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 |D'| 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	TTGAGGAGTTCC/TGGAAAATGCAGTT			0.026
MPJ6_DPD003[a]		5′-Flank	52206137	–266	CTCCCTCCCTCC/ATTCTGCTTGCAG			0.045
MPJ6_DPD004[a]		5′-Flank	52206114	–243	AGGCTGGGGCGC/GAGAGCGGGCTGAA			0.0059
MPJ6_DPD005[a]		Exon 1	52205843	29	GTAAGGACTCGG/CAGGACATCGAGGT	Ala10Glu		0.0015
MPJ6_DPD006[g]	rs1801265	Exon 2	52168278	85	CATGCAACTCTG/TCGTTCCACTTCGG	Cys29Arg	*9	0.029
MPJ6_DPD007[a]		Intron 2	52168055	IVS2 + 158	TTTGAAAGTGTA/TCTTTTAATTACAC			0.0015
MPJ6_DPD008[a]		Intron 3	52113040	IVS3 + 23	GTCACCATAGCAA/GCAGTCACAGATG			0.0029
MPJ6_DPD009[a]		Exon 5	52006617	325	ATTTTGCAGAAC/TAATTATGGAGCTG	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*.

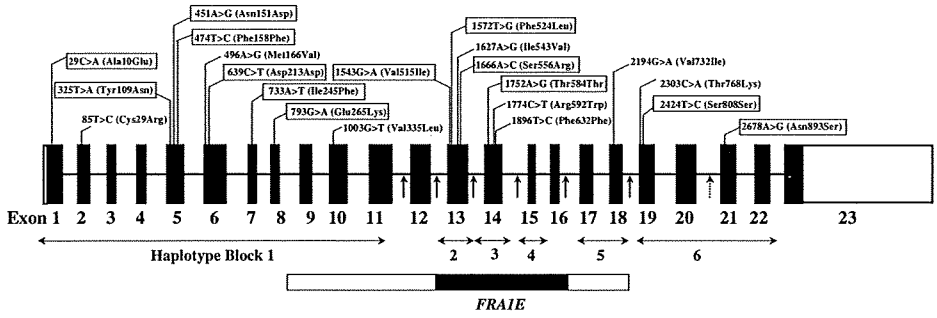

Fig. 1 Twenty-one variations detected in the coding exons are depicted in the schematic diagram of the *DPYD* gene. Fourteen novel variations are enclosed by *squares*. The recombination spots were estimated based on the LD profiles obtained from Japanese data in the HapMap project and indicated by *arrows*. 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

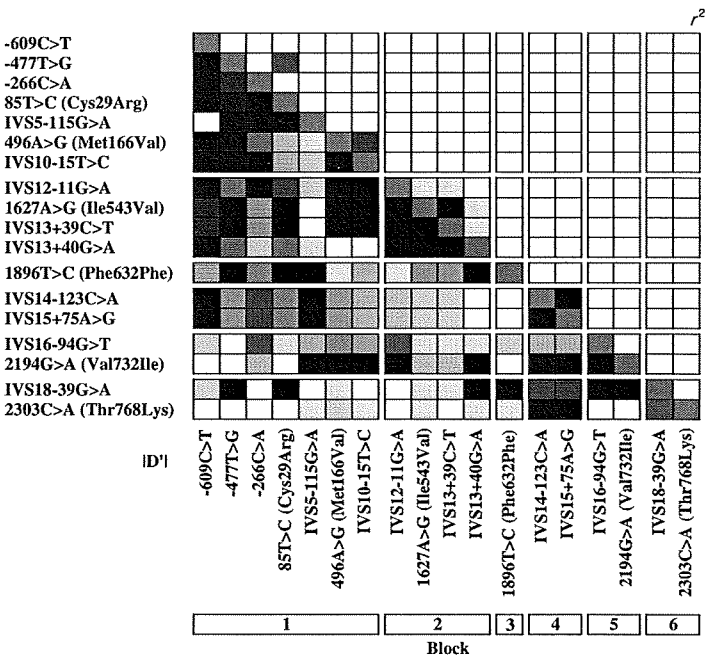

Fig. 2 Linkage disequilibrium (*LD*) analysis of *DPYD*. Pairwise LD between 18 common SNPs (>0.01 in allele frequencies) is expressed as r^2 (*upper*) and |D'| (*lower*) by a 10-graded *blue color*. The *denser color* indicates higher linkage. The haplotype block partition based on LD measure |D'| of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

a) Block 1

Nucleotide change			-609C>T	-477T>G	-266C>A	-243G>A	29C>A	85T>C	IVS2+158 T>C	IVS3+23 A>G	325T>A	451A>G	474T>C	IVS5-115G>A	496A>G	639C>T	733A>T	IVS7+64 G>T	793G>A	IVS8+91 C>T	IVS9-120A>T	1003G>T	IVS10+24 A>G	IVS10-15T>C	Number	Frequency
Amino acid change							A10E	C29R			Y109N	N151D	F158F		M166V	D213D	I245F		E265K			V335L				
Haplotype	*1	*1a																							558	0.818
		*1b																							31	0.045
		*1c																							14	0.021
		*1d																							13	0.019
		*1e																							6	0.0088
		*1f																							6	0.0088
		*1g																							3	0.0044
		*1h																							2	0.0029
		*1i																							2	0.0029
		*1j																							2	0.0029
		*1k																							1	0.0015
		*1l																							1	0.0015
	*9	*9c																							15	0.022
		*9d																							4	0.0059
		*9e§																							1	0.0015
	*11	*11a																							1	0.0015
	*166V	*166Va																							8	0.012
		*166Vb																							3	0.0044
	*151D	*151D																							6	0.0088
	*109N	*109N																							2	0.0029
	*10E	*10E																							1	0.0015
	*265K	*265K																							1	0.0015
	*245F	*245F§																							1	0.0015

b) Block 2

Nucleotide change			IVS12-11G>A	IVS12-9A>G	1543G>A	1572T>G	1627A>G	1666A>C	IVS13+39 C>T	IVS13+40 G>A	Number	Frequency
Amino acid change					V515I	F524L	I543V	S556R				
Haplotype	*1	*1a									361	0.529
		*1b									120	0.176
		*1c									5	0.0073
	*5	*5a					5				167	0.245
		*5b					5				26	0.038
	*515I	*515I			515I						1	0.0015
	*524L	*524L				524L					1	0.0015
	*556R	*556R§						556R			1	0.0015

c) Block 3

Nucleotide change			IVS13-47_48insTA	1752A>G	1774C>T	1896T>C	IVS14+19 C>A	IVS14+100 T>G	Number	Frequency
Amino acid change				T584T	R592W	F632F				
Haplotype	*1	*1a							578	0.848
		*1b							94	0.138
		*1c							6	0.0088
		*1d							1	0.0015
		*1e							1	0.0015
		*1f§							1	0.0015
	*592W	*592W			592W				1	0.0015

d) Block 4

Nucleotide change			IVS14-123C>A	IVS14-21C>A	IVS15+75 A>G	Number	Frequency
Amino acid change							
Haplotype	*1	*1a				576	0.845
		*1b				105	0.154
		*1c§				1	0.0015

e) Block 5

Nucleotide change			IVS16-127A>G	IVS16-94G>T	IVS17+34del T	IVS17+47C>T	2194G>A	Number	Frequency
Amino acid change							V732I		
Haplotype	*1	*1a						414	0.607
		*1b						255	0.374
		*1c§						1	0.0015
		*1d§						1	0.0015
		*1e§						1	0.0015
	*6	*6a					6	10	0.015

f) Block 6

Nucleotide change			IVS18-39G>A	2303C>A	2424T>C	2678A>G	IVS21+80 C>G	IVS21+113 T>A	IVS21+136 G>C	IVS21+162 T>G	IVS22+129 A>G	IVS22-69 G>A	IVS22-58 G>C	Number	Frequency
Amino acid change				T768K	S808S	N893S									
Haplotype	*1	*1a												624	0.915
		*1b												22	0.032
		*1c												5	0.0073
		*1d												2	0.0029
		*1e												2	0.0029
		*1f												2	0.0029
		*1g												2	0.0029
		*1h												1	0.0015
		*1i												1	0.0015
	*768K	*768K												19	0.028
	*893S	*893S												2	0.0029

Fig. 3 Block haplotypes in *DPYD* of block 1 (**a**), block 2 (**b**), block 3 (**c**), block 4 (**d**), block 5 (**e**), and block 6 (**f**) in a Japanese population. The nucleotide positions were numbered based on the cDNA sequence (A of the translational start codon is +1) or from the nearest exon. *White cell* wild-type, *gray cell* nucleotide alteration. §The haplotypes were inferred in only one patient and ambiguous except for marker SNPs

Table 3 Linkages of haplotype-tagging SNPs with HapMap SNPs for *DPYD*

Haplotype-tagging SNPs in *DPYD*	dbSNP ID (NCBI)	Block haplotype in this paper	HapMap SNPs with close linkages ($r^2 > 0.8$)[a]
85T>C (Cys29Arg)	rs1801265	Block 1 #*9*	rs10747488, rs7526108, rs4421623, rs4379706, rs4523551,rs11165921, rs9661794, rs6677116, rs6604093, rs17379561, rs10747491, rs10747492, rs12062845, rs7524038, rs10875112, rs4394693, rs10875113, rs4970722, rs9727548, rs10875118, rs9662719, rs12077442, rs4394694, rs9727976, rs4246515, rs6692580
496A>G (Met166Val)	rs2297595	Block 1 #*166V*	rs2786543, rs2811215, rs2811214, rs2786544, rs2248658, rs11165897, rs2786490, rs2811203, rs2811202, rs2811200, rs2811198, rs2786503, rs2811196, rs2786505, rs2811195, rs2811194, rs12073839, rs6663670, rs7512910, rs2151563, rs2786509, rs3790387, rs3790389
1627A>G (Ile543Val)	rs1801159	Block 2 #*5*	rs1415682, rs952501, rs2811187, rs2786778, rs2786774, rs2811183, rs17116806, rs2786780, rs1801159, rs2786771, rs2297780, rs2297779, rs12729863
1896T>C (Phe632Phe)	rs7556439	Block 3 #*1b*	rs12073650
IVS16-94G>T	rs7556439	Block 5 #*1b*	rs693680, rs827500, rs499009, rs7518848, rs553388, rs507170, rs628959, rs991544, rs526645, rs1609519
IVS18-39G>A	rs12137711	Block 6 #*1b*	rs12120068, rs12116905

[a] All SNPs are in the same block

Taken together, our data demonstrated considerable differences in the haplotype distributions in blocks 1, 3 and 6 between Japanese and Caucasians.

Discussion

This study provides Japanese data on the genetic variations of *DPYD*, a gene encoding a key enzyme catalyzing degradation of the well-known anticancer drug 5-FU. Nine novel (Ala10Glu, Tyr109Asn, Asn151Asp, Ile245Phe, Glu265Lys, Val515Ile, Phe524Leu, Ser556Arg, and Asn893Ser) and seven known nonsynonymous variations (Cys29Arg, Met166Val, Val335Leu, Ile543Val, Arg592Trp, Val732Ile, and Thr768Lys) were found in our Japanese population (Table 2 and Fig. 1). The association analysis between the genotypes and 5-FU pharmacodynamics is now on-going.

Uneven distributions of coding SNPs over 23 *DPYD* exons were pointed out in the previous review by van Kuilenburg (2004). The author indicated that 81% of all reported variations were confined to exons 2–14, representing 61% of the coding sequences, and typical hotspots of variation were localized in exons 2, 6, and 13. Our Japanese data also revealed that 17 out of 21 coding variations (81%) were localized in exons 1–14, and that more than three variations were detected in exons 5, 13, and 14 (Fig. 1). Recently, Hormozian et al. (2007) have reported that the common chromosomal fragile site on 1p21.2, *FRA1E*, spans 370 kb of genomic sequence between introns 8 and 18 of *DPYD*, and that its core region with the highest fragility is located between introns 12 and 16. The instability at the core of *FRA1E* might be associated with the high mutational rates and recombinogenic nature from intron 12 to 14 of *DPYD* (Fig. 1).

To estimate potential functional consequences of the amino acid substitutions, we examined whether the positions of amino acid changes are located in highly conserved areas or potentially critical regions of the molecule (for example, substrate recognition sites or binding regions of prosthetic groups). We also considered the locations of the residues in a three-dimensional (3D) framework provided by the crystal structures of pig DPD, which have recently been determined in complexes with NADPH and substrate (5-FU) (Dobritzsch et al. 2001) or inhibitors (Dobritzsch et al. 2002). The amino acid sequences of pig and human DPD are 93% identical (Mattison et al. 2002), and the substituted residues and their neighboring residues are conserved between both enzymes. From these points of view, it is speculated that at least two substitutions (Glu265Lys and Arg592Trp) might impact the structure and function of DPD as discussed below.

Glu265 is located on the loop following to the third β sheet (IIβ3) in the FAD binding domain II (Dobritzsch et al. 2001). Glu265 is conserved among four mammalian species (human, mouse, rat, and pig), although it is replaced with aspartic acid in bovine and *Drosophila melanogaster* DPDs (Mattison et al. 2002). In the 3D structure of pig DPD (Fig. 5a), Glu265 is in close proximity to Lys259. The substitution, Lys259Glu, was

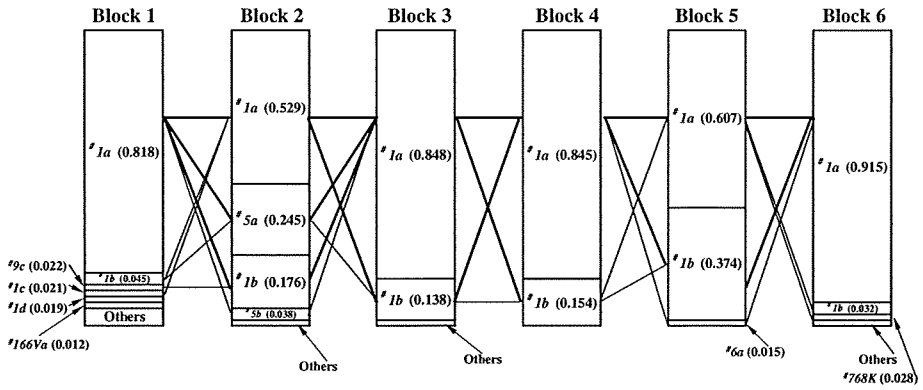

Fig. 4 The combinations of block haplotypes in Japanese. *Thick lines* represent combinations with frequencies over 10%, and *thin lines* represent combinations with frequencies of 1.0–9.9%

detected in the patient exhibiting severe mucositis during cyclophosphamide/methotrexate/5-FU chemotherapy (Gross et al. 2003). Furthermore, the adjacent Leu261 interacts via the main chain atoms with the N6, N1, and N3 atoms of adenine of FAD, and has an important role in the proper orientation of the adenine moiety in the FAD-binding pocket (Dobritzsch et al. 2001). Moreover, the carboxyl group (Glu265-Oε)might form hydrogen bonds to the main chain nitrogen of Ser260 next to Leu261. Thus, the change in polarity from negative to positive by the novel Glu265Lys substitution is likely to cause structural changes affecting proper binding of FAD.

Arg592 is located at one (IVβc) of the additional four-stranded antiparallel β sheets (IVβc-βf) inserted at the top of a typical $(\alpha/\beta)_8$ barrel fold in the FMN-binding domain IV (Dobritzsch et al. 2001). Arg592 is completely conserved among the above-mentioned six species (Mattison et al. 2002), suggesting its functional importance. Arg592 closely contacts Met599 (2.9 Å) and Gln604 (2.8 Å) in the same subunit and Ser994 (2.9 Å) in another subunit (Fig. 5B). The substitution of tryptophan for Arg592 is likely to weaken these interactions due to altered hydrophobicity and electrostatic changes. Arg592Trp was recently reported from a Korean population with an allele frequency of 0.004, although its functional significance remains to be confirmed(Cho et al. 2007).

As for known *DPYD* alleles, their distributions in several populations are becoming more evident by recent reports. For example, IVS14 + 1G>A (*2) (van Kuilenburg 2004), 295_298delTCAT (Phe100SerfsX15, *7) (Seck et al. 2005), 1679T>G (Ile560Ser, *13) (Collie-Duguid et al. 2000; Morel et al. 2006) 2846A>T (Asp949Val) (Seck et al. 2005; Morel et al. 2006), all of which are associated with decreased DPD activities, are detected in Caucasians with allele frequencies of 0.01–0.02, 0.003, 0.001 and 0.006–0.008, respectively. However, none of them were detected in our Japanese samples, while 1003G>T (Val335Leu, *11) and 2303C>A (Thr768Lys) have been found only in Japanese, indicating that variations with clinical relevance do not overlap between Caucasians and Japanese.

2303C>A (Thr768Lys), which was originally found in a Japanese female volunteer with very low DPD activity (Ogura et al. 2005), is relatively frequent in Japanese (allele frequency = 0.0279). Functional characterization in vitro revealed that 768Lys caused thermal instability of the variant protein without changing its affinity for NADPH or kinetic parameters toward 5-FU. Therefore, they might cause 5-FU-related toxicities in Japanese.

1003G>T (Val335Leu, *11) was found in a Japanese family with decreased DPD activity by Kouwaki et al. (1998). By in vitro expression in *E. coli*, they demonstrated that the variant protein with Leu335 showed a significant loss of activity (about 17% of the wild-type protein). Dobritzsch et al. (2001) suggested from the 3D structure of pig DPD that Val335Leu, in spite of a conservative change, disturbs packing interactions in the hydrophobic core formed by IIIβ3 and IIIα3 within the Rossman-motif, thereby affecting NADPH binding. In our study, heterozygous 1003G>T (Val335Leu) was found from a patient administrated 5-FU (allele frequency = 0.0015), who also has seven other variations: IVS12–11G>A, 1896T>C (Phe632Phe), and IVS16–94G>T are heterozygous, and 1627A>G (Ile543Val), IVS13 + 39C>T, IVS14–123C>A, and IVS15 + 75A>G are homozygous, indicating that at least Val335Leu is linked to Ile543Val (*5).

On the other hand, Caucasians and Japanese share four variations: *5 (Ile543Val), *9 (Cys29Arg), Met166Val, and *6 (Val732Ile), although their allele frequencies were different, especially for *9 (Table 4). Because they have not necessarily correlated with phenotypic changes (e.g., differences in DPD enzyme activity, 5-FU pharmacokinetics and pharmacodynamics) (Collie-Duguid et al. 2000; Johnson et al. 2002; Zhu et al. 2004; Seck et al. 2005; Ridge et al. 1998a, 1998b; Hsiao et al. 2004), all of these variations are generally accepted as common polymorphisms that result in unaltered function. Consistent with this, van Kuilenburg et al. (2002) suggested that the

Table 4 Allele frequencies of common *DPYD* SNPs in different populations

Nucleotide change (amino acid change)	Allele or tagged haplotypes	Population	Allele frequency	Number of subjects	Reference
85T>C	*9	Caucasian	0.194	157	Seck et al. 2005
(Cys29Arg)	(Block 1 #9)	French Caucasian	0.185	487	Morel et al. 2006
		Japanese	0.037	107	Yamaguchi et al. 2001
		Japanese	0.029	341	This study
		Taiwanese	0.022	300	Hsiao et al. 2004
496A>G	Block 1 #166V	Caucasian	0.080	157	Seck et al. 2005
(Met166Val)		Japanese	0.022	341	This study
IVS10-15T>C	Block 1 #166Va, #9d	Caucasian	0.127	157	Seck et al. 2005
		Japanese	0.018	341	This study
1627A>G	*5	Caucasian	0.140	157	Seck et al. 2005
(Ile543Val)	(Block 2 #5)	Caucasian	0.275	60	Ridge et al. 1998a
		Finnish	0.072	90	Wei et al. 1998
		African-American	0.227	105	Wei et al. 1998
		Japanese	0.352	50	Wei et al. 1998
		Japanese	0.283	341	This study
		Taiwanese	0.210	131	Wei et al. 1998
		Taiwanese	0.283	300	Hsiao et al. 2004
1896T>C	Block 3 #1b	Caucasian	0.035	157	Seck et al. 2005
(Phe632Phe)		Japanese	0.098	107	Yamaguchi et al. 2001
		Japanese	0.139	341	This study
		Han Chinese	0.133	45	HapMap
IVS15 + 75A>G	Block 4 #1b	Caucasian	0.166	157	Seck et al. 2005
		Japanese	0.155	341	This study
IVS16-94G>T	Block 5 #1b	Caucasian	0.415	59	HapMap
		Yorba	ND	60	HapMap
		Japanese	0.455	44	HapMap
		Japanese	0.378	341	This study
		Han Chinese	0.333	45	HapMap
2194G>A	*6	Caucasian	0.022	157	Seck et al. 2005
(Val732Ile)	(Block 5 #6)	Caucasian	0.058	60	Ridge et al. 1998a
		Finnish	0.067	90	Wei et al. 1998
		African-American	0.019	105	Wei et al. 1998
		Japanese	0.044	50	Wei et al. 1998
		Japanese	0.015	341	This study
		Taiwanese	0.014	131	Wei et al. 1998
		Taiwanese	0.012	300	Hsiao et al. 2004
IVS18-39G>A	Block 6 #1b	Caucasian	0.105	157	Seck et al. 2005
		Caucasian	0.100	60	HapMap
		Yorba	0.017	60	HapMap
		Japanese	0.044	45	HapMap
		Japanese	0.032	341	This study
		Han Chinese	0.022	45	HapMap
IVS22-69G>A	Block 6 #1f	Caucasian	0.183	60	HapMap
		Yorba	0.400	60	HapMap
		Japanese	ND	45	HapMap
		Japanese	0.003	341	This study
		Han Chinese	ND	45	HapMap

ND not detected

substitution Cys29Arg on the protein surface was unlikely to alter DPD activity. However, conflicting results were reported regarding **9* (Vreken et al. 1997, van Kuilenburg et al. 2000), **6* (van Kuilenburg et al. 2000), and Met166Val (van Kuilenburg et al. 2000; Gross et al. 2003). To interpret these inconsistencies, haplotype analysis of *DPYD* might be helpful. Especially for **9* and Met166Val in Japanese, functional involvement of –477T>G (block 1 #9c and #9e), –243G>A (block 1 #9d), IVS10-15T>C (block 1 #9d and #166Va) and many other HapMap SNPs linked to **9* and Met166Val (Table 3) needs clarification.

The HapMap project provides genotype data of more than 1,000 sites located mostly in the intronic regions of *DPYD* for four different populations (Nigerian, Chinese,

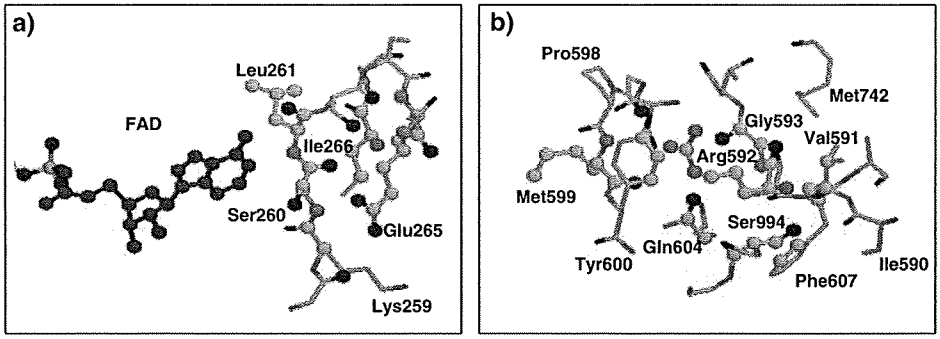

Fig. 5 Stereo view of the variation sites in pig DPD (accession code of the Protein Data Bank: 1gth). Glu265 (**a**), Arg592 (**b**) and their adjacent residues are shown as *ball-and-stick* models with oxygens in *red*, nitrogens in *blue*, carbons in *gray* and sulfur in *yellow*. The adenosine moiety of the cofactor FAD is also shown in *pink* (**a**)

Japanese and Caucasians). HapMap data on 44 unrelated Japanese subjects showed that 476 variations are polymorphic, whereas 529 are monomorphic, and the average density of polymorphic markers is 1 SNP per 1,772 bp. In contrast, our study focused on exons and surrounding introns to detect variations, and only nine variations overlapped with the HapMap data. Therefore, we could not utilize the HapMap data to further identify common subtypes of *#1* to be discriminated by many intronic HapMap SNPs in each block. However, most of the frequent SNPs are unlikely to be associated with substantially decreased DPD activity because DPD activity in the healthy Japanese population (N = 150) showed a unimodal Gaussian distribution (Ogura et al. 2005).

On the other hand, in 60 unrelated Caucasian subjects in the HapMap project, 617 are polymorphic, whereas 383 are monomorphic. LD profiles of these polymorphisms were compared between Caucasians and Japanese by using the program Marker (http://www.gmap.net/marker). Strong LD ($|D'| > 0.75$) clearly decays within introns 11, 12, 13, 14, 16, 18, and 20 in Japanese, whereas, similar decays are observed within introns 13, 14, 18, and 20, but are not obvious within introns 11, 12, and 16 in Caucasians (data not shown). Moreover, strong LD decays within intron 3 in Caucasians. Therefore, the LD blocks are considerably different between Japanese and Caucasians. Along with the marked differences in allele frequencies of several variations (Table 4), these results suggest that the haplotype structures in *DPYD* are quite different between the two populations.

In conclusion, we found 55 variations, including 38 novel ones, in *DPYD* from 341 Japanese subjects. Nine novel nonsynonymous SNPs were found, some of which were assumed to have impact on the structure and function of DPD. As for known variations, we obtained their accurate allele frequencies in a Japanese population of a large size and showed that variations with clinical relevance do not overlap between Caucasians and Japanese. In Japanese, 2303C>A (Thr768Lys) and 1003G>T (Val335Leu) might play important roles in 5-FU-related toxicity. Along with differences in haplotype structures between Japanese and Caucasians, these findings suggest that ethnic-specific tagging SNPs should be considered on genotyping *DPYD*. Thus, the present information would be useful for pharmacogenetic studies for evaluating the efficacy and toxicity of 5-FU in Japanese and probably in East Asians.

Acknowledgments We thank Ms. Chie Sudo for her secretarial assistance. This study was supported in part by the Program for the Promotion of Fundamental Studies in Health Sciences (05–25) of the National Institute of Biomedical Innovation and in part by the Health and Labor Sciences Research Grants from the Ministry of Health, Labor and Welfare.

References

Bakkeren JA, De Abreu RA, Sengers RC, Gabreels FJ, Maas JM, Renier WO (1984) Elevated urine, blood and cerebrospinal fluid levels of uracil and thymine in a child with dihydrothymine dehydrogenase deficiency. Clin Chim Acta 140:247–256

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Cho HJ, Park YS, Kang WK, Kim JW, Lee SY (2007) Thymidylate synthase (TYMS) and dihydropyrimidine dehydrogenase (DPYD) polymorphisms in the Korean population for prediction of 5-fluorouracil-associated toxicity. Ther Drug Monit 29:190–196

Collie-Duguid ES, Etienne MC, Milano G, McLeod HL (2000) Known variant DPYD alleles do not explain DPD deficiency in cancer patients. Pharmacogenetics 10:217–223

Dobritzsch D, Schneider G, Schnackerz KD, Lindqvist Y (2001) Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. Embo J 20:650–660

Dobritzsch D, Ricagno S, Schneider G, Schnackerz KD, Lindqvist Y (2002) Crystal structure of the productive ternary complex of dihydropyrimidine dehydrogenase with NADPH and 5-iodouracil. Implications for mechanism of inhibition and electron transfer. J Biol Chem 277:13155–13166

Etienne MC, Lagrange JL, Dassonville O, Fleming R, Thyss A, Renee N, Schneider M, Demard F, Milano G (1994) Population study of dihydropyrimidine dehydrogenase in cancer patients. J Clin Oncol 12:2248–2253

Gross E, Ullrich T, Seck K, Mueller V, de Wit M, von Schilling C, Meindl A, Schmitt M, Kiechle M (2003) Detailed analysis of five mutations in dihydropyrimidine dehydrogenase detected in

cancer patients with 5-fluorouracil-related side effects. Hum Mutat 22:498

Grem JL (1996) Fluoropyrimidines. In: Chabner BA, Longo DL (eds) Cancer chemotherapy and biotherapy, 2nd edn. Lippincott-Raven, Philadelphia, pp 149–197

Heggie GD, Sommadossi JP, Cross DS, Huster WJ, Diasio RB (1987) Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. Cancer Res 47:2203–2206

Hormozian F, Schmitt JG, Sagulenko E, Schwab M, Savelyeva L (2007) *FRA1E* common fragile site breaks map within a 370 kilobase pair region and disrupt the dihydropyrimidine dehydrogenase gene (*DPYD*). Cancer Lett 246:82–91

Hsiao HH, Yang MY, Chang JG, Liu YC, Liu TC, Chang CS, Chen TP, Lin SF (2004) Dihydropyrimidine dehydrogenase pharmacogenetics in the Taiwanese population. Cancer Chemother Pharmacol 53:445–451

Johnson MR, Wang K, Diasio RB (2002) Profound dihydropyrimidine dehydrogenase deficiency resulting from a novel compound heterozygote genotype. Clin Cancer Res 8:768–774

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

Kouwaki M, Hamajima N, Sumi S, Nonaka M, Sasaki M, Dobashi K, Kidouchi K, Togari H, Wada Y (1998) Identification of novel mutations in the dihydropyrimidine dehydrogenase gene in a Japanese patient with 5-fluorouracil toxicity. Clin Cancer Res 4:2999–3004

Lu Z, Zhang R, Diasio RB (1993) Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. Cancer Res 53:5433–5438

Lu Z, Zhang R, Carpenter JT, Diasio RB (1998) Decreased dihydropyrimidine dehydrogenase activity in a population of patients with breast cancer: implication for 5-fluorouracil-based chemotherapy. Clin Cancer Res 4:325–329

Martz E (2002) Protein explorer: easy yet powerful macromolecular visualization. Trends Biochem Sci 27:107–109

Mattison LK, Johnson MR, Diasio RB (2002) A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. Pharmacogenetics 12:133–144

McLeod HL, Collie-Duguid ES, Vreken P, Johnson MR, Wei X, Sapone A, Diasio RB, Fernandez-Salguero P, van Kuilenberg AB, van Gennip AH, Gonzalez FJ (1998) Nomenclature for human DPYD alleles. Pharmacogenetics 8:455–459

Morel A, Boisdron-Celle M, Fey L, Soulie P, Craipeau MC, Traore S, Gamelin E (2006) Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. Mol Cancer Ther 5:2895–2904

Naguib FN, el Kouni MH, Cha S (1985) Enzymes of uracil catabolism in normal and neoplastic human tissues. Cancer Res 45:5405–5412

Nishiyama T, Ogura K, Okuda H, Suda K, Kato A, Watabe T (2000) Mechanism-based inactivation of human dihydropyrimidine dehydrogenase by (E)-5-(2-bromovinyl)uracil in the presence of NADPH. Mol Pharmacol 57:899–905

Ogura K, Ohnuma T, Minamide Y, Mizuno A, Nishiyama T, Nagashima S, Kanamaru M, Hiratsuka A, Watabe T, Uematsu T (2005) Dihydropyrimidine dehydrogenase activity in 150 healthy Japanese volunteers and identification of novel mutations. Clin Cancer Res 11:5104–5111

Ridge SA, Sludden J, Brown O, Robertson L, Wei X, Sapone A, Fernandez-Salguero PM, Gonzalez FJ, Vreken P, van Kuilenburg AB, van Gennip AH, McLeod HL (1998a) Dihydropyrimidine dehydrogenase pharmacogenetics in Caucasian subjects. Br J Clin Pharmacol 46:151–156

Ridge SA, Sludden J, Wei X, Sapone A, Brown O, Hardy S, Canney P, Fernandez-Salguero P, Gonzalez FJ, Cassidy J, McLeod HL (1998b) Dihydropyrimidine dehydrogenase pharmacogenetics in patients with colorectal cancer. Br J Cancer 77:497–500

Seck K, Riemer S, Kates R, Ullrich T, Lutz V, Harbeck N, Schmitt M, Kiechle M, Diasio R, Gross E (2005) Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in a cohort of Caucasian individuals. Clin Cancer Res 11:5886–5892

Shestopal SA, Johnson MR, Diasio RB (2000) Molecular cloning and characterization of the human dihydropyrimidine dehydrogenase promoter. Biochim Biophys Acta 1494:162–169

van Kuilenburg AB (2004) Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. Eur J Cancer 40:939–950

van Kuilenburg AB, Haasjes J, Richel DJ, Zoetekouw L, Van Lenthe H, De Abreu RA, Maring JG, Vreken P, van Gennip AH (2000) Clinical implications of dihydropyrimidine dehydrogenase (DPD) deficiency in patients with severe 5-fluorouracil-associated toxicity: identification of new mutations in the DPD gene. Clin Cancer Res 6:4705–4712

van Kuilenburg AB, Dobritzsch D, Meinsma R, Haasjes J, Waterham HR, Nowaczyk MJ, Maropoulos GD, Hein G, Kalhoff H, Kirk JM, Baaske H, Aukett A, Duley JA, Ward KP, Lindqvist Y, van Gennip AH (2002) Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. Biochem J 364:157–163

Vreken P, Van Kuilenburg AB, Meinsma R, van Gennip AH (1997) Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and expression of missense mutations C29R, R886H and R235W. Hum Genet 101:333–338

Wei X, Elizondo G, Sapone A, McLeod HL, Raunio H, Fernandez-Salguero P, Gonzalez FJ (1998) Characterization of the human dihydropyrimidine dehydrogenase gene. Genomics 51:391–400

Yamaguchi K, Arai Y, Kanda Y, Akagi K (2001) Germline mutation of dihydropyrimidine dehydrogenese gene among a Japanese population in relation to toxicity to 5-Fluorouracil. Jpn J Cancer Res 92:337–342

Zhang K, Qin Z, Chen T, Liu JS, Waterman MS, Sun F (2005) HapBlock: haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms. Bioinformatics 21:131–134

Zhu AX, Puchalski TA, Stanton VP Jr, Ryan DP, Clark JW, Nesbitt S, Charlat O, Kelly P, Kreconus E, Chabner BA, Supko JG (2004) Dihydropyrimidine dehydrogenase and thymidylate synthase polymorphisms and their association with 5-fluorouracil/leucovorin chemotherapy in colorectal cancer. Clin Colorectal Cancer 3:225–234

Int. J. Radiat. Biol., Vol. 83, No. 7, July 2007, pp. 433–442

informa
healthcare

Analysis of Common Deletion (CD) and a novel deletion of mitochondrial DNA induced by ionizing radiation

LU WANG[1], YOSHIKAZU KUWAHARA[1], LI LI[1], TAISUKE BABA[1], RYONG-WOON SHIN[2], YASUHITO OHKUBO[3], KOJI ONO[4], & MANABU FUKUMOTO[1]

[1]*Department of Pathology, Institute of Development, Aging and Cancer, Tohoku University, Sendai,* [2]*Department of Neurological Science, Graduate School of Medicine, Sendai,* [3]*Department of Radiopharmacology, Tohoku Pharmaceutical University, Sendai, and* [4]*Research Reactor Institute, Kyoto University, Kumatori, Osaka, Japan*

(Received 19 June 2006; revised 22 March 2007; accepted 28 March 2007)

Abstract

Purpose: In order to identify supportive evidence of radiation exposure to cells, we analyzed the relationship between exposure to ionizing radiation and the induction of deletions in mitochondrial DNA (mtDNA).

Materials and methods: Using human hepatoblastoma cell line, HepG2 and its derivatives, HepG2-A, -89 and -400, established after long term exposure to X-ray, mtDNA deletions were analyzed by polymerase chain reaction (PCR) and real-time PCR after cells were subjected to radiation and genotoxic treatments.

Results: Common Deletion (CD), the most extensively studied deletion of mtDNA, was induced within 24 h after exposure to 5 Gray (Gy) of X-rays and was associated with replication of mtDNA. CD became undetectable several days after the exposure due to the death of cells containing mitochondria within which CD had been induced. Furthermore, we found a novel mtDNA deletion that consisted of a 4934 base-pair deletion (4934del) between nucleotide position 8435 and 13,368. A lower dose of ionizing radiation was required to induce the 4934del than for CD and this was independent of the quality of radiation used and was not induced by treatments with hydrogen peroxide (H_2O_2) and other genotoxic reagents including bleomycin.

Conclusion: CD is induced by ionizing radiation, however, the amount of CD detected at a certain point in time after radiation exposure is dependent on the initial frequency of CD induced and the death rate of cells with mtDNA containing CD. The novel mtDNA deletion found in this study, therefore, will be used to determine whether cells were exposed to ionizing radiation.

Keywords: *Mitochondrial DNA, ionizing radiation, deletion, cell death*

Introduction

DNA is known to be one of the major targets for radiation DNA damage due to radiation exposure result in cell death, mutation and carcinogenesis. Radiation induces various types of genetic changes such as point mutations, small and large deletions and complex rearrangements. We have studied molecular genetic mechanisms of radiation carcinogenesis by analyzing Thorotrast-induced liver tumors which are caused by chronic exposure to ionizing radiation (Fukumoto et al. 2006). However, a characteristic spectrum of associated DNA mutations has not been determined for radiation-induced human cancers. In order to estimate the genetic risk resulting from radiation exposure, we need to know the differences between spontaneous and radiation-induced mutations in terms of their natures and mechanisms of induction. Although the most frequently induced mutations reported are multiloci deletions, the full spectrum of genetic changes induced by radiation remains unknown (Sankaranarayanan 1991, Brooks 2005). This may be partially due to the fact that most studies have focused on the effects of radiation on nuclear DNA.

In order to more accurately identify biological changes specifically induced by radiation, markers that are more specific than changes in nuclear DNA are required. Mitochondria are essential for respiration and oxidative energy production in aerobic cells

Correspondence: Manabu Fukumoto, Department of Pathology, Institute of Development, Aging and Cancer, Tohoku University, 4-1 Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan. Tel: +81 22 717 8507. Fax: +81 22 717 8512. E-mail: fukumoto@idac.tohoku.ac.jp

ISSN 0955-3002 print/ISSN 1362-3095 online
DOI: 10.1080/09553000701370878

and are involved in apoptosis and probably also in tumorigenesis (Petros et al. 2005). Human mitochondrial DNA (mtDNA) is a 16,569 base-pair (bp) double-strand circular DNA molecule which contains 37 genes, encoding 13 polypeptides for the mitochondrial electron transport chain, 2 ribosomal RNA and 22 transfer RNA for mitochondrial protein synthesis (Anderson et al. 1981). Somatic cells have an average of 100–500 mitochondria with 1–15 mtDNA molecules per mitochondrion (Satoh & Kuroiwa 1991). Mitochondrial DNA is more susceptible than nuclear DNA to structural damage due to the lack of protective histones and a limited capacity for DNA repair. Therefore, mtDNA has a 10-fold higher mutation rate than nuclear DNA (Brown et al. 1979). These factors suggest that mtDNA may be a more sensitive indicator of exposure to ionizing radiation than nuclear DNA. The most abundant large-scale deletion reported in mtDNA is called a 'Common Deletion' (CD). CD is a 4977-bp deletion specifically occurring from nucleotide position (nt) 8470 to nt13,446 in the mtDNA sequence and can be used as a marker of oxidative damage to mtDNA. CD is a very sensitive marker of mtDNA damage because the lesion is amplified during mtDNA replication and is easily detected and quantified by polymerase chain reaction (PCR) using primers adjacent to the CD site. It is assumed that mtDNA may be particularly susceptible to radiation damage, resulting in CD formation and may therefore be a good indicator for total DNA damage. Induction of CD has been reported 72 h after the exposure to X-ray. However, the number of deletions induced was not dose dependent. In addition, the relationship between the radiation doses required to induce CD and radiosensitivity of the cell was inconsistent (Kubota et al. 1997, Prithivirajsingh et al. 2004).

Ionizing radiation is known to induce genomic instability that is transmitted through many generations after irradiation via the progeny of surviving cells. Radiation-induced genomic instability, including chromosome aberrations, is commonly observed in various types of mammalian cells (Suzuki et al. 2003). These imply that the biological effects and cellular alterations induced by chronic exposure to ionizing radiation could be different from those merely resulting from an accumulation of acute exposures. Therefore, in order to understand the biological effects of radiation, we need to examine the temporal fluctuation of deleted mtDNA after exposure to radiation. We also thought that the high copy number of mtDNA per cell and the high mutation rate of mtDNA might be advantageously exploited for the detection of radiation-specific DNA alterations. In this study, using cells chronically exposed to X-ray, we performed time course observations to measure the quantity of mtDNA present and the number of CD induced after radiation exposure, and sought to identify novel mtDNA deletions specifically induced by radiation.

Materials and methods

Cell lines

Human hepatoblastoma cell line, HepG2 was obtained from the Cell Resource Center for Biomedical Research in our institute and established 3 sub-cell lines by exposing HepG2 cells to 0.5 Gy of X-rays twice a day, everyday, for more than 4 years: this comprised a total dose of more than 1600 Gy of X-rays. HepG2-A was treated only by X-ray, HepG2-400 was treated by 400 nM N-Methyl-N′-Nitro-N-Nitrosoguanidine (MNNG; Fluka Chemika Biochemika, Buchs, Switzerland) 24 h once before starting X-ray exposure and HepG2-89 was exposed once to 2 Gy of alpha particles by boron-neutron capture before continuous X-ray treatment. HepG2-89 was the most radioresistant followed by HepG2-A. HepG2-400 was found to have the same sensitivity as HepG2. All the cells were maintained in Roswell Park Memorial Institute (RPMI1640) medium (Sigma-Aldrich Inc., St Louis, MO, USA) supplemented with 5% fetal bovine serum (FBS, Gibco Invitrogen Corp., Carlsbad, CA, USA). Establishment and characteristics of the cell lines will be published elsewhere (Kuwahara et al. manuscript in preparation). Maintenance radiation was applied to long-term irradiated cells at a level of 0.5 Gy twice a day. This procedure was used for the present study and exposure experiments were performed 12 h after the last maintenance irradiation. For X-ray irradiation, we used a 150-KVp X-ray generator in our institute (Model MBR-1520R, Hitachi, Tokyo, Japan) with a total filtration of 0.5 mm aluminum plus 0.1 mm copper filter. The X-ray dose rate measured by a thimble ionization chamber (IC 17A, Far West Technology, Goleta, CA, USA) at the same position as the samples was about 1 Gy/min. HepG2 cells were maintained in culture without any treatment for the same period as other 3 sub-cell lines.

Irradiation, treatments with hydrogen peroxide (H_2O_2) and anti-cancer drugs

HepG2, HepG2-A, -89 and -400 cells were irradiated by X-ray at the doses indicated. Cells were harvested at 24, 48, 72 h, and 10, 20, 30 and 45 days after exposure to 5 Gy of X-rays for time course studies of mtDNA deletions. H_2O_2 treatment was performed according to the studies by Prithivirajsingh et al. (2004) and Wang et al.

(2002). Briefly, a suitable aliquot of 30% H_2O_2 (Wako Chemical, Tokyo, Japan) solution was freshly diluted into phosphate-buffered saline (PBS) immediately before the experiment. After treatment with 200 μM H_2O_2 for 1 h in RPMI1640 without FBS, cells were washed once with PBS. Cells were harvested at 24, 48, 72 h, and 10, 20, 30 and 45 days after treatments.

For the treatment with anticancer reagents, HepG2 cells were washed once with PBS and incubated in RPMI1640 without serum for 1 h with one of following chemicals:

- 5 μg/ml bleomycin (Nihon Kayaku Co., Ltd., Tokyo, Japan)
- 20 μg/ml etoposide (Nihon Kayaku Co., Ltd., Tokyo, Japan)
- 200 μg/ml fluorouracil (5-Fu, Kyowa Hakko, Co., Ltd., Tokyo, Japan)
- 10 μM vincristine (Shionogi Pharmaceutical Co., Ltd., Tokyo, Japan)

Bleomycin treatment was carried out according to the report of Paul et al. (2001). Before bleomycin treatment, cells were washed with PBS containing 1 mM Calcium Chloride ($CaCl_2$) and permiabilized with 12.5 μg/ml L-α-lysophosphatidylcholine (Sigma-Aldrich, St Louis, MO, USA) in PBS-1 mM $CaCl_2$ for 1 min on ice. After treatment with anticancer drugs, the medium was removed and cells were washed once with PBS. Cells were harvested at 24, 48 and 72 h after treatment. From treatment to cell harvest, cells were cultured in RPMI1640 without irradiation. All the post-treatment conditions were the same for both radiation and chemical treatments, unless otherwise specified. The medium was changed 24 h before cell harvest. As a control for all treatments, cells without any treatment were used. The control was subjected to a change of medium 24 h before cell harvest, as was used for the treatments. In order to examine mtDNA deletion after exposure to different qualities of radiation, cells were irradiated by alpha particles, neutrons and gamma rays, respectively. Boron neutron capture (BNC) was used for the exposure to alpha particles. A ^{10}B compound, Sodium mercaptoundeca-Hydrododecaborate (BSH; Katchem Ltd., Prague, Czech Republic) was suspended in physiological saline at a concentration of 50 parts per million (ppm) in the culture medium, the cells were exposed to neutron radiation at the Research Reactor Institute, Kyoto University (KURR). The average fluency of the thermal neutron source was 2.1×10^{12} n/cm^2 and the average flux was 2.3×10^9 n/cm^2/s at 5 MW. We determined the exposure period at the calculated dose of 5 Gy of alpha particles. For the neutron-exposed group, the cells were exposed to the neutron source for the same period as the BNC group corresponding to 0.35 Gy. For the gamma ray exposed group, the cells were irradiated at a dose of 5 Gy (0.34 Gy/min) with a ^{60}Co gamma ray source at KURR. Twenty-four hours after exposure cells were analyzed for mtDNA deletions. All the exposure experiments were carried out using 25 cm^2 flasks (Nunc, Roskilde, Denmark) in triplicate.

Mitochondrial staining

Mitochondria of live cells were stained by incubating with 500 nM MitoTracker Green FM (Molecular Probes Inc., Eugene, OR, USA) for 30 min at 37°C. Cells were fixed with 3.7% formaldehyde (Wako chemical, Tokyo, Japan) for 30 min at room temperature and washed three times with PBS. Images were captured by fluorescence microscopy (Keyence BZ-8000, Keyence Co., Osaka, Japan) and fluorescence intensity was calculated using Photoshop software 7.0.1.1 (Adobe Systems Inc., San Jose, CA, USA).

Flow cytometry

Twenty four hours after exposure to 5 Gy of X-rays, HepG2 cells were trypsinized to make a single cell suspension. Pre-apoptotic and dying cells were stained by an Annexin V-fluoresceinisothiocyanate (FITC) apoptosis detection kit (BD Biosciences Japan, Tokyo, Japan) according to the manufacture's protocol. Annexin V positive pre-apoptotic cells were combined with propidium iodide positive and dying cells were separated from surviving/viable cells in a BD Fluorescence Activated Cell Sorting (FACS) Vantage (Becton, Dickinson Co., Franklin Lakes, NJ, USA).

Primers and PCR

Primers were synthesized by Gene Design Inc., Osaka, Japan. To detect CD, primer F1 and R1 sites were chosen at the region flanking the 4977-bp CD region and PCR was carried out under conditions that allowed a product to form only if the deletion had occurred. For the assessment of mtDNA copy number, a primer set to amplify the mitochondrial cytochrome *b* gene, where mutations rarely exist, was designed (F: 5′-tatccgccatcccatacatt-3′, R: 5′-ggtgattcctaggggttgt-3′). Primers for amplifying the nuclear β-actin gene (forward primer (F): 5′-ttctacaatgagctgcgtgtgg-3′, reverse primer (R): 5′-tcctacggaaaacggcagaaga-3′) were used to normalize the copy number of PCR products from mtDNA in each sample relative to the copy number of nuclear DNA. We designed the primers Fnd and Rnd to amplify the novel deletion that we found in this

study. The sequence of mtDNA and the annealing site for the primers, that is, F1, R1, Fnd and Rnd are shown in Figure 1.

Total DNA was extracted from cells using a DNeasy Tissue Kit (Qiagen, Valencia, CA, USA). Real-time PCR analysis was performed in a BIO-RAD icycler iQ, SA-THK Real-time PCR analysis system (BIO-RAD, Hercules, CA, USA). Amplification was conducted in a final volume of 15 μl containing 10 ng DNA, 0.15 μM of each primer, and AB solute syber green ROX mix (ABgene, Epsom, Surrey, UK). The PCR program included initial denaturation at 95°C for 15 min followed by 40 amplification cycles consisting of denaturation at 94°C for 30 sec, annealing at a suitable temperature (61°C for CD, 66°C for β-actin and 58°C for cytchrome *b*) for 30 sec, and extension at 72°C for 30 sec. Ordinary PCR was performed in a GeneAmp PCR system 9600 (Perkin-Elmer Corp., Norwalk, CT, USA). The reactions were carried out in a 10 μl reaction volume containing 10 ng DNA, 0.25 U platinum™ *Taq* DNA polymerase High Fidelity (Invitrogen Corp., Carlsbad, CA, USA), 1x High Fidelity PCR Buffer (Invitrogen Corp., Carlsbad, CA, USA), 2 mM $MgSO_4$, 0.2 mM of each dNTP (Roche Applied Science, Penzberg, Germany) and 0.2 μM primers. Amplification was performed under the same conditions as for real-time PCR, but with 35 amplification cycles. PCR products were electrophoresed on a 10% polyacrylamide gel in Tris borate ethylenediaminetetraacetic acid (EDTA) (Wako chemical, Tokyo, Japan) buffer, and stained with ethidium bromide (Wako chemical, Tokyo, Japan) or SYBR Green Nucleic Acid Gel Stain Starter Kit (Molecular Probes Inc., Eugene, OR, USA). All PCR analyses were performed in triplicate.

Cloning and sequencing

After PCR products were electrophoresed in a 1% agarose gel, strips of agarose corresponding to DNA bands were cut out and purified using the Wizard SV Gel and PCR Clean-Up System (Promega Corp., Madison, WI, USA). Sequencing was performed in an ABI Prism 310 (Perkin-Elmer, Foster City, CA, USA) according to the manufacture's protocol after subcloning into TOPO vectors (TOPO TA Cloning Kit, Invitrogen Corp., Carlsbad CA, USA).

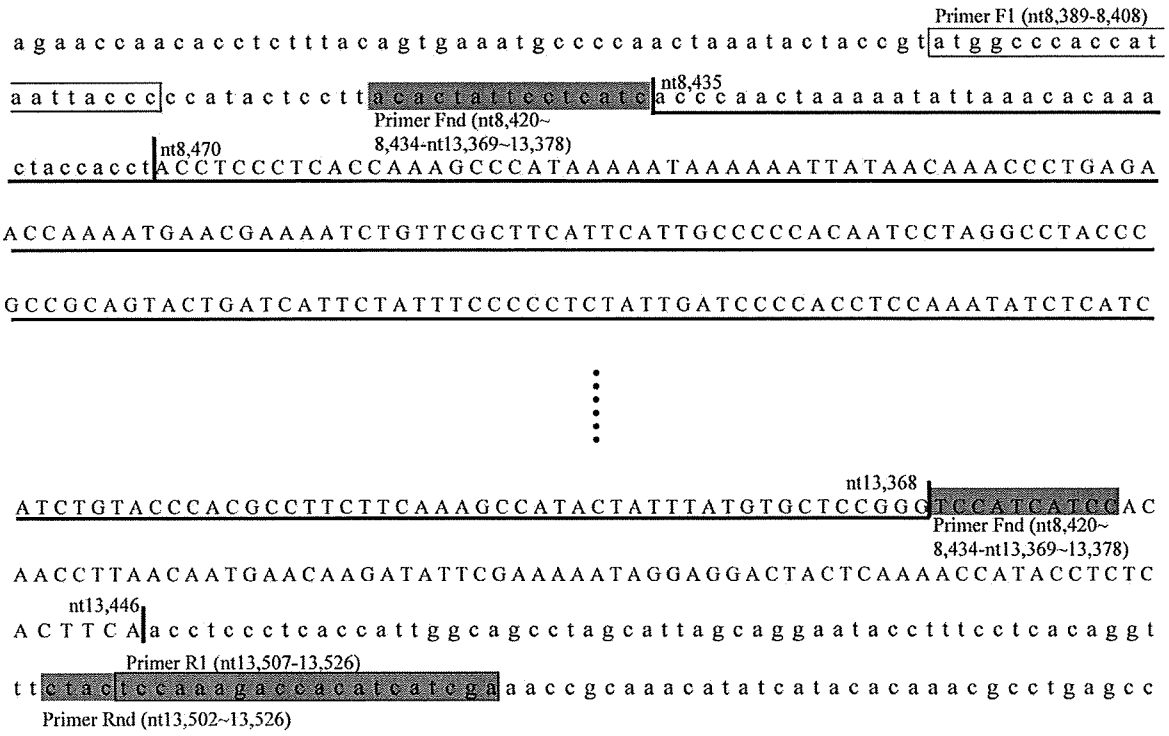

Figure 1. Mitochondrial DNA (mtDNA) sequence around common deletion (CD). Capital letters indicate nucleotides within CD. Solid underlines (__) indicate a novel deletion (4934del) specific to the radiation exposure found in this study. Vertical lines (|) indicate the boundaries of deletions. DNA sequences where PCR primers anneal are enclosed with squares (□) or hatched (■). Primer set F1-R1 was used to detect CD. Primer set Fnd-Rnd was used for detection of the novel deletion found in this study. The 'nt' number indicates nucleotide position of mtDNA. CD associated 13-nucleotide direct repeats are shown in bold letters.