups (Marsh et al., 2004; Wu et al., 2004). The frequency (0.269) of the **1b/*1c*-tagging SNP in Japanese, IVS10-108G>A, was comparable to that in African-Americans (0.263), but much higher than that in Asian-Americans (0.06) and European-Americans (0.063) (Wu et al., 2004). However, the **1b*-tagging SNP 1749A>G (0.239 in this study) was detected only in Asian-Americans with a low frequency (0.03) (Wu et al., 2004). The frequency of the **1c*-tagging SNP, −363C>G, also showed marked ethnic differences between Japanese (0.031) and Europeans (0.12) or Africans (0.33) (our data; Marsh et al., 2004). These findings indicate the existence of large ethnic difference in haplotype structures among African, European, and Japanese populations.

In this study, the relationship between the *CES2* genotypes and the (SN-38 + SN-38G)/CPT-11 AUC ratios of irinotecan-administered patients was analyzed. First, the relationship between the genotypes and the AUC ratios among the **1/*1* diplotypes in the patient group with or without coadministered drugs was assessed, and no significant differences in the AUC ratios were observed among the **1/*1* diplotypes in each group (Fig. 2). Wu et al. (2004) reported that the haplotype harboring SNP −363C>G that was homozygous appeared to have lower mRNA levels than the other haplotype groups. In this study, the haplotype having the SNP −363C>G was assigned haplotype **1c*. However, no functional differences were found between haplotype **1c* and the other **1* group haplotypes. Marsh et al. (2004) reported that IVS10-88C>T was associated with reduced RNA expression in colon tumor tissues. However, this SNP was not found in the present study with Japanese subjects.

The major **1* group haplotypes, **1a*, **1b*, and **1c*, account for 94% of Japanese *CES2* haplotypes. The current study revealed no association between the major *CES2* genotypes and changes in the AUC ratio, indicating that the variability in AUC ratio could not be interpreted by these haplotypes alone.

In irinotecan-administered patients, three nonsynonymous SNPs, 100C>T ($Arg^{34}Trp$, **2*), 1A>T ($Met^{1}Leu$, **5*), and 617G>A ($Arg^{206}His$, **6*), and a SNP at the splice acceptor site of intron 8 (IVS8-2A>G, **4*) were found as single heterozygotes. The patients heterozygous for $Arg^{34}Trp$ or $Met^{1}Leu$ showed substantially reduced AUC ratios. These results were consistent with in vitro functional analysis for the nonsynonymous SNPs by Kubo et al. (2005).

In the case of haplotype **6* harboring the nonsynonymous SNP, 617G>A ($Arg^{206}His$), the AUC ratio of the patient who received cisplatin was lower than the median value but within the range for the **1/*1* group treated with platinum-containing drugs. The protein expression level of the 206His variant was 82 ± 4%, and the $Arg^{206}His$ substitution itself showed no functional differences in intrinsic enzyme activity by in vitro functional analysis. Thus, the impact of the 617G>A ($Arg^{206}His$) SNP on irinotecan pharmacokinetics might be small.

On the other hand, the AUC ratio of the patient carrying the haplotype **4* was not different from the median value of the **1/*1* group treated with platinum-containing drugs. It is possible that other genetic factors might have increased the AUC ratio in this patient.

The patients with **4*, **5*, or **6* were found as single heterozygotes. Thus, further studies are needed to elucidate in vivo importance of the three haplotypes.

In conclusion, we have identified a panel of haplotypes of the *CES2* gene in a Japanese population using 21 genetic polymorphisms detected in this study and found that some rare haplotypes with nonsynonymous SNPs show a decreasing tendency toward enzymatic levels or activity. In vitro functional analysis for nonsynonymous SNPs showed that the 1A>T ($Met^{1}Leu$) SNP was a defective allele. These findings will be useful for further pharmacogenetic studies on efficacy and adverse reactions to CES2-activated prodrugs.

Acknowledgments. We thank Chie Sudo for secretarial assistance. We also thank Yakult Honsha Co. Ltd. for kindly providing CPT-11, SN-38, and SN-38G.

References

Bencharit S, Morton CL, Howard-Williams EL, Danks MK, Potter PM, and Pedinbo MR (2002) Structural insight into CPT-11 activation by mammalian carboxylesterases. *Nat Struct Biol* **9**:337–342.

Cecchin E, Corona G, Masier S, Binson P, Cattarossi G, Frustaci S, Buonadonna A, Colussi A, and Toffoli G (2005) Carboxylesterase isoform 2 mRNA expression in peripheral blood mononuclear cells is a predictive marker of the irinotecan to SN38 activation step in colorectal cancer patients. *Clin Cancer Res* **11**:6901–6907.

Charasson V, Bellott R, Meynard D, Longy M, Gorry P, and Robert J (2004) Pharmacogenetics of human carboxylesterase 2, an enzyme involved in the activation of irinotecan into SN-38. *Clin Pharmacol Ther* **76**:528–535.

Dodds HM, Haaz MC, Riou JF, Robert J, and Rivory LP (1998) Identification of a new metabolite of CPT-11 (irinotecan): pharmacological properties and activation to SN-38. *J Pharmacol Exp Ther* **286**:578–583.

Frank J, McGrath JA, Poh-Fitzpatrick MB, Hawk JL, and Christiano AM (1999) Mutations in the translation initiation codon of the protoporphyrinogen oxidase gene underlie variegate porphyria. *Clin Exp Dermatol* **24**:296–301.

Fukao T, Matsuo N, Zhang GX, Urasawa R, Kubo T, Kohno Y, and Kondo N (2003) Single base substitutions at the initiator codon in the mitochondrial acetoacetyl-CoA thiolase (ACAT1/T2) gene result in production of varying amounts of wild-type T2 polypeptide. *Hum Mutat* **21**:587–592.

Hanioka N, Jinno H, Nishimura T, Ando M, Ozawa S, and Sawada J (2001) High-performance liquid chromatographic assay for glucuronidation activity of 7-ethyl-10-hydroxycamptothecin (SN-38), the active metabolite of irinotecan (CPT-11), in human liver microsomes. *Biomed Chromatogr* **15**:328–333.

Humerickhouse R, Lohrbach K, Li L, Bosron WF, and Dolan ME (2000) Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2. *Cancer Res* **60**:1189–1192.

Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, Coffman BL, and Ratain MJ (1998) Genetic predisposition to the metabolism of irinotecan (CPT-11): role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest* **101**:847–854.

Kim SR, Nakamura T, Saito Y, Sai K, Nakajima T, Saito H, Shirao K, Minami H, Ohtsu A, Yoshida T, et al. (2003) Twelve novel single nucleotide polymorphisms in the *CES2* gene encoding human carboxylesterase 2 (hCE-2). *Drug Metab Pharmacokinet* **18**:327–332.

Kitamura Y, Moriguchi M, Kaneko H, Morisaki H, Morisaki T, Toyama K, and 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.

Kozak M (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* **115**:887–903.

Kubo T, Kim SR, Sai K, Saito Y, Nakajima T, Matsumoto K, Saito H, Shirao K, Yamamoto N, Minami H, et al. (2005) Functional characterization of three naturally occurring single nucleotide polymorphisms in the *CES2* gene encoding carboxylesterase 2 (hCE-2). *Drug Metab Dispos* **33**:1482–1487.

Langenhoven E, Warnich L, Thiart R, Rubinsztein DC, van der Westhuyzen DR, Marais AD, and Kotze MJ (1996) Two novel point mutations causing receptor-negative familial hypercholesterolemia in a South African Indian homozygote. *Atherosclerosis* **125**:111–119.

Marsh S, Xiao M, Yu J, Ahluwalia R, Minton M, Freimuth RR, Kwok PY, and McLeod HL (2004) Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* **84**:661–668.

Mathijssen RH, van Alphen RJ, Verweij J, Loos WJ, Nooter K, Stoter G, and Sparreboom A (2001) Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). *Clin Cancer Res* **7**:2182–2194.

Minami H, Sai K, Saeki M, Saito Y, Ozawa S, Suzuki K, Kaniwa N, Sawada J, Hamaguchi T, Yamamoto N, et al. (2007) Irinotecan pharmacokinetics/pharmacodynamics and *UGT1A* genetic polymorphisms in Japanese: roles of *UGT1A1*6* and **28*. *Pharmacogenet Genomics* **17**:497–504.

Nebert DW (2000) Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics* **10**:279–290.

Pindel EV, Kedishvili NY, Abraham TL, Brzezinski MR, Zhang J, Dean RA, and Bosron WF (1997) Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. *J Biol Chem* **272**:14769–14775.

Quinteiro C, Castro-Feijoo L, Loidi L, Barreiro J, de la Fuente M, Dominguez F, and Pombo M (2002) Novel mutation involving the translation initiation codon of the growth hormone receptor gene (GHR) in a patient with Laron syndrome. *J Pediatr Endocrinol Metab* **15**:1041–1045.

Sai K, Kaniwa N, Ozawa S, and Sawada J (2002) An analytical method for irinotecan (CPT-11) and its metabolites using a high-performance liquid chromatography: parallel detection with fluorescence and mass spectrometry. *Biomed Chromatogr* **16**:209–218.

Sai K, Saeki M, Saito Y, Ozawa S, Katori N, Jinno H, Hasegawa R, Kaniwa N, Sawada J, Komamura K, et al. (2004) UGT1A1 haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer. *Clin Pharmacol Ther* **75**:501–515.

Santos A, Zanetta S, Cresteil T, Deroussent A, Pein F, Raymond E, Vernillet L, Risse ML, Boige V, Gouyette A, et al. (2000) Metabolism of irinotecan (CPT-11) by CYP3A4 and CYP3A5 in humans. *Clin Cancer Res* **6**:2012–2020.

Satoh T and Hosokawa M (1998) The mammalian carboxylesterases: from molecules to functions. *Annu Rev Pharmacol Toxicol* **38**:257–288.

Satoh T, Taylor P, Bosron WF, Sanghani SP, Hosokawa M, and LaDu BN (2002) Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* 30:488–493.
Schwer H, Langmann T, Daig R, Becker A, Aslanidis C, and Schmit G (1997) Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. *Biochem Biophys Res Commun* 233:117–120.
Shibata F, Takagi Y, Kitajima M, Kuroda T, and Omura T (1993) Molecular cloning and characterization of a human carboxylesterase gene. *Genomics* 17:76–82.
Takai S, Matsuda A, Usami Y, Adachi T, Sugiyama T, Katagiri Y, Tatematsu M, and Hirano K (1997) Hydrolytic profile for ester- or amide-linkage by carboxylesterases pI 5.3 and 4.5 from human liver. *Biol Pharm Bull* 20:869–873.
Wu MH, Chen P, Remo BF, Cook EH Jr, Das S, and Dolan ME (2003) Characterization of multiple promoters in the human carboxylesterase 2 gene. *Pharmacogenetics* 13:425–435.
Wu MH, Chen P, Wu X, Liu W, Strom S, Das S, Cook EH Jr, Rosner GL, and Dolan ME (2004) Determination and analysis of single nucleotide polymorphisms and haplotype structure of the human carboxylesterase 2 gene. *Pharmacogenetics* 14:595–605.
Xie M, Yang D, Liu L, Xue B, and Yan B (2002) Human and rodent carboxylesterases: immunorelated-ness, overlapping substrate specificity, differential sensitivity to serine enzyme inhibitors, and tumor-related expression. *Drug Metab Dispos* 30:541–547.
Xu G, Zhang W, Ma MK, and McLeod HL (2002) Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with activation of irinotecan. *Clin Cancer Res* 8:2605–2611.

Address correspondence to: Dr. Su-Ryang Kim, Project Team for Pharmacogenetics, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. E-mail: kim@nihs.go.jp

J Hum Genet (2007) 52:804–819
DOI 10.1007/s10038-007-0186-6

ORIGINAL ARTICLE

Genetic variations and haplotype structures of the *DPYD* gene encoding dihydropyrimidine dehydrogenase in Japanese and their ethnic differences

Keiko Maekawa · Mayumi Saeki · Yoshiro Saito · Shogo Ozawa · Kouichi Kurose · Nahoko Kaniwa · Manabu Kawamoto · Naoyuki Kamatani · Ken Kato · Tetsuya Hamaguchi · Yasuhide Yamada · Kuniaki Shirao · Yasuhiro Shimada · Manabu Muto · Toshihiko Doi · Atsushi Ohtsu · Teruhiko Yoshida · Yasuhiro Matsumura · Nagahiro Saijo · Jun-ichi Sawada

Received: 30 May 2007 / Accepted: 26 July 2007 / Published online: 9 September 2007

Abstract Dihydropyrimidine dehydrogenase (DPD) is an inactivating and rate-limiting enzyme for 5-fluorouracil (5-FU), and its deficiency is associated with a risk for developing a severe or fatal toxicity to 5-FU. In this study, to search for genetic variations of *DPYD* encoding DPD in Japanese, the putative promoter region, all exons, and flanking introns of *DPYD* were sequenced from 341 subjects including cancer patients treated with 5-FU. Fifty-five genetic variations, including 38 novel ones, were found and consisted of 4 in the 5′-flanking region, 21 (5 synonymous and 16 nonsynonymous) in the coding exons, and 30 in the introns. Nine novel nonsynonymous SNPs, 29C>A (Ala10Glu), 325T>A (Tyr109Asn), 451A>G (Asn151Asp), 733A>T (Ile245Phe), 793G>A (Glu265Lys), 1543G>A (Val515Ile), 1572T>G (Phe524Leu), 1666A>C (Ser556-Arg), and 2678A>G (Asn893Ser), were found at allele frequencies between 0.15 and 0.88%. Two known nonsynonymous variations reported only in Japanese, 1003G>T (**11*, Val335Leu) and 2303C>A (Thr768Lys), were found at allele frequencies of 0.15 and 2.8%, respectively. SNP and haplotype distributions in Japanese were quite different from those reported previously in Caucasians. This study provides fundamental information for pharmacogenetic studies for evaluating the efficacy and toxicity of 5-FU in Japanese and probably East Asians.

Keywords *DPYD* · SNP · Haplotype · Japanese · 5-fluorouracil

K. Maekawa (✉) · Y. Saito · J. Sawada
Division of Biochemistry and Immunochemistry,
National Institute of Health Sciences,
1-18-1 Kamiyoga, Setagaya-ku,
Tokyo 158-8501, Japan
e-mail: maekawa@nihs.go.jp

K. Maekawa · M. Saeki · Y. Saito · S. Ozawa · K. Kurose · N. Kaniwa · J. Sawada
Project Team for Pharmacogenetics,
National Institute of Health Sciences, Tokyo, Japan

S. Ozawa
Division of Pharmacology,
National Institute of Health Sciences, Tokyo, Japan

K. Kurose · N. Kaniwa
Division of Medicinal Safety Science,
National Institute of Health Sciences, Tokyo, Japan

M. Kawamoto · N. Kamatani
Division of Genomic Medicine,
Department of Advanced Biomedical Engineering and Science,
Tokyo Women's Medical University, Tokyo, Japan

K. Kato · T. Hamaguchi · Y. Yamada · K. Shirao · Y. Shimada
Gastrointestinal Oncology Division, National Cancer Center Hospital, National Cancer Center, Tokyo, Japan

M. Muto
Gastrointestinal Oncology Division,
National Cancer Center Hospital East, Kashiwa, Japan

T. Doi · A. Ohtsu
Division of GI Oncology/Digestive Endoscopy,
National Cancer Center Hospital East, Kashiwa, Japan

T. Yoshida
Genetics Division, National Cancer Center Research Institute,
National Cancer Center, Tokyo, Japan

Y. Matsumura
Research Center of Innovative Oncology,
National Cancer Center Hospital East, Kashiwa, Japan

N. Saijo
Deputy Director, National Cancer Center Hospital East, Kashiwa, Japan

Introduction

Dihydropyrimidine dehydrogenase (DPD) is an inactivating and rate-limiting enzyme for 5-fluorouracil (5-FU), which is used in various therapeutic regimens for gastrointestinal, breast and head/neck cancers (Grem 1996). While the antitumor effect of 5-FU is exerted via anabolic pathways responsible for its intracellular conversion into anti-proliferative nucleotides, DPD affects 5-FU availability by rapidly degrading it to 5, 6-dihydrofluorouracil (DHFU) (Heggie et al. 1987). The importance of DPD in 5-FU metabolism was also highlighted by a lethal drug interaction between 5-FU and the antiviral agent sorivudine. Due to inhibition of DPD by a sorivudine metabolite, severe systemic exposure to 5-FU caused several acute deaths in Japan (Nishiyama et al. 2000).

5-FU catabolism occurs in various tissues, including tumors, but is highest in the liver (Naguib et al. 1985; Lu et al. 1993). Wide variations in DPD activity (8- to 21-fold) were shown in Caucasians, and 3–5% of Caucasians had reduced DPD activity (Etienne et al. 1994; Lu et al. 1998). This variability, which is partially attributed to genetic defects of the DPD gene (*DPYD*), leads to differential responses of cancer patients, resistance to or increased toxicity of 5-FU (van Kuilenburg 2004). Complete DPD deficiency is also associated with the inherited metabolic disorder, thymine-uraciluria, which is characterized by neurological problems in pediatric patients (Bakkeren et al. 1984).

To date, at least 30 variant *DPYD* alleles have been published, with or without deleterious impact upon DPD activity (Gross et al. 2003; Ogura et al. 2005; Seck et al. 2005; van Kuilenburg 2004; Zhu et al. 2004). Of these variations, a splice site polymorphism, IVS14 + 1G>A, which causes skipping of exon 14, is occasionally detected in North Europeans with allele frequencies of 0.01–0.02 (van Kuilenburg 2004). Detection of IVS14 + 1G>A in patients suffering from 5-FU-associated grade 3 or 4 toxicity revealed that 24–28% of them were heterozygous or homozygous for this single nucleotide polymorphism (SNP) (van Kuilenburg 2004). However, this SNP has not been reported in Japanese and African-Americans. Recently, Ogura et al. (2005) have shown that a Japanese population exhibits a large degree of interindividual variations in DPD activity of peripheral blood mononuclear cells. They also identified a novel variation, 1097G>C (Gly366Ala), in a healthy volunteer with the lowest DPD activity and demonstrated that the 366Ala variant has reduced activity towards 5-FU in vitro. At present, however, information on variant alleles with clinical relevance in Japanese is limited and cannot fully explain polymorphic DPD activity.

In this study, we searched for genetic variations in *DPYD* by sequencing 5′ regulatory regions, all exons and surrounding introns from 341 Japanese subjects. Fifty-five variations including nine novel nonsynonymous ones were identified. Then, linkage disequilibrium (LD) and haplotype analyses were performed to clarify the *DPYD* haplotype structures in Japanese.

Materials and methods

Human DNA samples

Three hundred and forty-one Japanese subjects in this study included 263 cancer patients and 78 healthy volunteers. All 263 patients were administered 5-FU or tegafur for treatment of various cancers (mainly stomach and colon) at the National Cancer Center, and blood samples were collected prior to the fluoropyrimidine chemotherapy. The healthy volunteers were recruited at the Tokyo Women's Medical University. DNA was extracted from the blood of cancer patients and Epstein-Barr virus-transformed lymphoblastoid cells derived from healthy volunteers. Written informed consent was obtained from all participating subjects. The ethical review boards of the National Cancer Center, the Tokyo Women's Medical University and the National Institute of Health Sciences approved this study.

PCR conditions for DNA sequencing

To amplify 22 exons (exons 2–23) of *DPYD*, multiplex PCRs were performed by using four sets of mixed primers (mix 1 to mix 4 of "first PCR" in Table 1). Namely, five exonic fragments were simultaneously amplified from 50 ng of genomic DNA using 0.625 units of Ex-Taq (Takara Bio. Inc., Shiga, Japan) with 0.20 μM primers. Because of the high GC content in exon 1 of *DPYD*, this region was separately amplified from 50 ng of genomic DNA with 2.5 units of LA-Taq and 0.2 μM primers (listed in Table 1) in GC buffer I (Takara Bio. Inc.). The first PCR conditions were 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 58°C for 1 min, and 72°C for 2 min; and then a final extension for 7 min at 72°C. Next, each exon was amplified separately from the first PCR products by nested PCR (2nd PCR) using the primer sets (0.2 μM) listed in "second PCR" of Table 1. The second PCR conditions were the same as those of the first PCR, and LA-Taq (2.5 units) for exon 1 and Ex-Taq (0.625 units) for exons 2–23 were used. All PCR primers were designed in the flanking intronic sites to analyze the exon-intron splice junctions. The PCR products were treated with a PCR Product Pre-Sequencing Kit (USB Co., Cleveland, OH) and sequenced directly on both strands using an ABI BigDye Terminator Cycle Sequencing Kit (Applied Biosystems,

Table 1 Primer sequences for human *DPYD*

Amplified and sequenced region			Forward primer		Reverse primer		PCR product (bp)
			Sequences (5' to 3')	Position[a]	Sequences (5' to 3')	Position[a]	
First PCR	5'-UTR to exon 1		GTTCTGGAAGGTAATCTGATGG	52207178	ACGACATACAGGAGGTGAAG	52205443	1,736
	Mix 1	Exon 2	CTACTTGGGAGACTAAGGTG	52168526	GTATCATTGTGTCATTAGGC	52167832	695
		Exon 3	TCCCTTCATCTTAGTCAATG	52113605	CTGAGGCTTAACATTTATGC	52112876	730
		Exon 4	TCTGAGAGGAGGGACAGTTA	52025660	AATCACAACTTGGAAGTGCT	52025165	496
		Exon 5	AAATGGAGGATAACCTGAGT	52007046	TAATAACCTGCTGGGATTGC	52006234	813
		Exon 6	AGAGGAGAGGCACTTAATGT	51984772	TGCTTCAAGCCAACTGCAAA	51984115	658
	Mix 2	Exon 7	CTCAAATAATAGTGCCATAGG	51977410	CAGTAGACAGACAAATGCCC	51976498	913
		Exon 8	CACATCGTGCTTTGAACATA	51964415	CCAACTCCATCCTTTATGAT	51963667	749
		Exons 9 and 10	TGAGGCAAGAATATAACCTG	51880431	TCCGTATGTGTCTTATTACC	51877795	2,637
		Exon 11	AGAAATACCTTATGATGCCG	51859160	GCCTTTTGAATCAAGATTGC	51858562	599
		Exon 16	CTCCCTATGCTTCAGTTCAC	51658925	TGCCGTGCCCCATTTACTAC	51658114	812
	Mix 3	Exon 12	CCGCTCTGAAACATTGACCA	51834944	CTGGGATTATAGGCATTAGG	51834279	666
		Exon 13	GCCCATATCTCTGAGCACTA	51801258	ATCTTTGTTGCTTCCTAGAC	51800450	809
		Exon 14	CCTTCACTGATTTACATCGG	51735640	CCAGCCACATACAGTGAAAA	51734704	937
		Exon 15	AGCCAGTAAAATCCTCTCTA	51667711	TATGGAAAACCTGCTGACTA	51666815	897
		Exon 23	TGGAAAGACCCGAACTCTGC	51364409	AGCGAAGGGGATTTTACTTA	51363336	1,074
	Mix 4	Exons 17 and 18	TTCTAAAGGCTCTGTTGAGG	51591491	TGGCAAAAGAACTGAGAGAC	51589933	1,559
		Exon 19	CGTGGATTCAAGCAGTTTTC	51520500	AGACAGTGGGTTCGTAAGCC	51519586	915
		Exon 20	CTGTGACACCATTACCATTG	51478435	TGCCAGTCATCACCACAGTA	51477733	703
		Exon 21	GAACCTGATACCGAGAAGAC	51383758	AAATGTCCAGGCTTTCCAGA	51382987	772
		Exon 22	GCCATAACAACTCACACGGG	51367740	TTGGCAGAAGGAATCATAGC	51366885	856
Second PCR		5'-UTR to exon 1	TGTGGATGTTTTTGCTCGC	52206503	AGTAAACAGGTCCCGACGC	52205586	918
		Exon 2	GTGAACTGAGATTGTACCACTGC	52168471	CATATCCCTTATCAAAAATGCTT	52167924	548
		Exon 3	GAATGCTACCCAATTAAAGTGG	52113285	TTCAAAACCAAATACAGCCTC	52112899	387
		Exon 4	TGCCAAAGATGAAACACAGA	52025601	ACCCACAGATAATAGAGAACAAGA	52025273	329
		Exon 5	TGATGGTTCCTGATAGTAGTATTG	52006775	TGTCACACTAAAAATGTTGGG	52006348	428
		Exon 6	AAGGAAAGACTGAAAGTTAGCC	51984688	GAGCCTGAAGTTCCTATATGAT	51984201	488
		Exon 7	TTCTACTGTATCTTCACTCCACG	51976953	GCTTCTGCCTGATGTAGC	51976541	413
		Exon 8	GGCTGACTTTTCATTCTTTTT	51964221	CATCTTGCCGAAATCTCTCC	51963831	391
		Exon 9	TGTGATTTACGATGTGTACTTGG	51880335	GCAAGGTTGGGTGTGAGAG	51879895	441
		Exon 10	AAAATGGGAATAAAACTGTCTT	51878507	TCAGGATATGGAAGACTTAGCAC	51877859	649
		Exon 11	ACTGGTAACTGAAACTCAG	51859069	CAATTCCCTGAAAGCTAG	51858628	442
		Exon 12	TCAGTGCCCTTCAAATGTGT	51834881	ACCAAATAGAAATGCTCTTATAGA	51834414	468
		Exon 13	TCGGATGCTGTGTTGAAGTG	51800982	TGTGTAATGATAGGTCGTGTC	51800543	440

Table 1 Primer sequences for human *DPYD*

Amplified and sequenced region		Forward primer		Reverse primer		PCR product (bp)
		Sequences (5' to 3')	Position[a]	Sequences (5' to 3')	Position[a]	
	Exon 14	TGCAAATATGTGAGGAGGGACC	51735287	CAGCAAAGCAACTGGCAGATTC	51734877	411
	Exon 15	GCTATCTTACCCTGCTATTTTC	51667571	TAGGTAGTGTGTGAAATCCAAGG	51667107	465
	Exon 16	CCCCTTATGAGCACTGAGTAAAT	51658821	TAGTAACTATCCATACGGGGG	51658440	382
	Exon 17	AGTCTAGGTGTAATACTGAGGAGG	51591407	ATCAAGTGCTCAACTGGAAACT	51590986	422
	Exon 18	GTGAAGAACTTTGAGGAGAAGAC	51590461	CATCCTGTGCTGTCACTTGA	51590026	436
	Exon 19	ATTTGTCCAGTGACGCTGTC	51520048	TCAGGTCTCTTCATAACTTGTCAG	51519629	420
	Exon 20	GAGAAGTGAATTTGTTTGGAG	51478265	TTTGTTAGTGAGAATGTGAGATGG	51477926	340
	Exon 21	AGTGGTCCAAAACAATGAGTG	51383737	TGCTTGCCAGTGTTCTAAAA	51383221	517
	Exon 22	GGGTGTCATTTATTCTTTCTGTC	51367723	GGCTGATGAAATGGTATAAAAA	51367033	691
	Exon 23	GTTGTCTCATAGTGTGGCTCCTC	51364206	TTTTTCACATAAGACAACTGGCA	51363641	566
Sequencing	5'-UTR to exon 1	TGTGGATGTTTTTGCTCGC	52206503			
	5'-UTR to exon 1	CGGACTGCTTTTACCTTTGC	52206258	CCAGAGAGCCAAGTGACAGC	52205933	
	5'-UTR to exon 1	CCCTAGTCTGCCTGTTTTCG	52205987	AGTAAACAGGTCCCGACGC	52205586	
	Exon 2	GTGACAAAGTGAGAGAGACCGT	52168436	GCCTTACAATGTGTGGAGTGAG	52168152	
	Exon 3	GAATGCTACCCAATTAAAGTGG	52113285	TTCAAAACCAAATACAGCCTC	52112899	
	Exon 4	TGCCAAAGATGAAACACAGA	52025601	ACCCACAGATAATAGAGAACAAGA	52025273	
	Exon 5	TGATGGTTCCTGATAGTAGTATTG	52006775	TGTCACACTAAAAATGTTGGG	52006348	
	Exon 6	AAAATATGTTTGAGGATGTAAGC	51984560	GAGCCTGAAGTTCCTATATGAT	51984201	
	Exon 7	TTCTACTGTATCTTCACTCCACG	51976953	GCTTCTGCCTGATGTAGC	51976541	
	Exon 8	GGCTGACTTTTCATTCTTTTT	51964221	CATCTTGCCGAAATCTCTCC	51963831	
	Exon 9	TGTGATTTACGATGTGTACTTGG	51880335	GCAAGGTTGGGTGTGAGAG	51879895	
	Exon 10	AAAATGGGAATAAAACTGTCTT	51878507	TTCATCTCCTAAAATCTGTTGG	51878109	
	Exon 11	ACTGGTAACTGAAACTCAG	51859069	CAATTCCCTGAAAGCTAG	51858628	
	Exon 12	TCAGTGCCCTTCAAATGTGT	51834881	GAGTATCAAAAATAAATGAAGCAC	51834439	
	Exon 13	TCGGATGCTGTGTTGAAGTG	51800982	TGTGTAATGATAGGTCGTGTC	51800543	
	Exon 14	TGCAAATATGTGAGGAGGGACC	51735287	CAGCAAAGCAACTGGCAGATTC	51734877	
	Exon 15	GCTATCTTACCCTGCTATTTTC	51667571	TAGGTAGTGTGTGAAATCCAAGG	51667107	
	Exon 16	CCCCTTATGAGCACTGAGTAAAT	51658821	TAGTAACTATCCATACGGGGG	51658440	
	Exon 17	AGTCTAGGTGTAATACTGAGGAGG	51591407	ATCAAGTGCTCAACTGGAAACT	51590986	
	Exon 18	GTGAAGAACTTTGAGGAGAAGAC	51580461	CATCCTGTGCTGTCACTTGA	51590026	
	Exon 19	ATTTGTCCAGTGACGCTGTC	51520048	CGAATCTATTTTTTTTTTGTCAC	51519715	
	Exon 20	GAGAAGTGAATTTGTTTGGAG	51478265	TTTGTTAGTGAGAATGTGAGATGG	51477926	
	Exon 21	TATCTTCCCATTTTTCTCTTCTC	51383644	TGCCAGTGTTCTAAAAAGTATAAA	51383225	
	Exon 22	GTATAAAAACAGGAAAATGCTGA	51367510	ATAAGGGTGACAGGACAGAAAG	51367125	
	Exon 23	GTTGTCTCATAGTGTGGCTCCTC	51364206	TATTTGTTTTAATTTGGAAAGAG	51363821	

[a] Nucleotide position of the 5' end of each primer on NT_032977.7

Foster City, CA) with the primers listed in "sequencing" of Table 1. Excess dye was removed with a DyeEx96 kit (Qiagen, Hilden, Germany). The eluates were analyzed on an ABI Prism 3700 DNA Analyzer (Applied Biosystems). All novel SNPs were confirmed by sequencing of PCR products generated from new genomic DNA amplifications. The genomic and cDNA sequences of *DPYD* obtained from GenBank (NT_032977.7 and NM_000110.2, respectively) were used as reference sequences. SNP positions were numbered based on the cDNA sequence, and adenine of the translational initiation site in exon 1 was numbered +1. For intronic polymorphisms, the position was numbered from the nearest exon.

Linkage disequilibrium (LD) and haplotype analyses

Hardy-Weinberg equilibrium and LD analyses were performed by SNPAlyze software (Dynacom Co., Yokohama, Japan), and pairwise LD parameters between variations were obtained as the lD'l and rho square (r^2) values. Some haplotypes were unambiguously identified from subjects with homozygous variations at all sites or a heterozygous variation at only one site. Diplotype configurations were inferred by LDSUPPORT software, which determines the posterior probability distribution of the diplotype for each subject based on the estimated haplotype frequencies (Kitamura et al. 2002). Although the nomenclature for nonsynonymous *DPYD* alleles (*DPYD*1* to *DPYD*13*) have been already publicized (McLeod et al. 1998; Collie-Duguid et al. 2000; Johnson et al. 2002), several reported alleles remain unassigned. To avoid confusion with the previous *DPYD* allele nomenclature, our block haplotypes in this study were tentatively defined by using '#' instead of '*'. A group of haplotypes without any amino acid change is designated as *#1*, and the haplotype groups bearing already defined alleles, *DPYD*5* (Ile543Val), *DPYD*6* (Val732Ile), *DPYD*9* (Cys29Arg) and *DPYD*11* (Val335Leu), were numbered by using the corresponding Arabic numerals, *#5*, *#6*, *#9*, and *#11*, respectively. Other haplotypes with known nonsynonymous SNPs such as 496A>G (Met166Val) or with the novel nonsynonymous SNP were represented by '#' plus amino acid positions followed by variant residues (for example, *#166V*). Subtypes within each haplotype group were consecutively named with small alphabetical letters depending on their frequencies. Haplotypes ambiguously inferred in only one patient were indicated in the Fig. 3 legend. Combinations of block haplotypes were analyzed by Haploview software (http://www.broad.mit.edu/mpg/haploview/index.php) (Barrett et al. 2005), and the long-range (whole gene) haplotypes spanning all blocks were inferred by Hapblock software (www.cmb.usc.edu/msms/HapBlock/) (Zhang et al. 2005).

Typing data on *DPYD* from unrelated 44 Japanese and 30 Caucasian trios were also obtained from the HapMap project (HapMap release 19: http://www.hapmap.org/). The LD profiles and haplotypes of the HapMap data were obtained by Marker beta in Gmap Net (http://www.gmap.net/marker) using its four (1254711, 1254712, 1254713, and 1254714) and six (1166276, 1166277, 1166278, 1166279, 1166280, and 1166281) datasets covering *DPYD* genomic regions for Japanese and Caucasians, respectively.

Drawing of protein structures

The coordinate data (1gth) of the crystal structure of pig DPD (Dobritzsch et al. 2002) was obtained from the Protein Data Bank. Protein Explorer (http://proteinexplorer.org) (Martz 2002) was used to display the structural features of pig DPD and depict three-dimensional views.

Results

DPYD variations found in a Japanese population

We identified 55 variations, including 38 novel ones by sequencing the promoter regions (up to 613 bp upstream from the translational initiation site), all 23 exons and their flanking regions of *DPYD* from 341 Japanese subjects (Table 2). The distribution of the variations consisted of 4 in the 5' flanking region, 21 (5 synonymous and 16 nonsynonymous ones) in the coding exons (Fig. 1) and 30 in the introns. Since we did not find any significant differences in allele frequencies between healthy volunteers and cancer patients ($P > 0.05$ by χ^2 test or Fisher's exact test) except for one variation, IVS14 + 19C>A, ($P = 0.027$ by Fisher's exact test); the data for all subjects were analyzed as one group. All detected variations except for 451A>G (Asn151Asp) and IVS13 + 40G>A were in Hardy-Weinberg equilibrium ($P \geq 0.24$).

Thirteen novel variations in the coding region (enclosed by a square in Fig. 1) contain four synonymous SNPs, 474T>C (Phe158Phe), 639C>T (Asp213Asp), 1752A>G (Thr584Thr), and 2424T>C (Ser808Ser) and nine nonsynonymous SNPs, 29C>A (Ala10Glu), 325T>A (Tyr109Asn), 451A>G (Asn151Asp), 733A>T (Ile245Phe), 793G>A (Glu265Lys), 1543G>A (Val515Ile), 1572T>G (Phe524Leu), 1666A>C (Ser556Arg), and 2678A>G (Asn893Ser). 451A>G (Asn151Asp), 325T>A (Tyr109Asn), and 2678A>G (Asn893Ser) were found at frequencies of 0.009, 0.003 and 0.003, respectively. The others were detected as single heterozygotes (allele frequencies = 0.0015).

Table 2 Summary of *DPYD* SNPs detected in a Japanese population

SNP ID		Location	Position		Nucleotide change and flanking sequences (5′ to 3′)	Amino acid change	Reported alleles	Allele frequency (341 subjects)
This study	dbSNP (NCBI)		NT_032977.7	From the translational initiation site or from the end of nearest exon				
MPJ6_DPD001[a]		5′-Flank	52206480	−609	TTGCTCGCCTCCC/TTCCCCTCCCCGC			0.021
MPJ6_DPD002[a]		5′-Flank	52206348	−477	TTGAGGAGTTCCT/GGAAAATGCAGTT			0.026
MPJ6_DPD003[a]		5′-Flank	52206137	−266	CTCCCTCCCTCCC/ATTCTGCTTGCAG			0.045
MPJ6_DPD004[a]		5′-Flank	52206114	−243	AGGCTGGGGCGCG/AGAGCGGGCTGAA			0.0059
MPJ6_DPD005[a]		Exon 1	52205843	29	GTAAGGACTCGGC/AGGACATCGAGGT	Ala10Glu		0.0015
MPJ6_DPD006[g]	rs1801265	Exon 2	52168278	85	CATGCAACTCTGT/CGTTCCACTTCGG	Cys29Arg	*9	0.029
MPJ6_DPD007[a]		Intron 2	52168055	IVS2 + 158	TTTGAAAGTGTAT/CTTTTAATTACAC			0.0015
MPJ6_DPD008[a]		Intron 3	52113040	IVS3 + 23	GTCACCATAGCAA/GCAGTCACAGATG			0.0029
MPJ6_DPD009[a]		Exon 5	52006617	325	ATTTTGCAGAACT/AATTATGGAGCTG	Tyr109Asn		0.0029
MPJ6_DPD010[a]		Exon 5	52006491	451	GAGGGACCCATTA/GATATTGGTGGAT	Asn151Asp		0.0088
MPJ6_DPD011[a]		Exon 5	52006468	474	ATTGCAGCAATTT/CGCTACTGAGGTA	Phe158Phe		0.0044
MPJ6_DPD012[a]		Intron 5	51984611	IVS5−115	CATATTAATACTG/AAAAATGTACTGC			0.021
MPJ6_DPD013[g]	rs2297595	Exon 6	51984484	496	GTATTCAAAGCAA/GTGAGTATCCCAC	Met166Val		0.022
MPJ6_DPD014[a]		Exon 6	51984341	639	GGGGTACTCTGAC/TATCACTATATTT	Asp213Asp		0.0088
MPJ6_DPD015[a]		Exon 7	51976695	733	GTGAATTTTGAGA/TTTGAGCTAATGA	Ile245Phe		0.0015
MPJ6_DPD016[a]		Intron 7	51976602	IVS7 + 64	CTCTACACTAAAG/TATTAACAGCAAA			0.0015
MPJ6_DPD017[a]		Exon 8	51964101	793	CTTTCAGTGAATG/AAAATGACTCTTA	Glu265Lys		0.0015
MPJ6_DPD018[a]		Intron 8	51963953	IVS8 + 91	TTCAGACATTTTC/TGTGATGAAAGTT			0.0088
MPJ6_DPD019[a]		Intron 9	51878456	IVS9−120	TTTGATAGTGACA/TCTTCATCCTGGA			0.0029
MPJ6_DPD020[b]		Exon 10	51878292	1003	ATACGGGGAGTCG/TTGATTGTACTTG	Val335Leu	*11	0.0015
MPJ6_DPD021[a]		Intron 10	51878143	IVS10 + 24	CCATCAGAAAATA/GTGGAGTTGTACT			0.0015
MPJ6_DPD022[c]		Intron 10	51858934	IVS10-15	TTTCTTCTCTGTT/CCTGTTTTGTTTT			0.018
MPJ6_DPD023[a]		Intron 12	51800901	IVS12-11	AAGTATTGGTTTG/ATATTTTGCAGTC			0.038
MPJ6_DPD024[a]		Intron 12	51800899	IVS12-9	GTATTGGTTTGTA/GTTTTGCAGTCAC			0.0073
MPJ6_DPD025[a]		Exon 13	51800872	1543	TATGGAGCTTCCG/ATTTCTGCCAAGC	Val515Ile		0.0015
MPJ6_DPD026[a]		Exon 13	51800843	1572	ACTACCCCTCTTT/GTACACTCCTATT	Phe524Leu		0.0015
MPJ6_DPD027[g]	rs1801159	Exon 13	51800788	1627	GGATTGAAGTTTA/GTAAATCCTTTTG	Ile543Val	*5	0.283
MPJ6_DPD028[a]		Exon 13	51800749	1666	ACTCCAGCCACCA/CGCACATCAATGA	Ser556Arg		0.0015
MPJ6_DPD029	rs2786783	Intron 13	51800636	IVS13 + 39	AGAAATGTCTATC/TATATATTTTAAT			0.283
MPJ6_DPD030	rs2811178	Intron 13	51800635	IVS13 + 40	GAAATGTCTATCG/ATATATTTTAATT			0.179
MPJ6_DPD031[a]		Intron 13	51735220_51735219	IVS13-47_-48	ATAAAGATTATA−/TAGCTTTTCTTTGT			0.0015
MPJ6_DPD032[a]		Exon 14	51735161	1752	GGACATTGTGACA/GAATGTTTCCCCC	Thr584Thr		0.0015
MPJ6_DPD033[f]		Exon 14	51735139	1774	CCCAGAATCATCC/TGGGGAACCACCT	Arg592Trp		0.0015
MPJ6_DPD034[g]	rs17376848	Exon 14	51735017	1896	AAAGGCTGACTTT/CCCAGACAACGTA	Phe632Phe		0.139
MPJ6_DPD035[a]		Intron 14	51734989	IVS14 + 19	GTGATTTAACATC/ATAAAACAAGAGA			0.0088
MPJ6_DPD036[a]		Intron 14	51734908	IVS14 + 100	TTAATGTGTATAT/GTTTATTAAAGAA			0.0015
MPJ6_DPD037[a]		Intron 14	51667533	IVS14-123	GATTTATTTTTTC/AACAGTTTGAAAA			0.155
MPJ6_DPD038[a]		Intron 14	51667431	IVS14-21	TGAACTTATATTC/ATTTTGTTTTTCT			0.0015
MPJ6_DPD039[d]		Intron 15	51667267	IVS15 + 75	TAAAGAGCTGCCA/GTGAGAAATAATA			0.155
MPJ6_DPD040[a]		Intron 16	51591373	IVS16-127	GGAATTTGAGAAA/GTATATCATGTAG			0.0015

Table 2 continued

SNP ID		Location	Position		Nucleotide change and flanking sequences (5′ to 3′)	Amino acid change	Reported alleles	Allele frequency (341 subjects)
This study	dbSNP (NCBI)		NT_032977.7	From the translational initiation site or from the end of nearest exon				
MPJ6_DPD041[g]	rs7556439	Intron 16	51591340	IVS16-94	CAAGTTGGATTTG/TTCTTGCACGTCT			0.378
MPJ6_DPD042[a]		Intron 17	51591092	IVS17 + 34	GTTGCCCGCTATT/-GTAAATATTGGC			0.0015
MPJ6_DPD043[a]		Intron 17	51591079	IVS17 + 47	GTAAATATTGGCC/TCACATTATGTAG			0.0015
MPJ6_DPD044	rs1801160	Exon 18	51590313	2194	GGTGCCAATGGCG/ATTACAGCCACCA	Val732Ile	*6	0.015
MPJ6_DPD045[g]	rs12137711	Intron 18	51519982	IVS18-39	TATACTCAAGTGG/ATCAGTGTGCTAA			0.032
MPJ6_DPD046[c]		Exon 19	51519940	2303	TTTGTGTAGGGAC/AAGCAATCAGACC	Thr768Lys		0.028
MPJ6_DPD047[a]		Exon 19	51519819	2424	GTTTCTCCATAGT/CGGTGCTTCCGTC	Ser808Ser		0.0029
MPJ6_DPD048[a]		Exon 21	51383526	2678	TCATAGCAGAAAA/GCAAGATTAGACT	Asn893Ser		0.0029
MPJ6_DPD049[a]		Intron 21	51383358	IVS21 + 80	GTTTATTTACTGC/GTTAAATGTATCA			0.0015
MPJ6_DPD050[a]		Intron 21	51383325	IVS21 + 113	GTTTTAGAATTAT/AAATGAAAGTTTT			0.0015
MPJ6_DPD051[g]	rs11165777	Intron 21	51383302	IVS21 + 136	TTAAAAACATCTG/CTCCATGGTGAAA			0.0029
MPJ6_DPD052[a]		Intron 21	51383276	IVS21 + 162	CTGCATTTAAATT/GATAAAATAACCT			0.0029
MPJ6_DPD053[a]		Intron 22	51367150	IVS22 + 129	TTCTGCAACAGTA/GCATCTTTCTGTC			0.0073
MPJ6_DPD054[g]	rs290855	Intron 22	51364164	IVS22-69	GAGAAAAATGTTG/AACGCTAAAATGG			0.0029
MPJ6_DPD055[g]	rs17116357	Intron 22	51364153	IVS22-58	TAACGCTAAAATG/CGGGACATTGTTG			0.0029

[a] Novel variations detected in this study

[b] Kouwaki et al. 1998

[c] Collie-Duguid et al. 2000

[d] Seck et al. 2005

[e] Ogura et al. 2005

[f] Cho et al. 2007

[g] Variations overlapping with the HapMap project

In the 5′ flanking region, all four detected SNPs (-609C>T, -477T>G, -266C>A, -243G>A) were newly found at relatively high allele frequencies (0.006–0.05). However, these SNPs were not located near the proposed *cis*-regulatory promoter elements (Shestopal et al. 2000). The remaining 21 novel variations were found in intronic regions. Of these SNPs, IVS5–115G>A, IVS12–11G>A, and IVS14-123C>A were detected with allele frequencies of 0.021, 0.038, and 0.155, respectively, but others were rare (<0.01). They were not located in the exon-intron splicing junctions or branch sites.

Seventeen variations were already reported. The ID numbers in the dbSNP databases or references for these SNPs are described in Table 2. The well-known nonsynonymous SNPs, 1627A>G (**5*, Ile543Val), 2194G>A (**6*, Val732Ile), 85T>C (**9*, Cys29Arg), and 1003G>T (**11*, Val335Leu), were found in this study at allele frequencies of 0.283, 0.015, 0.029, and 0.0015, respectively. The allele frequencies of two reported SNPs, 496A>G (Met166Val) and 2303C>A (Thr768Lys), were 0.022 and 0.028, respectively. Recently, 1774C>T (Arg592Trp) was reported from a Korean population (Cho et al. 2007), and its allele frequency was 0.0015 in this study. Nine intronic variations, IVS10-15T>C, IVS13 + 39C>T, IVS13 + 40 G>A, IVS15 + 75A>G, IVS16-94G>T, IVS18-39G>A, IVS21 + 136G>C, IVS22-58G>C, and IVS22-69G>A, and one synonymous variation, 1896T>C (Phe632Phe), were found with various allele frequencies (0.003–0.378, Table 2). The variations previously detected in Japanese (Kouwaki et al. 1998; Yamaguchi et al. 2001; Ogura et al. 2005), 62G>A (Arg21Gln, **12*), 74G>A (His25Arg), 812delT (Leu271X), 1097G>C (Gly366Ala), 1156G>T (Glu386X, **12*), and 1714C>G (Leu572Val), were not found in our study. This might be due to their low frequencies.

Linkage disequilibrium (LD) analysis and haplotype block partition

LD analysis was performed by r^2 and |D′| using 18 SNPs (allele frequency ≥0.01) (Fig. 2). Strong linkages were observed in four pairs of SNPs: between -477T>G and 85T>C (Cys29Arg) ($r^2 = 0.7025$), between 496A>G (Met166Val) and IVS10-15T>C ($r^2 = 0.7964$), between 1627A>G (Ile543Val) and IVS13 + 39C>T ($r^2 = 1.0$), and between IVS14-123C>A and IVS15 + 75A>G ($r^2 = 1.0$). In addition, two known rare SNPs, IVS22-69G>A (rs290855) and IVS22-58G>C (rs17116357), were perfectly linked ($r^2 = 1.0$) (data not shown). As for |D′| values, only 43 pairs (28%) out of 153 pairs gave |D′| = 1.0, indicating that a number of recombinations had occurred within this gene. This is not surprising because *DPYD* is a huge gene of at least 950 kb in length with 3 kb of coding sequences. However, it was difficult to estimate past recombination events in *DPYD* from our data alone because our variations were mostly limited to exons and surrounding introns.

To define haplotype blocks, we utilized the HapMap data because SNPs were comprehensively genotyped with an average density of 1 SNP per 1.8 kb. Of 1,002 variations of *DPYD* genotyped by the HapMap project, 474 SNPs were polymorphic for 44 unrelated Japanese subjects. When the LD profiles for Japanese were obtained by Marker using the HapMap data, strong LD (|D′| > 0.75) clearly decays within introns 11, 12, 13, 14, 16, 18, and 20 (data not shown), suggesting that recombination had occurred in these regions. Based on these findings, the SNPs detected in our study were divided into six haplotype blocks (Figs. 1, 2). Block 1, the largest block, ranges from the 5′-untranslated region (5′-UTR) to intron 10 (347 kb), and includes 22 variations. Block 2 includes eight variations from IVS12-11G>A in intron 12 to IVS13 + 40G>A in intron 13. Block 3 includes six variations from IVS13-47_48insTA in intron 13 to IVS14 + 100T>G in intron 14. Block 4 contains only three SNPs, IVS14-123C>A, IVS14-21C>A and IVS15 + 75A>G, and ranges from intron 14 to intron 15. Block 5 consists of IVS16-94G>T and four rare variations from intron 16 to exon 18. Although the HapMap data showed a decline in LD in intron 20, we defined a block ranging from intron 18 to intron 22 as block 6 because only rare variations (allele frequencies <0.01) were detected downstream of intron 20 (exon 21, intron 21, and intron 22). The block partitioning based on the HapMap data fitted our SNPs well: more than 70% of SNP pairs in each block (block 1–6) gave pair-wise |D′| values greater than 0.8 (Fig. 2).

Haplotype estimation

Using 22, 8, 6, 3, 5, and 11 variations in blocks 1 to 6, 23 (block 1), 8 (block 2), 7 (block 3), 3 (block 4), 6 (block 5), and 11 (block 6) haplotypes were identified or inferred (Fig. 3). Probabilities of diplotype configurations in all six blocks were 100% for over 97% of the subjects. To discriminate our block haplotypes from the previously assigned alleles or haplotypes (*DPYD*1* to **13*), the mark, #, was used to indicate block haplotypes.

In block 1, the most dominant haplotype without any variation was #*1a* (0.818 in frequency), followed by #*1b* (0.045), #*9c* (0.022), and #*1c* (0.021). As suggested by LD (Fig. 2), #*9c*, the major subtype of the #*9* group bearing 85T>C (Cys29Arg), also harbored -477T>G in the 5′-UTR. Known nonsynonymous SNP, 496A>G (Met166Val), was assigned to three haplotypes, #*9d*, #*166Va*, and #*166Vb*.

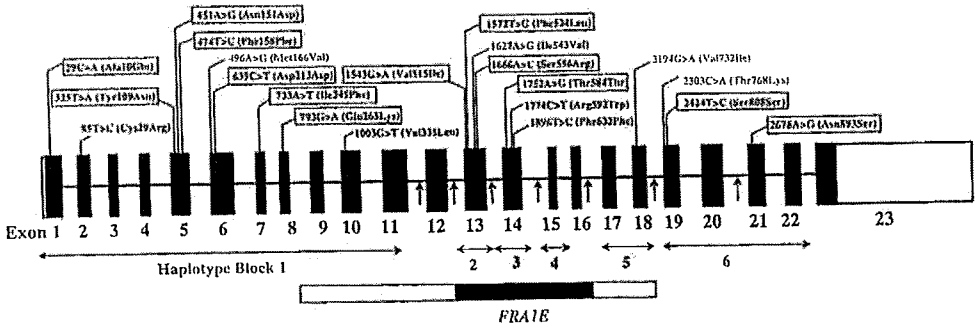

Fig. 1 Twenty-one variations detected in the coding exons are depicted in the schematic diagram of the *DPYD* gene. Fourteen novel variations are enclosed by *squares*. The recombination spots were estimated based on the LD profiles obtained from Japanese data in the HapMap project and indicated by *arrows*. The borders (between introns 8 and 18 of the *DPYD*) and core region (between introns 12 and 16) of *FRA1E* identified by Hormozian et al. (2007) are indicated as an *open* and *closed box*, respectively

In block 2, four haplotypes, #*1a* (0.529), #*5a* (0.245), #*1b* (0.176), and #*5b* (0.038), were major in Japanese and accounted for 99% of all inferred haplotypes. Two subtypes of the #*5* group, #*5a* and #*5b*, both of which harbored Ile543Val (*5) and IVS13 + 39C>T, were distinguished by a novel intronic SNP, IVS12-11G>A.

As for block 3, in addition to #*1a* (0.848), #*1b* harboring the synonymous SNP, 1896T>C (Phe632Phe), was found at a relatively high frequency (0.138).

Block 4 is simple and comprises only three haplotypes, #*1a* (0.845), #*1b* (0.154) and #*1c* (0.0015). The second frequent haplotype, #*1b*, harbored perfectly linked SNPs, IVS14-123C>A and IVS15 + 75A>G.

Block 5 contained IVS16-94G>T, the most frequent SNP among the 55 SNPs found in this study, which was assigned to #*1b* with a frequency of 0.374. This block also contained the known nonsynonymous SNP, 2194G>A (Val732Ile, *6), which was assigned to #*6a* (0.015).

In block 6, the most dominant haplotype was #*1a* (0.915). It was followed by #*1b* (0.032) with IVS18-39G>A and #*768K* (0.028) with 2303C>A (Thr768Lys).

The HapMap data include nine SNPs that we detected (Table 2). Of them, six, 85T>C (rs1801265), 496A>G (rs2297595), 1627A>G (rs1801159), 1896T>C (rs17376848), IVS16-94G>T (rs7556439) and IVS18-39G>A (rs12137711), were suitable for haplotype tagging SNPs (htSNPs) to capture the block haplotypes, block 1 #*9*, block 1 #*166V*, block 2 #*5*, block 3 #*1b*, block 5 #*1b*, and block 6 #*1b*, respectively. IVS21 + 136G>C (rs11165777) and IVS22-69G>A (rs290855)/IVS22-58G>C (rs1711 6357), were the marker SNPs for block 6 #*1e* and #*1f*, respectively, but very rare (allele frequencies = 0.003) in Japanese. The six SNPs, especially 85T>C (rs1801265) and 496A>G (rs2297595), were in strong LD ($r^2 > 0.8$) with other HapMap SNPs in Japanese (Table 3), indicating that many HapMap SNPs were concurrently linked on the same haplotypes.

Next, the combinations of block haplotypes (inter-block haplotypes) were analyzed focusing on the haplotypes with frequencies of >0.01 in each block (Fig. 4). Between blocks 1 and 2, both #*1a* and #*1b* in block 1 were complicatedly associated with various haplotypes in block 2. It should be noted that #*9c* in block 1 was linked either with block 2 #*1b* (0.016 in absolute frequency) or with block 2 #*5a* (0.006, not shown in Fig. 4). #*1c* in block 1 was completely linked with block 2 #*1a*. #*151D* in block 1 (not shown in Fig. 4), which was a rare haplotype (0.009) harboring 451A>G (Asn151Asp), was completely linked with #*5a* in block 2.

Between blocks 2 and 3, both #*5b* and #*1b* in block 2 were mostly linked with #*1a* in block 3, whereas both #*1a* and #*5a* in block 2 were complicatedly linked with #*1a*, #*1b*, or other rare haplotypes such as #*1c* (not shown in Fig. 4) in block 3. Between blocks 3 and 4 and between blocks 4 and 5, no strong associations of block haplotypes were observed except for the linkage of block 5 #*6a* to block 4 #*1a*. Between blocks 5 and 6, most of #*1b* and all of #*6a* in block 5 were linked with #*1a* in block 6. Although #*1a* in block 6 was associated with various haplotypes in block 5, #*1b* in block 6 was completely linked with #*1a* in block 5.

Among the six blocks, the following combinations were major: #*1a* (block 1)–#*1a* (block 2) –#*1a* (block 3)–#*1a* (block 4)–#*1a* (block 5)–#*1a* (block 6) (0.239 in frequency), #*1a*–#*5a*–#*1a*–#*1a*–#*1b*–#*1a* (0.081), #*1a*–#*1a*–#*1a*–#*1a*–#*1b*–#*1a* (0.075), #*1a*–#*5a*–#*1a*–#*1a*–#*1a*–#*1a* (0.070), #*1a*–#*1b*–#*1a*–#*1a*–#*1a*–#*1a* (0.060) and #*1a*–#*1a*–#*1b*–#*1a*–#*1a*–#*1a* (0.051).

Ethnic differences in distributions of *DPYD* SNPs and haplotypes

We compared SNP and haplotype distributions in Japanese with those in other ethnic groups reported in the literature

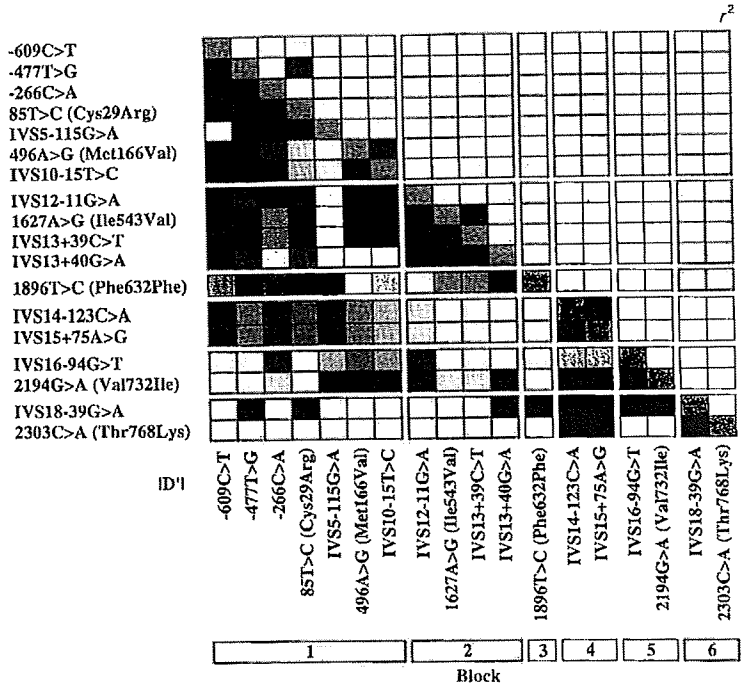

Fig. 2 Linkage disequilibrium (*LD*) analysis of *DPYD*. Pairwise LD between 18 common SNPs (>0.01 in allele frequencies) is expressed as r^2 (*upper*) and ID'I (*lower*) by a 10-graded *blue color*. The *denser color* indicates higher linkage. The haplotype block partition based on LD measure ID'I of HapMap data in Japanese is also indicated

or HapMap project. Notably, IVS14 + 1G>A (*2), 1897delC (Pro633GlnfsX5, *3), 1601G>A (Ser534Asn, *4), 295_298delTCAT (Phe100SerfsX15, *7), 703C>T (Arg235Trp, *8), 2983G>T (Val995Phe, *10), 62G>A (Arg21Gln, *12), 1156G>T (Glu386X, *12), and 1679T>G (Ile560Ser, *13) were not found in this study. Furthermore, several SNPs showed marked differences in allele frequencies among Japanese and other ethnic groups (Table 4).

The allele frequency of 85T>C (Cys29Arg, *9), the tagging SNP for block 1 #9, was quite different between Asians and Caucasians. Its allele frequency in Japanese (0.029 in this study) and Taiwanese (0.022) (Hsiao et al. 2004) was much lower than that in Caucasians (0.185–0.194) (Seck et al. 2005; Morel et al. 2006).

The SNP 496A>G (Met166Val) in block 1 is found at a lower allele frequency in Japanese (0.022) than in Caucasians (0.080) (Seck et al. 2005). Seck et al. (2005) inferred two haplotypes harboring 496A>G (Met166Val) from 157 Caucasians: *hap5* (#9d in this study) harboring additional 85T>C (Cys29Arg) and IVS10-15T>C and *hap11* concurrently harboring IVS10-15T>C alone with frequencies of 0.040 and 0.014, respectively. In our haplotype analysis, #166Va (0.012) corresponding to *hap11* (0.014) was found with a similar frequency in Japanese, whereas the frequency of #9d (0.006) was much lower than that of the corresponding haplotype, *hap5* (0.040) in Caucasians.

1627A>G (Ile543Val, *5) in block 2 was found with comparable allele frequencies among Japanese (0.283 in this study), Caucasians (0.14-0.275) (Seck et al. 2005; Ridge et al. 1998a), African-Americans (0.227) (Wei et al. 1998), and Taiwanese (0.210–0.283) (Wei et al. 1998; Hsiao et al. 2004).

The allele frequency (0.015) of 2194G>A (Val732Ile, *6) in block 5 in our Japanese population is slightly lower than that previously reported in Caucasians (0.022-0.058) (Seck et al. 2005; Ridge et al. 1998a) and Finish (0.067) (Wei et al. 1998), but is comparable to that in Taiwanese (0.012-0.014) (Wei et al. 1998; Hsiao et al. 2004) and African-Americans (0.019) (Wei et al. 1998).

Ethnic differences in the allele frequencies were also observed with synonymous and intronic variations (Table 4). The allele frequency of 1896T>C (Phe632Phe), which tags block 3 #1b, was higher in Japanese (0.139 in this study) than in Caucasians (0.035) (Seck et al. 2005). *Hap13* assigned in 157 Caucasians by Seck et al. (2005) is the counterpart of block 3 #1b, and its frequency (0.012) was much lower than that in Japanese (0.138).

In contrast, IVS10-15T>C linked to 85T>C (*9) or 496A>G (#166V) within block 1 showed a lower allele frequency in Japanese (0.018) than in Caucasians (0.127). Seck et al. (2005) assigned *hap7* as the haplotype containing IVS10-15T>C alone with a haplotype frequency of 0.03 in Caucasians. In Japanese, however, the corresponding haplotype was not found.

Allele frequencies of IVS18-39G>A and IVS22-69G>A, which are tagging SNPs for block 6 #1b and #1f, respectively, are lower in Japanese (0.032 and 0.003, respectively) than in Caucasians (0.105 and 0.183, respectively).