Association between genetic polymorphisms of the base excision repair gene *MUTYH* and increased colorectal cancer risk in a Japanese population

Hong Tao,[1,14] Kazuya Shinmura,[1] Masaya Suzuki,[1] Suminori Kono,[2] Ryuichi Mibu,[3] Masao Tanaka,[3] Yoshihiro Kakeji,[4] Yoshihiko Maehara,[4] Takeshi Okamura,[5] Kouji Ikejiri,[6] Kitaroh Futami,[7] Youichi Yasunami,[8] Takafumi Maekawa,[9] Kenji Takenaka,[10] Hitoshi Ichimiya,[11] Nobutoshi Imaizumi[12] and Haruhiko Sugimura[1,13]

[1]First Department of Pathology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192; Departments of [2]Preventive Medicine, [3]Surgery and Oncology, and [4]Surgery and Science, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582; [5]Department of Gastroenterological Surgery, National Kyushu Cancer Center, 3-1-1 Notame, Minami-ku, Fukuoka 811-1395; [6]Division of Surgery, National Kyushu Medical Center, 1-8-1 Jigyohama, Chuo-ku, Fukuoka 810-8563; [7]Department of Surgery, Fukuoka University Chikushi Hospital, 377-1 Oaza-zokumyoin, Chikushino-shi 818-0067; [8]The First and [9]Second Departments of Surgery, Fukuoka University School of Medicine, 4-45-1 Nanakuma, Jonan-ku, Fukuoka 814-0180; [10]Division of Surgery, Fukuoka City Hospital, 13-1 Yoshizuka-honmachi, Hataka-ku, Fukuoka 812-0046; [11]Division of Surgery, Hamanomachi General Hospital, 3-5-27 Maizuru, Chuo-ku, Fukuoka 810-8539; [12]Division of Surgery, Fukuoka Red Cross Hospital, 3-1-1 Ogusu, Minami-ku, Fukuoka 815-8555, Japan

(Received May 28, 2007/Revised October6/October 17, 2007/Accepted October 24, 2007/Online publication February 4, 2008)

The *MUTYH* gene encodes a DNA glycosylase that can initiate the base excision repair pathway and prevent G:C > T:A transversion by excising adenine mispaired with 8-hydroxyguanine. Biallelic germline mutations of *MUTYH* have been shown to predict familial and sporadic multiple colorectal adenomas and carcinomas, however, whether there is an association between single nucleotide polymorphisms (SNPs) of *MUTYH* and sporadic colorectal cancer (CRC) risk has remained unclear. In this study we investigated four *MUTYH* SNPs, IVS1+11C > T, IVS6+35G > A, IVS10–2A > G, and 972G > C (Gln324His), for an association with increased CRC risk in a population-based series of 685 CRC patients and 778 control subjects from Kyushu, Japan. A statistically significant association was demonstrated between IVS1+11T and increased CRC risk (odds ratio [OR]: 1.43; 95% confidence interval [CI]: 1.012–2.030; *P* = 0.042) and one of the five haplotypes based on the four SNPs, the IVS1+11T – IVS6+35G – IVS10–2A – 972C (TGAC) haplotype containing IVS1+11T, was demonstrated to be associated with increased CRC risk (OR, 1.43; 95% CI, 1.005–2.029; *P* = 0.046). Subsite-specific analysis showed that the TGAC haplotype was statistically significantly (*P* = 0.013) associated with an increased risk of distal colon, but not proximal colon or rectal cancer. Furthermore, IVS1+11C > T was found to be in complete linkage disequilibrium with –280G > A and 1389G > C (Thr463Thr). The results indicated that Japanese individuals with –280A/IVS1+11T/1389C genotypes or the TGAC haplotype are susceptible to CRC. (*Cancer Sci* 2008; 99: 355–360)

Intracellular DNA is at risk of damage by reactive oxygen species (ROS) generated by normal metabolism and environmental exposure, and 8-hydroxyguanine (8-ohG) is one of the products induced by ROS damage and is known to be a mutagenic lesion.[1,2] The base excision repair (BER) pathway plays an important role in repairing oxidative-damage-induced mutations, and the *MUTYH* gene encodes the glycosylase capable of initiating the BER pathway by catalyzing the removal of adenine residues mispaired with 8-ohG.[3–5] It has been indicated that defects in the BER pathway may contribute to tumorigenesis by increasing mutation frequency in oncogenes and tumor suppressor genes.[6] In fact, it has been reported that some cases of autosomal recessive inherited multiple colorectal adenomatous polyposis and carcinoma with an increased frequency of somatic G:C > T:A mutations in *APC* are attributable to biallelic germline mutations in the *MUTYH* gene.[7–10] The disease-causing mutations, Y165C, G382D, 466delE, E466X, and Y90X have been reported in Caucasians, Indian, Pakistani and other ethnic groups.[8,9,11,12] The frequencies of Y165C and G382D have been investigated in several colorectal cancer (CRC) case-control studies, and monoallelic carriers of these variants were found in 0.0–2.6% of the cases and 0.0–2.1% of the controls and biallelic carriers of these variants were found in 0.0–0.8% of the cases and 0% of the controls, respectively.[13–18] However, neither of these two variants has ever been detected in East Asians, including Japanese,[19–22] suggesting that they are ethnicity-specific alleles. Based on the above findings, we hypothesized that *MUTYH* variants other than Y165C and G382D act as low-penetrance susceptibility alleles in Japanese CRC, similar to a situation previously reported for the *APC* and *CHEK2* gene variants.[23,24]

We conducted a CRC case-control study to evaluate the significance of *MUTYH* variants in a Japanese population. In the single-nucleotide polymorphisms (SNPs) reported in the Japanese population,[19,20] four SNPs (IVS1+11C > T, IVS6+35G > A, IVS10–2A > G and 972G > C [Gln324His]), were selected, and all 685 cases and 778 matched controls were genotyped to detect these four SNPs. Statistically significant association was found between the IVS1+11C > T SNP and increased CRC risk in the Japanese population. A haplotye-based association study was also carried out, and a statistically significant association was found between the IVS1+11T – IVS6+35G – IVS10–2A – 972C (TGAC) haplotype containing the IVS1+11T allele and CRC risk. In the subsite-specific analysis, the IVS1+11C > T SNP was detected to be nearly statistically significantly associated and the TGAC haplotype was found to be statistically significantly associated with an increased risk of distal colon, but not proximal colon or rectal cancer. We also found that a novel –280G > A SNP in the 5′ flanking region of *MUTYH* and a previously reported 1389G > C (Thr463Thr) SNP were both in complete linkage disequilibrium with the IVS1+11C > T. Our results suggest that the –280A/IVS1+11T/1389C or the TGAC haplotype of *MUTYH* may be novel CRC susceptibility alleles.

Materials and Methods

Specimens. Blood specimens from 685 CRC cases and 778 controls were collected in a previous study. DNA was extracted from these specimens and written informed consent was obtained

[13]To whom correspondence should be addressed.
E-mail: hsugimur@hama-med.ac.jp
[14]Hong Tao is a scholarship student (2006–08) of the Rotary Yoneyama Scholarship endowed by Rotary International, Japan.

doi: 10.1111/j.1349-7006.2007.00694.x

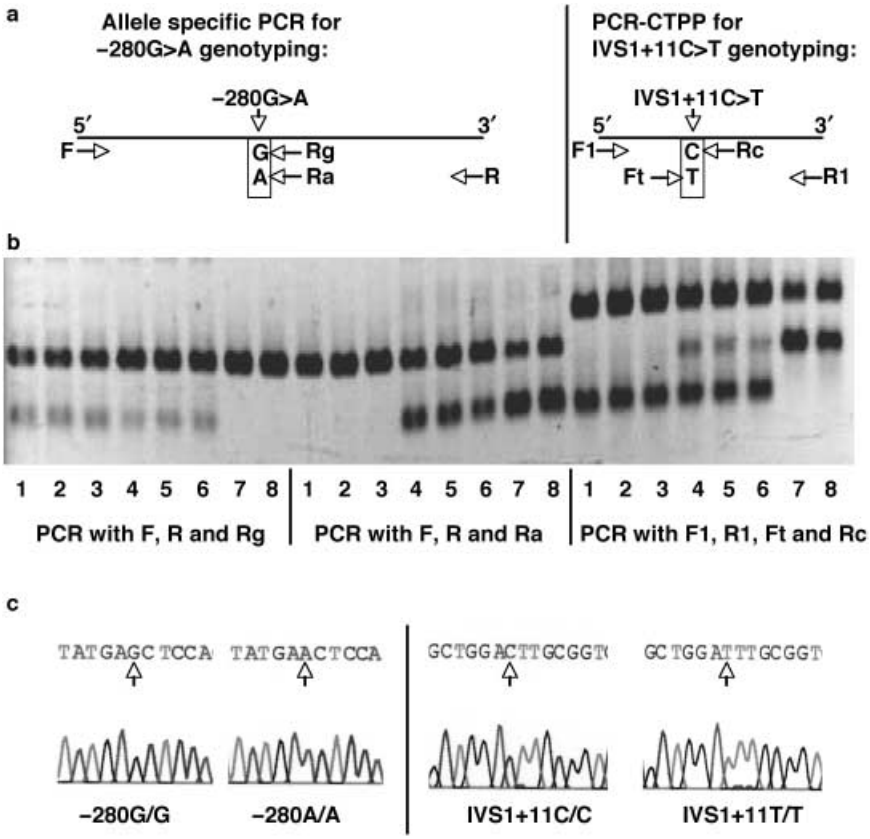

Fig. 1. Genotyping of the –280G > A and IVS1+11C > T single nucleotide polymorphisms (SNPs) of the *MUTYH* gene. (a) The schematic diagrams of the allele-specific polymerase chain reaction (PCR) used to genotype the –280G > A SNP (left) and the PCR with confronting two-pair primers (PCR-CTPP) used to genotype the IVS1+11C > T SNP (right). PCR primers are indicated by the horizontal arrows, and F and R mean forward primer and reverse primer, respectively. The location of each SNP is indicated by a vertical arrow. (b) Agarose gel electrophoresis of the PCR products. Eight samples, three from homozygous carriers of the wild-type allele (No. 1–3), three from heterozygous (No. 4–6) and two from homozygous (No. 7 and 8) carriers of the variation, were genotyped for –280G > A (left and middle) and IVS1+11C > T (right). (c) Sequence electropherograms of the region containing the –280G/G and A/A (left two) and IVS1+11C/C and T/T (right two). The positions of the SNPs are indicated by vertical arrows.

from each individual patient.[25] The characteristics of the cases and controls have been described previously.[25–28] In brief, the cases were composed of a consecutive series of patients with histologically-confirmed incident colorectal adenocarcinomas, and controls were composed of individuals that had no diagnosis of CRC. Other eligibility criteria were as follows: age 20–74 years at the time of diagnosis for the cases or at the time of selection for the controls, residents of the study area (Fukuoka City and three adjacent areas), no prior history of partial or total removal of the colorectum, familial adenomatous polyposis or inflammatory bowel disease, and mental competence to give informed consent and participate in the interview. The number of control candidates by gender and 10-year age class was determined in accordance to the expected sex- and age-specific number of incident cases of colorectal cancer. For the reverse transcriptase-polymerase chain reaction (RT-PCR) experiment, total RNA was extracted from the non-cancerous colorectal mucosa of six CRC patients and converted to cDNA, as described previously.[19] This study was approved by the Institutional Review Board (IRB) of Hamamatsu University School of Medicine (12–14, 18-4).

Target SNPs and genotyping. The six SNPs genotyped in this study were as follows: IVS1+11C > T (rs2275602), IVS6+35G > A (rs3219487), IVS10–2A > G (5′-flanking sequence: 5′-CAC TCA ACC CTG TGC CTC TC-3′; 3′-flanking sequence: 5′-GGT GGA GCA GGA ACA GCT CT-3′), 972G > C (Gln324His) (rs3219489), G382D (rs36053993), and –280G > A (5′-flanking sequence: 5′-ATT ACT ACT AAC CGT TAT GA-3′; 3′-flanking sequence: 5′-CTC CAG ACT ACA TCT CCC GC-3′). The IVS10–2A > G and –280G > A had not been presented in the SNP database (dbSNP) of the National Center for Biotechnology Information (NCBI) Entrez system. Genotyping of the four target SNPs, namely, IVS1+11C > T, IVS6+35G > A, IVS10–2A > G and 972G > C (Gln324His), was carried out by PCR with confronting two-pair primers (PCR-CTPP), as described previously (Fig. 1a,b),[19] and genotyping of the G382D SNP was carried out by PCR-restriction fragment length polymorphism (PCR-RFLP). Genotyping of the –280G > A SNP was carried out by two independent allelic-specific PCRs (Fig. 1a,b). The PCR primers used were: IVS1+11C > T SNP: F1 (5′-AAC TAT GAG CCC GAG GCC TTC C-3′), R1 (5′-CAG CAG AAC ACG GAG GCC C-3′), F2 (5′-AGT CGT CTG TGG GTA CGC TGG AT-3′), and R2 (5′-CCA GGA GAC GGA CCG CAA G-3′); IVS6+35G > A: F1 (5′-CCA GTG TGG GTC TCA GAG G-3′), R1 (5′-CCC TAG CTC CTC TAC CAC CTG-3′), F2 (5′-CTA GGG TAG GGG AAA TAG GAA CA-3′), and R2 (5′-CAC CCG TCA GTC CCT CTA TC-3′); IVS10–2A > G SNP, those described previously,[19] 972G > C (Gln324His) SNP: F1 (5′-CCT GTC GGG CAG TCC TGA CG-3′), R1 (5′-CGC TGA AGC TGC TCT GAG GGC-3′), F2 (5′-CCC AGC TCC CAA CAC TGG ACA C-3′), and R2 (5′-GAG GCA GGC ACA GGT GGC AC-3′); G382D SNP: F (5′-GCC CAA ATT CTG CTG GTG C-3′) and R (5′-GCC CAA CGC TGT AGT TCC TG-3′); –280G > A SNP, F (5′-TAC TGT TCT CAT GGT GCC CC-3′), R (5′-GCC TCG GGC TCA TAG TTC TAG-3′), Ra (5′-GCG GGA GAT GTA GTC TGG AGT-3′), and Rg (5′-CGG GAG ATG TAG TCT GGA GC-3′). PCR products were fractionated by electrophoresis on a 2.0% agarose gel and stained with ethidium bromide. All the cases and controls were genotyped for all of the above SNPs.

Statistical analysis. χ^2 tests were used for deviation from the Hardy–Weinberg equilibrium (HWE) among the controls, and the significance level was set at 0.05. Associations between *MUTYH* genotypes or haplotypes and risk of CRC were assessed by calculating the odds ratio (OR) and 95% confidence interval (CI). SAS version 8.2 software (SAS institute, Inc., Cary, NC, USA) was used to carry out the statistical analysis. A *P*-value less than 0.05 was accepted as statistically significant in all cases. Adjustment for multiple testing was performed using false discovery rate (FDR) principle.[29] Haplotypes were inferred by

doi: 10.1111/j.1349-7006.2007.00694.x

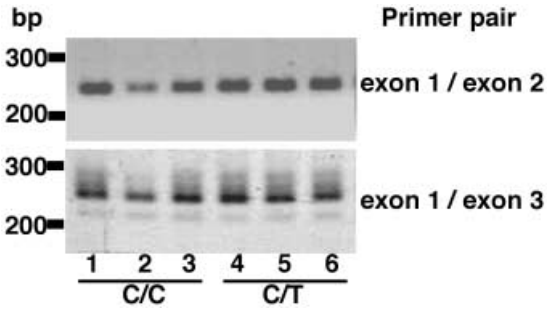

Fig. 2. Reverse transcription-polymerase chain reaction (RT-PCR) analysis. RT-PCR was carried out with a set of primers located at exon 1 and exon 2 (upper panel) and a set of primers located at exon 1 and exon 3 (lower panel). cDNAs from three homozygous carriers of the wild-type allele (lanes 1–3) and three heterozygous carriers of the variation (lanes 4–6) were used as the templates.

the expectation maximization algorithm with the SNPAlyze Version 5.0 software (DYNACOM, Yokohama, Japan). Five haplotypes with a frequency of greater than 1% were selected for further statistical analysis. The linkage disequilibrium analysis of the haplotypes was carried out using the SNPAlyze Version 5.0 software (DYNACOM).

RT-PCR analysis. Reverse transcription-polymerase chain reaction was carried out for the IVS1+11C/C and C/T genotype, respectively, with 1 μL of each cDNA prepared from the non-cancerous colorectal mucosa of six CRC patients. The two primer pairs shown in Fig. 2 were used: one pair was composed of a forward primer in exon 1 and a reverse primer in exon 2, and the other pair was composed of the same forward primer in exon 1 and a reverse primer in exon 3. The sequences of the primer sets are available on request.

Results

Target SNP selection. Among the variants registered in the dbSNPs of the NCBI Entrez system, there were six *MUTYH* SNPs, namely, IVS1+11C > T (rs2275602), IVS1+1841G > A (rs3219472), IVS1+3221T > G (rs3219476), IVS6+35G > A (rs3219487), IVS14−40G > C (rs3219493) and 972G > C (Gln324His) (rs3219489), that have been detected in the Japanese population. Among the SNPs reported in previous publications,[19,20] five *MUTYH* SNPs, namely, IVS1+11C > T (rs2275602), IVS6+35G > A (rs3219487), IVS10−2A > G (5′-flanking sequence: 5′-CAC TCA ACC CTG TGC CTC TC-3′; 3′-flanking sequence: 5′-GGT GGA GCA GGA ACA GCT CT-3′), 972G > C (Gln324His) (rs3219489), and 1389G > C (Thr463Thr) (5′-flanking sequence: 5′-CCA GGT GCT CGC TGG CTG AC-3′; 3′-flanking sequence: 5′-CAG GAG GAA TTT CAC ACC GC-3′) have been reported in the Japanese population. However, since the IVS6+35G > A and IVS1+11C > T had been found to be in complete linkage disequilibrium with IVS14−40G > C and 1389G > C (Thr463Thr), respectively. The remaining six SNPs were initially selected as candidates. For the haplotype association analysis, a pilot study was carried out by genotyping the six SNPs in 30 healthy Japanese individuals. Analysis with the SNPAlyze Version 5.0 software revealed the five haplotypes with a frequency of more than 1%, and they comprised all of the total predicted haplotype variation. As we were able to distinguish these five haplotypes with four SNPs, namely, IVS1+11C > T, IVS6+35G > A, IVS10−2A > G, and 972G > C (Gln324His), the 4 SNPs were ultimately chosen as the haplotype-tagging SNPs.

Association between the IVS1+11C > T SNP and increased risk of CRC. The 685 cases and 778 controls were genotyped for the *MUTYH* SNPs by PCR-CTPP, and the accuracy of the genotyping was verified by sequencing five specimens for each genotype of each SNP. The concordance rate was 100% (data not shown). The frequencies of each SNP are summarized in Table 1. The genotypic distributions of all the SNPs detected were in HWE. The IVS1+11C > T SNP, whose functional role has never been investigated, was shown to be statistically significantly associated with increased CRC risk. The crude OR was 1.43 (95% CI, 1.012–2.030; $P = 0.042$). After adjustments for gender, age and place of residence, the OR was estimated to be 1.46 (95% CI, 1.024–2.069; $P = 0.036$) (Table 1). The *P*-value remained less than 0.05 after FDR adjustment (Table 1). No statistically significant differences in the frequency of any of the other three SNPs, IVS6+35G > A, IVS10−2A > G, and 972G > C, were observed between the cases and controls (Table 1). Furthermore, the association between the SNPs of *MUTYH* and the risk of CRC was examined by the anatomic subsite of the CRC. It showed that the IVS1+11 A/T + T/T genotypes were nearly statistically significantly associated with an increased risk of distal colon cancer risk (OR, 1.58; 95% CI, 0.984–2.544; $P = 0.058$) (Table 2). Since monoallelic mutation of G382D has recently been shown to be associated with CRC risk in Caucasians,[16] the 685 cases and 778 control subjects were also examined for G382D, but no homozygotes or heterozygotes for this mutation were detected (data not shown). Because the complete linkage disequilibrium between IVS1+11C > T and 1389G > C had already been reported,[19] the results suggested that the IVS1+11C > T and 1389G > C variants of *MUTYH* may confer susceptibility to CRC in the Japanese population.

Table 1. Genotypes of the four *MUTYH* single nucleotide polymorphisms (SNPs) and risk of colorectal cancer

Variation†	Genotype	No. of controls (%)/cases (%)	Not adjusted		Adjusted‡		
			OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	FDR adjusted *P*-value§
IVS1+11C > T	C/C	714 (91.8)/607 (88.6)	1.00 (reference)	–	1.00 (reference)	–	–
	C/T + T/T	64 (8.2)/78 (11.4)	1.43 (1.012–2.030)	0.042	1.46 (1.024–2.069)	0.036	0.036
IVS6+35G > A	G/G	628 (80.7)/539 (78.7)	1.00 (reference)	–	1.00 (reference)	–	–
	G/A	143 (18.4)/140 (20.4)	1.14 (0.880–1.480)	0.321	1.14 (0.878–1.485)	0.321	0.963
	A/A	7 (0.9)/6 (0.9)	1.00 (0.334–2.990)	0.998	0.97 (0.320–2.926)	0.953	>1.0
IVS10−2A > G	A/A	741 (95.2)/662 (96.6)	1.00 (reference)	–	1.00 (reference)	–	–
	A/G + G/G	37 (4.8)/23 (3.4)	0.70 (0.409–1.183)	0.178	0.67 (0.390–1.139)	0.138	0.276
972G > C (Gln324His)	G/G	215 (27.6)/194 (28.3)	1.00 (reference)	–	1.00 (reference)	–	–
	G/C	395 (50.8)/350 (51.1)	0.98 (0.771–1.250)	0.883	0.96 (0.751–1.223)	0.733	>1.0
	C/C	168 (21.6)/141 (20.6)	0.93 (0.692–1.251)	0.632	0.90 (0.670–1.220)	0.511	>1.0

†Nucleotide +1 is the A of the ATG-translation initiation codon. ‡Adjustment was made for gender, 5-year age class, and residential area. §False discovery rate (FDR) adjusted *P*-value. CI, confidence interval; OR, odds ratio.

Table 2. *MUTYH* genotypes and the risk of colorectal cancer (CRC) stratified by anatomic subsite

Variation†	Genotype	Proximal colon‡ ($n = 150$)			Distal colon‡ ($n = 232$)			Rectum‡ ($n = 290$)		
		No.	OR (95% CI)	*P*-value	No.	OR (95% CI)	*P*-value	No.	OR (95% CI)	*P*-value
IVS1+11C > T	C/C	133	1.00 (reference)	–	203	1.00 (reference)	–	259	1.00 (reference)	–
	C/T + T/T	17	1.50 (0.843–2.664)	0.169	29	1.58 (0.984–2.544)	0.058	31	1.36 (0.861–2.154)	0.187
IVS6+35G > A	G/G	122	1.00 (reference)	–	178	1.00 (reference)	–	229	1.00 (reference)	–
	G/A	28	1.03 (0.655–1.629)	0.887	51	1.27 (0.884–1.835)	0.195	58	1.07 (0.757–1.510)	0.703
	A/A	0	–	0.984	3	1.52 (0.383–6.056)	0.551	3	1.15 (0.286–4.581)	0.848
IVS10–2A > G	A/A	143	1.00 (reference)	–	226	1.00 (reference)	–	281	1.00 (reference)	–
	A/G + G/G	7	0.98 (0.422–2.268)	0.959	6	0.48 (0.197–1.156)	0.101	9	0.62 (0.295–1.319)	0.216
972G > C (Gln324His)	G/G	34	1.00 (reference)	–	72	1.00 (reference)	–	83	1.00 (reference)	–
	G/C	95	1.57 (1.020–2.426)	0.040	105	0.78 (0.553–1.108)	0.167	144	0.90 (0.655–1.247)	0.538
	C/C	21	0.78 (0.434–1.405)	0.409	55	0.99 (0.654–1.484)	0.943	63	0.94 (0.639–1.392)	0.768

†Nucleotide +1 is the A of the ATG-translation initiation codon. ‡Adjustment was made for gender, 5-year age class, and residential area. CI, confidence interval; OR, odds ratio.

Table 3. Haplotype frequency based on the four *MUTYH* single nucleotide polymorphisms (SNPs) and risk of colorectal cancer

Haplotype†				Frequency (%)‡		Not adjusted		Adjusted§		
IVS1+11C > T	IVS6+35G > A	IVS10–2A > G	972G > C	Control	Case	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	FDR adjusted *P*-value¶
C	G	A	C	42.9	40.4	1.00 (reference)	–	1.00 (reference)	–	–
C	G	A	G	40.9	41.5	0.93 (0.791–1.089)	0.360	1.10 (0.936–1.293)	0.248	0.992
C	A	A	G	9.8	10.7	1.08 (0.840–1.397)	0.537	1.18 (0.910–1.519)	0.215	0.645
T	G	A	C	4.0	5.8	1.43 (1.005–2.029)	0.046	1.56 (1.098–2.228)	0.013	0.013
C	G	G	G	2.5	1.6	0.63 (0.370–1.079)	0.090	0.66 (0.387–1.136)	0.135	0.270

†Nucleotide +1 is the A of the ATG-translation initiation codon. ‡Inferred common haplotypes with frequency >1% are listed. §Adjustment was made for gender, 5-year age class, and residential area. ¶False discovery rate (FDR) adjusted *P*-values. CI, confidence interval; OR, odds ratio.

Table 4. *MUTYH* haplotypes and the risk of colorectal cancer (CRC) stratified by anatomic subsite

Haplotype†				Proximal colon‡		Distal colon‡		Rectum‡	
IVS1+11C > T	IVS6+35G > A	IVS10–2A > G	972G > C	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
C	G	A	C	1.00 (reference)	–	1.00 (reference)	–	1.00 (reference)	–
C	G	A	G	1.13 (0.855–1.484)	0.397	1.10 (0.871–1.389)	0.424	1.06 (0.856–1.311)	0.595
C	A	A	G	1.00 (0.634–1.587)	0.990	1.27 (0.885–1.816)	0.196	1.15 (0.822–1.594)	0.424
T	G	A	C	1.61 (0.904–2.879)	0.106	1.81 (1.131–2.884)	0.013	1.36 (0.854–2.166)	0.195
C	G	G	G	0.99 (0.430–2.296)	0.988	0.43 (0.166–1.116)	0.083	0.62 (0.296–1.316)	0.216

†Nucleotide +1 is the A of the ATG-translation initiation codon. ‡Adjustment was made for gender, 5-year age class, and residential area. CI, confidence interval; OR, odds ratio.

Association between the TGAC haplotype containing the IVS1+11T and increased risk of CRC. Haplotype-based association studies are known to have greater power than individual SNP-based association studies.[30] Haplotype analyses were carried out based on the genotyping data of the four SNPs of *MUTYH*, namely, IVS1+11C > T, IVS6+35G > A, IVS10–2A > G, and 972G > C. There were five haplotypes with a frequency greater than 1%, CGAC, CGAG, CAAG, TGAC, and CGGG (Table 3), and since the CGAC haplotype was detected in 42.7% of the controls (Table 3), the highest percentage, the CGAC haplotype was used as the reference haplotype, and the following statistical analysis was carried out using the SAS system. Consistent with the results for each SNP, the TGAC haplotype containing the IVS1+11T allele was statistically significantly associated with increased CRC risk. The crude OR was 1.43 (95% CI, 1.005–2.029; $P = 0.046$) and after adjustment for gender, age and place of residence, the OR was 1.56 (95% CI, 1.098–2.228; $P = 0.013$) (Table 3). The *P*-value remained less than 0.05 after FDR adjustment (Table 3). The results of subsite-specific analysis revealed a significant association between the TGAC haplotype and increased risk of distal colon cancer (OR, 1.81; 95% CI, 1.131–2.884; $P = 0.013$) (Table 4). These results suggested that the TGAC haplotype containing the c.36+11T SNP confers susceptibility to CRC, especially to distal colon cancer, in the Japanese population.

RT-PCR analysis and detection of a novel SNP –280G > A linked with IVS1+11C > T. The association between the IVS1+11C > T SNP of *MUTYH* and CRC risk suggested a functional difference between IVS1+11C and IVS1+11T. The IVS1+11C > T SNP is located in the boundary region between *MUTYH* exon 1 and intron 1, and many reports have suggested that gene variants in the neighborhood of the junction are often accompanied by abnormal splicing.[31–33] In order to investigate whether the IVS1+11C > T SNP affects the splicing of *MUTYH*, an RT-PCR analysis was carried out by using cDNAs from carriers of the IVS1+11C/T and C/C genotype. However, no splicing abnormalities were detected in the cases carrying the IVS1+11C/T genotype (Fig. 2). On the other hand, during our checking of the sequences around the first exon of *MUTYH* to exclude variation at the splice site, a novel SNP of –280G > A was detected in the sample carrying the IVS1+11C > T SNP (Fig. 1b,c). Further genotyping was carried out in all subjects,

doi: 10.1111/j.1349-7006.2007.00694.x

and –280G > A was found to be 100% linked with IVS1+11C > T. The –280G > A SNP was demonstrated to be in complete linkage disequilibrium with the IVS1+11C > T SNP ($r^2 = 1$) by the SNPAlyze Version 5.0 software. Since the –280G > A SNP and 1389G > C (Thr463Thr) SNP were both in complete linkage disequilibrium with the IVS1+11C > T SNP, the results suggested that the Japanese individuals with the –280 A/IVS1+11T/1389C alleles or the –280A – IVS1+11T – IVS6+35G – IVS10-2A – 972C – 1389C haplotype were significantly associated with increased CRC risk.

Discussion

In this Japanese population-based case-control study four *MUTYH* SNPs, namely, IVS1+11C > T, IVS6+35G > A, IVS10–2A > G, and 972G > C, were genotyped in 685 CRC cases and 778 controls. The frequency distribution of IVS1+11T and the IVS1+11T – IVS6+35G – IVS10-2A – 972C (TGAC) haplotype were significantly associated with increased CRC risk. Subsite-specific analysis showed that the IVS1+11C > T SNP was nearly statistically significantly associated ($P = 0.058$) and the TGAC haplotype were statistically significantly associated ($P = 0.013$) with an increased risk of distal colon, but not proximal colon or rectal cancer. Next, we found that the IVS1+11C > T SNP was in complete linkage disequilibrium with –280G > A. No aberrant splicing induced by IVS1+11T allele was detected by RT-PCR. Together with the previously detected 1389G > C (Thr463Thr), a SNP in complete linkage disequilibrium with IVS1+11C > T, our results suggested that individuals who have the *MUTYH* –280 A/IVS1+11T/1389C alleles or the TGAC haplotype are more susceptible to CRC in the Japanese population.

The present study is the first Japanese population-based CRC case-control study to evaluate the association between the SNPs of *MUTYH* and the risk of CRC by the anatomic subsite of the colorectal cancer. The results demonstrate that the IVS1+11C > T SNP and TGAC haplotype confer susceptibility to distal colon cancer in the Japanese population. Since this study investigated the association between four SNPs and five haplotypes of *MUTYH* and the risk of CRC on the same set of samples, a method for multiple testing is applicable to this study. Therefore, the method of FDR was used for all of the results, and the statistical significance of the associations was found to remain essentially unchanged (Tables 1 and 3). These results coincide with the fact that tumors arising from different subsites of the colorectum differ in their population distribution, clinical features as well as genetic pathways.[34–37] It was suggested from our results that the IVS1+11 > T SNPs and TGAC haplotype of *MUTYH* may be involved in distal colon carcinogenesis and that the risk of cancer arising from each anatomic subsite of the colorectum may be modified by different genetic pathways. Further studies need to be conducted to elucidate the underlying mechanisms.

The development of CRC is a multistep, multifactor process.[38] Some studies have demonstrated that environmental factors and physical conditions may modify the genetic risk of CRC associated with SNPs.[39,40] This may also hold true for the CRC risk associated with *MUTYH* SNPs. In the present study, adjustment was made for gender, age and place of residence to evaluate the association between the *MUTYH* SNPs and CRC risk. The adjusted OR and 95% CI remained essentially unchanged after the adjustments as compared with the values obtained without the adjustments (Tables 1 and 3). Furthermore, when the body mass index, disease history, physical activity, dietary factors, smoking and alcohol consumption status were taken into consideration, the *P*-value for IVS1+11C > T was 0.058, a nearly significant value, and the *P*-value for the TGAC haplotype remained less than 0.05 (data not shown). This result suggested that the association between the *MUTYH* SNPs and the risk of CRC was not significantly modified by environmental factors or the physical condition.

In the present study, the 972G/C genotype was statistically significantly associated with increased risk of proximal colon, but not distal colon or rectal cancer (Table 2). The functional analysis revealed no difference between the C/C type and G/G type,[4] and the 972C allele is more frequently detected in Japanese and Chinese than in European populations as shown in the dbSNPs of the NCBI Entrez system. Taking this into consideration with our result, it could be suggested that the 972C allele may be inversely associated with the development of at least proximal colon cancer in the Japanese population. Alternatively, this inverse association of the 972C allele with the risk of proximal colon cancer in the Japanese population may arise from its interaction with other allele(s). The IVS10–2A > G SNP had been demonstrated to generate a protein without nuclear expression and the IVS10–2G allele was suggested to be associated with a low BER function in the cell nuclei and thereby, act as a risk allele for cancer. However, the results of analyses in this study revealed an OR of less than 0.7 (except for cancer of the proximal colon) for the IVS10–2G allele and the CGGG haplotype, which contains the IVS10–2G allele, although the *P*-value did not reach statistical significance (Tables 1–4). These results remained essentially unchanged even after adjustments for environmental factors and physical conditions (as described above). Investigation of some other additional clinical factors, such as pathological stage, recurrence or survival, might yield some association. On the other hand, some studies have suggested that SNPs of repair genes may be associated with reduced cancer risk or fewer recurrences, and that effective host DNA repair capacity may be associated with poorer survival.[41–43] These observations suggest that mutations in the repair genes may also be inversely associated with malignant alterations, in addition to their more widely recognized association with increased cancer risk. The inverse association of the IVS10–2A > G SNP detected in this study with colorectal cancer risk might be explained by the contention that individuals with the A allele may be more resistant to ROS or other stresses than individuals with the G allele, and that the A allele has a protective effect on cells with mutations, similar to the situation suggested by Wang *et al*. for the XRCC1 Arg194Trp variant, a SNP associated with a reduced risk for various types of cancers.[44] Research on the effect of *MUTYHβ* isoforms on the cellular responses to various mutagens are expected help in clarifying this issue.

Besides the RT-PCR experiment, reporter assay was also carried out to investigate whether the –280G > A and IVS1+11C > T SNPs may affect the promoter activity of *MUTYH*. The dual-luciferase reporter assay experiments detected high transcriptional activity of the region (–411~+356) of *MUTYH* (data not shown). This information will be of use for future analyses. The reporter plasmids containing the wild-type and mutant sequence for the responses to oxidative stress were also investigated using the colon cancer cell line HCT116. ROS was induced by glucose oxidase, menadione or H_2O_2 at appropriate concentrations and treatment durations. However, the two linked SNPs, –280G > A and IVS1+11C > T, did not affect the promoter activity in our setting (data not shown). This study did not detect any functional differences in the –280G > A/IVS1+11C > T/1389G > C SNPs, and there remains the possibility that the three SNPs might be linked with other SNPs and these SNPs might affect the susceptibility to CRC.

Acknowledgments

This work was supported in part by Grants-in-Aid for Young Scientists Category B (17790258) from the Japan Society for the Promotion of Science, Scientific Research in a Priority Area (18014009) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Ministry of Health, Labour and Welfare (15-2, 15-22, 19-19) of Japan and the 21st century COE program 'Medical Photonics', from the Smoking Research Foundation, and from the Foundation for Promotion of Cancer Research.

References

1 Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 1991; **349**: 431–4.
2 Cadet J, Berger M, Douki T, Ravanat JL. Oxidative damage to DNA. formation, measurement, and biological significance. *Rev Physiol Biochem Pharmacol* 1997; **131**: 1–87.
3 Slupska MM, Luther WM, Chiang JH, Yang H, Miller JH. Functional expression of hMYH, a human homolog of the *Escherichia coli* MutY protein. *J Bacteriol* 1999; **181**: 6210–13.
4 Shinmura K, Yamaguchi S, Saitoh T *et al*. Adenine excisional repair function of MYH protein on the adenine: 8-hydroxyguanine base pair in double-stranded DNA. *Nucl Acids Res* 2000; **28**: 4912–8.
5 Tsuzuki T, Nakatsu Y, Nakabeppu Y. Significance of error-avoiding mechanisms for oxidative DNA damage in carcinogenesis. *Cancer Sci* 2007; **98**: 465–70.
6 Hirano S, Tominaga Y, Ichinoe A *et al*. Mutator phenotype of *MUTYH*-null mouse embryonic stem cells. *J Biol Chem* 2003; **278**: 38 121–4.
7 Al-Tassan N, Chmiel NH, Maynard J *et al*. Inherited variants of MYH associated with somatic G:X→T:A mutations in colorectal tumors. *Nat Genet* 2002; **30**: 227–32.
8 Jones S, Emmerson P, Maynard J *et al*. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G. X→T: a mutations. *Hum Mol Genet* 2002; **11**: 2961–7.
9 Sieber OM, Lipton L, Crabtree M *et al*. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med* 2003; **348**: 791–9.
10 Sampson JR, Dolwani S, Jones S *et al*. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 2003; **362**: 39–41.
11 Gismondi V, Meta M, Bonelli L *et al*. Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. *Int J Cancer* 2004; **109**: 680–4.
12 Halford SE, Rowan AJ, Lipton L *et al*. Germline mutations but not somatic changes at the MYH locus contribute to the pathogenesis of unselected colorectal cancers. *Am J Pathol* 2003; **162**: 1545–8.
13 Colebatch A, Hitchins M, Williams R, Meagher A, Hawkins NJ, Ward RL. The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. *Br J Cancer* 2006; **95**: 1239–43.
14 Peterlongo P, Mitra N, Chuai S *et al*. Colorectal cancer risk in individuals with biallelic or monoallelic mutations of MYH. *Int J Cancer* 2005; **114**: 505–7.
15 Croitoru ME, Cleary SP, Di Nicola N *et al*. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J Natl Cancer Inst* 2004; **96**: 1631–4.
16 Farrington SM, Tenesa A, Barnetson R *et al*. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* 2005; **77**: 112–9.
17 Enholm S, Hienonen T, Suomalainen A *et al*. Proportion and phenotype of MYH-associated colorectal neoplasia in a population-based series of Finnish colorectal cancer patients. *Am J Pathol* 2003; **163**: 827–32.
18 Webb EL, Rudd MF, Houlston RS. Colorectal cancer risk in monoallelic carriers of MYH variants. *Am J Hum Genet* 2006; **79**: 768–71.
19 Tao H, Shinmura K, Hanaoka T *et al*. A novel splice-site variant of the base excision repair gene MYH is associated with production of an aberrant mRNA transcript encoding a truncated MYH protein not localized in the nucleus. *Carcinogenesis* 2004; **25**: 1859–66.
20 Miyaki M, Iijima T, Yamaguchi T *et al*. Germline mutations of the MYH gene in Japanese patients with multiple colorectal adenomas. *Mutat Res* 2005; **578**: 430–3.
21 Zhang Y, Liu X, Fan Y *et al*. Germline mutations and polymorphic variants in MMR, E-cadherin and MYH genes associated with familial gastric cancer in Jiangsu of China. *Int J Cancer* 2006; **119**: 2592–6.
22 Kim IJ, Ku JL, Kang HC *et al*. Mutational analysis of OGG1, MYH, MTH1 in FAP, HNPCC and sporadic colorectal cancer patients: R154H OGG1 polymorphism is associated with sporadic colorectal cancer patients. *Hum Genet* 2004; **115**: 498–503.
23 Lamlum H, Al Tassan N, Jaeger E *et al*. Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. *Hum Mol Genet* 2000; **9**: 2215–21.
24 Meijers-Heijboer H, van den Ouweland A, Klijn J *et al*. Low-penetrance susceptibility to breast cancer due to CHEK2 (*) 1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* 2002; **31**: 55–9.
25 Kono S, Toyomura K, Yin G, Nagano J, Mizoue T. A case-control study of colorectal cancer in relation to lifestyle factors and genetic polymorphisms: design and conduct of the Fukuoka colorectal cancer study. *Asian Pac J Cancer Prev* 2004; **5**: 393–400.
26 Hagiwara T, Kono S, Yin G *et al*. Genetic polymorphism in cytochrome P450 7A1 and risk of colorectal cancer: the Fukuoka Colorectal Cancer Study. *Cancer Res* 2005; **65**: 2979–82.
27 Kimura Y, Kono S, Toyomura K *et al*. Meat, fish and fat intake in relation to subsite-specific risk of colorectal cancer: the Fukuoka Colorectal Cancer Study. *Cancer Sci* 2007; **98**: 590–7.
28 Yin G, Kono S, Toyomura K *et al*. Alcohol dehydrogenase and aldehyde dehydrogenase polymorphisms and colorectal cancer: the Fukuoka Colorectal Cancer Study. *Cancer Sci* 2007; **98**: 1248–53.
29 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc Series B* 1995; **57**: 289–300.
30 Johnson GC, Esposito L, Barratt BJ *et al*. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; **29**: 233–7.
31 Shinmura K, Tao H, Yamada H *et al*. Splice-site genetic polymorphism of the human kallikrein 12 (KLK12) gene correlates with no substantial expression of KLK12 protein having serine protease activity. *Hum Mutat* 2004; **24**: 273–4.
32 Chen X, Truong TT, Weaver J *et al*. Intronic alterations in BRCA1 and BRCA2: effect on mRNA splicing fidelity and expression. *Hum Mutat* 2006; **27**: 427–35.
33 den Hollander AI, Koenekoop RK, Yzer S *et al*. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet* 2006; **79**: 556–61.
34 Wei EK, Giovannucci E, Wu K *et al*. Comparison of risk factors for colon and rectal cancer. *Int J Cancer* 2004; **108**: 433–42.
35 Cheng X, Chen VW, Steele B *et al*. Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992–97. *Cancer* 2001; **92**: 2547–54.
36 Ward R, Meagher A, Tomlinson I *et al*. Microsatellite instability and the clinicopathological features of sporadic colorectal cancer. *Gut* 2001; **48**: 821–9.
37 Lindblom A. Different mechanisms in the tumorigenesis of proximal and distal colon cancers. *Curr Opin Oncol* 2001; **13**: 63–9.
38 Wakabayashi K, Nagao M, Esumi H, Sugimura T. Food-derived mutagens and carcinogens. *Cancer Res* 1992; **52**: 2092s–8s.
39 Marchand LL. Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. *J Natl Cancer Inst Monogr* 1999, 101–5.
40 Yeh CC, Hsieh LL, Tang R, Chang-Chieh CR, Sung FC. MS-920: DNA repair gene polymorphisms, diet and colorectal cancer risk in Taiwan. *Cancer Lett* 2005; **224**: 279–88.
41 Catto JW, Xinarianos G, Burton JL, Meuth M, Hamdy FC. Differential expression of hMLH1 and hMSH2 is related to bladder cancer grade, stage and prognosis but not microsatellite instability. *Int J Cancer* 2003; **105**: 484–90.
42 Bosken CH, Wei Q, Amos CI, Spitz MR. An analysis of DNA repair as a determinant of survival in patients with non-small-cell lung cancer. *J Natl Cancer Inst* 2002; **94**: 1091–9.
43 Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 1513–30.
44 Wang Y, Spitz MR, Zhu Y, Dong Q, Shete S, Wu X. From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. *DNA Repair (Amst)* 2003; **2**: 901–8.

doi: 10.1111/j.1349-7006.2007.00694.x