SPANDIDOS PUBLICATIONS

Differences in expression between transcripts using alternative promoters of hMLH1 gene and their correlation with microsatellite instability

RYOSUKE TAKAHASHI[1] and KEIICHI NAGAI[2]

[1]Advanced Research Laboratory, Hitachi, Ltd., Hatoyama, Saitama 350-0395; [2]Central Research Laboratory, Hitachi, Ltd., 1-280, Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan

Received April 3, 2009; Accepted May 11, 2009

DOI: 10.3892/or_00000433

Abstract. hMLH1 is involved in DNA mismatch repair and its defects cause hereditary non-polyposis colorectal cancer (HNPCC) as well as other types of cancer. A defective DNA mismatch repair system results in genetic instability, also referred to as microsatellite instability (MSI), which is a good indicator of HNPCC. Using *in silico* analysis of oligo-capping cDNA sequences, we initially identified a splicing (variant type 2) whose second exon is 5 bp shorter than that of a genuine hMLH1 transcript (variant type 1) and a transcript using alternative promoter (variant type 3) whose transcription starts about 300 bases downstream of variant type 1. We then compared the expression level of variant types 1 and 3 among six colorectal cancer cell lines using real-time PCR. As a result, we found that the cell lines that completely suppress the expression of variant type 1 by hypermethylation expressed variant type 3 to a certain extent. This result suggests that the expression of variant types 1 and 3 did not completely follow the same transcription mechanism. We also found that the cell lines showing MSI to be positive either expressed variant type 3 more than type 1 or expressed only variant type 3. These results showed the potential applicability of mRNA expression analysis to molecular diagnostic tests of MSI-positive cancer types.

Introduction

The expressional and translational diversity of mRNAs or proteins from each single gene are mainly controlled by alternative splicing or use of alternative promoters in the mRNA transcription and RNA processing stages (1,2). Alternative splicing is a significant mechanism for modulating the expression of genes and enables a single gene to increase its coding pattern, producing splicing variants and enabling the synthesis of several structurally and functionally distinct protein isoforms (1). Alternative promoters are alternative regions from which transcripts of genes originate. The existence of alternative promoters means that a gene has several transcriptional start sites (TSSs). Consequently, several types of transcriptional products that differ in their 5' terminal are produced (2). Recent research on 1,780,295 types of 5'-end cDNA sequences obtained from oligo-capping cDNA clones, showed that ~52% of genes assigned to the RefSeq database (http://ncbi.nih.gov/RefSeq/) were found to be under the influence of alternative promoters. Among them, there is an average of 3.1 alternative promoters per gene (3). These phenomena have been found to be associated with a variety of diseases, including cancer (4,5).

DNA mismatch repair system is the primary pathway for correcting replication errors. In >90% of hereditary non-polyposis colorectal cancer (HNPCC) syndrome cases, the representative disease of familial colorectal cancer and a defective DNA mismatch repair system resulted in genetic instability, also referred to as microsatellite instability (MSI) (6,7). Microsatellites are short repetitive sequences: two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D2S123, D5S346 and D17S250). These five nucleotide repeats are called microsatellite markers, and their instabilities indicate the number by which the repetitive sequences decrease or increase (8). MSI is easily observed using PCR, which detects microsatellite markers (9). A total of ~85% of HNPCC patients exhibit MSI, and this proportion is even higher in mutation-positive families (10). MSI is also found in 10-15% of sporadic colorectal tumor patients (10).

The genetic diagnosis of cancer is generally based on the mutation analysis of responsible gene(s). Unlike the APC gene, which is responsible for familial adenomatous polyposis syndrome (11), five mismatch repair genes (hMLH1, hMSH2, hMPS1, hPMS2 and hMSH6) have been identified for HNPCC (10). Of the five genes, hMLH1 (12,13) and hMSH2 (14) are the most extensively analyzed genes associated with HNPCC syndrome. hMLH1, for example, causes HNPCC by a nonsense mutation of codon 252 in exon 9 (TCA-TAA, Ser-Stop) (13) or by a lack of hMLH1 mRNA exons 6 and 7

Correspondence to: Dr Keiichi Nagai, Central Research Laboratory, Hitachi, Ltd., 1-280, Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan
E-mail: keiichi.nagai.mh@hitachi.com

Key words: alternative promoter, splicing variant, microsatellite instability, hereditary non-polyposis colorectal cancer

(15). In addition, loss of hMLH1 expression due to hypermethylation of the hMLH1 promoter region leads to colorectal cancer by means of MSI (16).

In the present study, using *in silico* analysis of oligo-capping cDNA sequences, we identified a new transcript using an alternative promoter which starts ~300 bases downstream of the genuine hMLH1 TSS. Using real-time PCR on three types of colorectal cancer cell lines, which have distinctive characteristics regarding hypermethylation and MSI, we compared the expression levels of transcripts derived from each promoter. We also evaluated the potential applicability of mRNA expression analysis to molecular diagnostic tests of MSI-positive cancer types.

Materials and methods

Cell lines and culture. SW48 (colorectal adenocarcinoma), RKO and HCT116 (colorectal carcinoma) were obtained from the American Tissue Culture Collection (Manassas, VA). LoVo, COLO320 DM and COLO201 (colorectal adenocarcinoma) were obtained from the Human Science Research Resources Bank (Osaka, Japan). Cells were cultured in advanced DMEM (Invitrogen, Carlsbad, CA) and supplemented with 10% fetal bovine serum (Invitrogen) and antibiotics. Cell cultures were maintained in a humidified incubator at 37˚C under 5% CO_2.

Acquisition of hMLH1-related genomic and cDNA alignments. 5'-end human cDNA sequences from full-length cDNA libraries derived from 164 kinds of human tissues, and cultured cells were isolated from the deposition sequences described in the literature (3). These cDNA libraries were constructed using oligo-capping methods (17). Using the RefSeq sequence of hMLH1 (NM_000249) as a query, a computer-based homology search using the blastn program from the BLAST package (18) was applied to these sequences that were multi-FASTA formatted. Sequences with a high degree of similarity (blast E value $<10^{-10}$) were selected as hMLH1-related cDNAs. To obtain the corresponding alignments, the bl2seq program from the BLAST package was carried out between the genomic sequences in the hMLH1 region dissected from the human genome chromosome 3 sequence (NT_022517) and each sequence of the hMLH1-related cDNAs described above. Subsequently, the alignments that were correctly aligned ahead of the second exonic region of the hMLH1 RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) sequence (NM_000249) and those that also closely matched (an identity of 98% or more) were selected from the whole set of alignments.

Analysis of hMLH1 mRNA expression. hMLH1 gene expression was measured by TaqMan-based real-time reverse transcriptase-PCR. Total RNA was extracted from cell lines using a DNeasy tissue kit (Qiagen, Hilden, Germany). Total RNA (1.0 μg) was reverse-transcribed with 2.0 μl of oligo-$dT_{(20)}$ primer (50 μM), 4.0 μl of dNTP mix (5 mM), 8.0 μl of 5X first-strand buffer, 2.0 μl of DTT (100 mM), 2.0 μl of DEPC-treated water, and SuperScript™ III reverse transcriptase (Invitrogen) with a total volume of 40 μl according to the manufacturer's instructions.

After reverse transcription, real-time PCR was conducted in an ABI PRISM 7900 HT (Applied Biosystems, Foster City, CA) to determine the expression levels of variant types 1 and 3 with a final reaction mixture of 1.0 μl of RT-products, 10 μl of TaqMan Universal Mastermix, 3.5 μl of forward primer (F, 5.0 μM), 3.5 μl of reverse primer (R, 5.0 μM), 3.5 μl of TaqMan probe (5.0 μM) and water with a total volume of 20 μl using a 384-well optical tray. PCR conditions were: initial denaturing at 95˚C for 10 min, 40 cycles of denaturing at 95˚C for 15 sec and annealing/extension at 60˚C for 1 min. The forward and reverse primer for variant type 1 and probe sequences were: 5' to 3', F: CAGCGGCCAGCTAATGCTAT, R: CCATTGTCTTGATCTGAATCAACTTC, and Probe: FAM-CAAGTATTCAAGTGATTGTTAAAGAGGGAG GCC-TAMRA (Applied Biosystems). The forward primer sequence for variant type 3 was 5' to 3', F: GGGTTGTTT GGAGTTTTAGATGCA. The reverse primer and the TaqMan probe for variant type 3 were the same as those for variant type 1. The fluorescence emitted by the reporter dye was detected in real time, and the threshold cycle (Ct) of each sample was recorded as a quantitative measure of the amount of PCR product in the sample.

The standard curve based on the Ct was generated using cDNA from the cell line COLO320 DM, which expresses high-level hMLH1 mRNA. PCR products were prepared before the real-time PCR using cDNA from COLO320 DM as a template with a final mixture of 2.0 μl of RT-products, 5.0 μl of 10X PCR buffer, 4.0 μl of dNTP mix (2.5 mM), 3.5 μl of forward primer (5.0 μM), 3.5 μl of reverse primer (5.0 μM), 0.25 μl of ExTaq (Takara Bio Inc., Japan) and water with a total volume of 50 μl. PCR conditions were: initial denaturing at 95˚C for 10 min, 35 cycles of denaturing at 95˚C for 30 sec, annealing at 63˚C for 1 min and extension at 72˚C for 1 min, as well as a final extension at 72˚C for 10 min. The forward and reverse primer sequences for a standard curve of variant type 1 were: 5' to 3', F: CTGGACGAGACA GTGGTGAAC and R: ATACTGGCTAAATCCTCAAAGG ACTG. The forward primer sequence for the standard curve of variant type 3 was 5' to 3', F: TTTCTTTACCGCTCTCC CCCG. The reverse primer was the same as that for variant type 1. After serial dilution, the PCR products were measured by real-time PCR and the Ct was determined. The Ct is the fractional cycle number at which the fluorescence generated by the reporter dye exceeds a fixed level above the baseline.

Results

Identification of an alternative promoter of the hMLH1 gene using in silico analysis of oligo-capping cDNA sequences. The human MLH1 (hMLH1, human mutL homologue 1) gene is an *E. coli* homologue of the DNA mismatch repair gene mutL and is recognized as a gene locus at which high frequency mutations are observed on HNPCC. RefSeq mRNA sequence of hMLH1 consists of 19 exons.

Using *in silico* analysis (see Materials and methods), we obtained 187 hMLH1-related cDNA sequences which were assumed to have TSSs. We classified these sequences into variant types 1, 2 and 3. The first three exons of the 19 exonic regions, variant types 1, 2 and 3, are shown in Fig. 1(a) (12,19). The hMLH1 RefSeq sequence belongs to variant

NA databank accession numbers and variant type.[a]

Type 1							Total 141
AU131219	DA031179	DA050403	DA110166	DB243737	DA707133	DA173423	DA262636
AU127758	DA716690	DA125431	DB040429	DA909925	AU128432	DA501808	DA819642
DA619769	DA472218	DA043349	DA880601	DA453337	DA555707	DA881017	DA356018
DA024420	DA983131	DA695612	DA475144	DA055019	DB092229	DB000482	DA919830
DA953555	DA702671	DB272759	DA762271	DA209147	DA555444	DA546685	DA751926
DA023897	DA701340	DB090691	DA503831	DA914770	DA880449	DA349545	DA207023
DA486475	DB200826	DA556556	DA494857	DA472331	DA033654	DB008432	DA659014
DA932049	DA304561	DB274178	DB252001	DA609156	DA305360	DA970996	DA599881
DA426942	DA658037	DA659112	DB120181	DA752464	DA013759	DA308229	DA503294
DA555173	DA624450	DB270772	DA666173	DB014982	DB266774	DA729732	DB011897
DA479521	DA489729	DB269584	DA631956	DA958303	DA877866	DA972163	DA582565
DA524635	DA476333	DA559708	DA518707	DA952805	DB055056	DA563477	DA287218
DA447945	AU280370	DA913772	DA698928	DA783187	DA887994	DA554567	DA765712
DA952842	DA374168	DA619525	DB246315	DA884873	DA997466	DA690019	DA656699
DA350100	DA201970	DA478068	DA106361	DA794400	DA874359	DA738557	DA592612
DA915432	DA577087	DA503277	DA712069	DA287058	DA478450	DA086654	DA004869
DA790495	DA926525	DA937827	DA493353	DB040297	DB089707	DA660079	DA510241
DA050020	DA814813	DA650335	DA911985	DA348650			
Type 2							Total 14
DA385243	DA868597	DA815001	DA070867	DB139980	DB287548	DB157098	DB012065
DB114802	DB018927	DA866298	DA352218	DA506267	DB252510		
Type 3							Total 32
AU127122	DA517219	DB204422	DA814707	DA927898	DA430514	DA922309	DB012054
DA935122	DA343206	DA930241	DA166268	DA921733	DA303777	DA476068	DA296090
DB253300	DA156778	DA566031	DA748219	DA963961	DA664790	DA580525	DB259565
DA073874	DA749519	DA750838	DA582466	DA172498	DA275297	DA821467	DA955738

[a]Types 1, 2 and 3 represent 75, 7.4 and 17.6% of the total, respectively.

type 1. Variant type 2 has almost the same exonic structure, but the 5'-end of the second exon is 5 bp shorter than that of variant type 1. We assumed that type 2 is a splicing variant. Variant type 3 has the same sequence as exons 2 and 3 of variant type 1, but the TSSs are located ~300 bp downstream of those of variant types 1 and 2 (Fig. 2). Of the 187 hMLH1-related cDNA sequences, 141 belonged to variant type 1, 14 belonged to variant type 2 and 32 belonged to variant type 3. The corresponding percentages of the total were 75, 7.5 and 17%, respectively. Even though variant type 1 is a significant transcript, variant types 2 and 3 are also transcribed to a large extent. Accession numbers of the hMLH1-related cDNA sequences in the public DNA sequence databanks are listed in Table I. The TSSs, whose sequences are shown in Fig. 2, were scattered over a relatively wide range for each variant type. As the promoter region of the hMLH1 gene lacked the canonical TATA box, this is not unusual (20).

Expressional change of variant types 1 and 3 of hMLH1 gene. A recent study showed that the range over which the TSSs are scattered is on average 62 bp, with a standard deviation of 20 bp. Moreover, it was assumed that the possibility of TSSs belonging to a single cluster scattered over a range of >500 bp would be very low (3). It would be interesting to know whether or not variant types 1 and 3 are subject to different expressional (or alternative promoter usage) regulations since the TSS intervals of these genes are only ~300 bp.

We examined the dependence of the expression on the position of the TSS using colorectal cancer cell lines. We used three types of cell lines based on the results in the literature (Table II) (13,14). COLO201 and COLO320 DM were methylation- and MSI-free. RKO and SW48 were methylation- and MSI-positive, while HCT116 and LoVo were methylation-free but MSI-positive. Since preliminary results showed that, under the present experimental conditions,

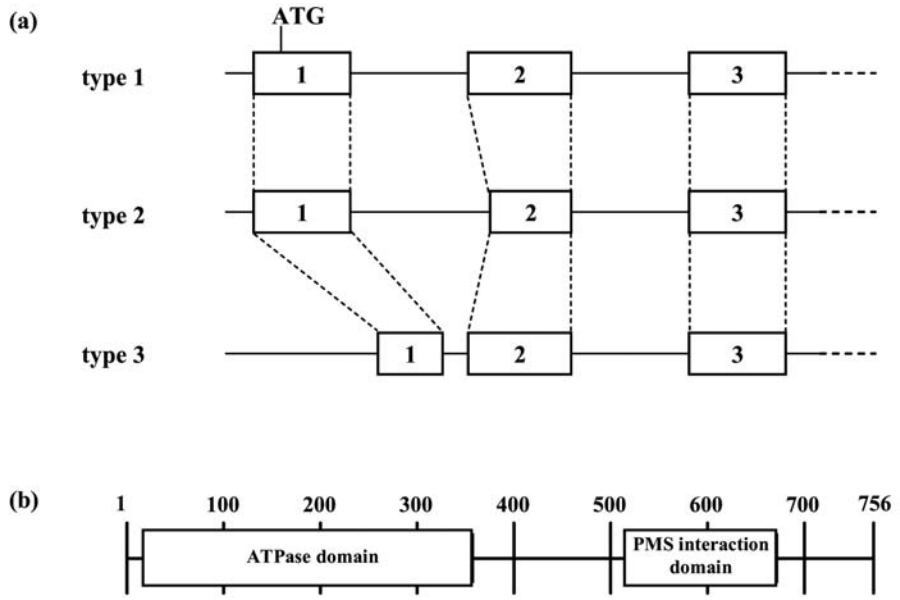

Figure 1. Variant types of the hMLH1 gene and hMLH1 functional domain. (a) Type 1 is considered to be a significant transcript structure for encoding protein (12,19). Type 2 lacks the first 5 nucleotides of exon 2 and will not create an intact protein. Type 3 has an alternative promoter and transcriptional starting site that starts after ~300 nucleotides of type 1. (b) Numbers represent amino acid position. hMLH1 encodes a protein consisting of 756 amino acids that include a highly conserved N-terminal domain (ATPase domain) and a variant C-terminal domain (PMS interaction domain) (12,21).

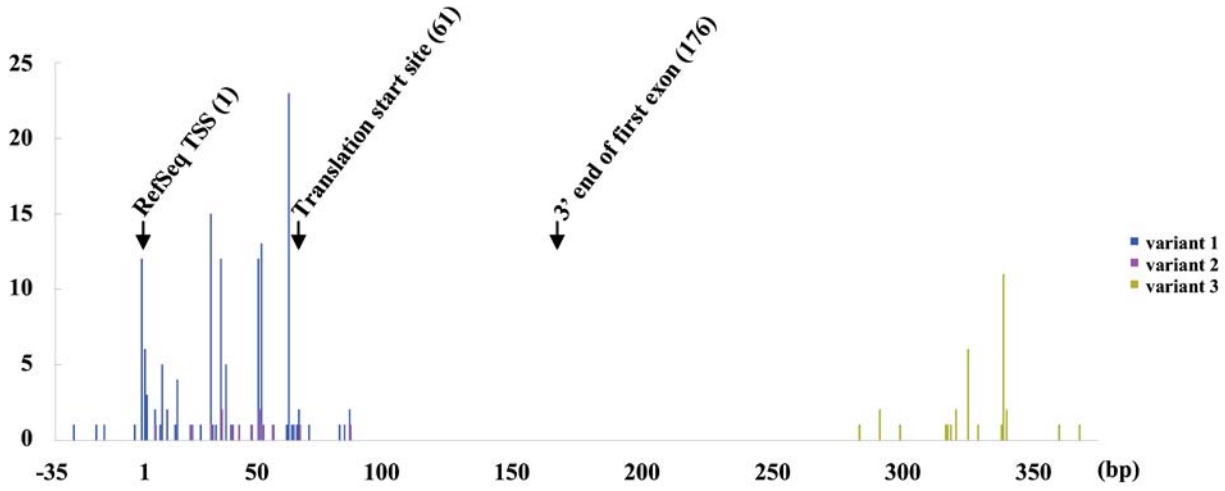

Figure 2. TSSs of the three variants and the number of clones. TSSs of the three variants were scattered in a range of ~400 bp. Number of hMLH1 clones was plotted against the TSS position. Blue, pink and yellow bars indicate variant types 1, 2 and 3 respectively. Arrows show RefSeq TSS, translation start site and the 3' end of the first exon.

Table II. Expression, microsatellite instability (MSI) and methylation status among cell lines.[a]

	COLO201	COLO320 DM	RKO	SW48	HCT116	LoVo
Expression	1>3	1>3	1(-), 3(+)	1(-), 3(+)	1<3	1<3
MSI	-	-	+	+	+	+
Methylation	-	-	+	+	-	-

[a]Data for expression were obtained from experimental results (Fig. 3). Data for MSI and methylation status are found in the literature (22), while the MSI status was determined by an analysis of the BAT25 and BAT26 loci. Methylation status of the hMLH1 promoter was investigated between bases -711 and +15 in each cell line. 1<3 means the copy number of variant type 3 is higher than that of type 1, and 1>3 the opposite. 1(-) means the expression of variant type 1 was negative and 3(+) means the expression of variant type 3 was positive.

riant Type 1

AAGAACGTGA GCACGAGGCA CTGAGGTGAT TGGCGAAAGG CACTTCCGTT GAGCATCTAG
ACGTTTCCTT GGCTCTTCTG GCGCCAAAAT GTCGTTCGTG GCAGGGGTTA TTCGGCGGCT
GGACGAGCA GTGGTGAACC GCATCGCGGC GGGGGAAGTT ATCCAGCGGC CAGCTAATGC
TATCAAAGAG ATGATTGAGA ACTG**TTTAGA** **TGCAAAATCC** **ACAAGTATTC** **AAGTGATTGT**
TAAAGAGGGA **GGCCTGAAGT** **TGATTCAGAT** **CCAAGACAAT** **GGCACCGGGA** **TCAG**GAAAGA
AGATCTGGAT ATTGTATGTG AAAGGTTCAC TACTAGTAAA CTGCAGTCCT TTTGAGGATTT
AGCCAGTATT TCTACCTATG GCTTTCGAGG TGAGG

(b) Variant Type 3

GCATGCCCAC AAGGGCGAG GCCGCCGGGT TCCCTGACGT GCCAGTCAGG CCTTCTCCTT
TTCCGCAGAC CGTGTGTTTC TTTACCGCTC TCCCCCGAGA CCTTTTAAGG GTTGTTTGGA
GTTTTAGATG **CAAAATCCAC** **AAGTATTCAA** **GTGATTGTTA** **AAGAGGGAGG** **CCTGAAGTTG**
ATTCAGATCC **AAGACAATGG** **CACCGGGATC** **AG**GAAAGAAG ATCTGGATAT TGTATGTGAA
AGGTTCACTA CTAGTAAACT GCAGTCCTTT GAGGATTTAG CCAGTATTTC TACCTATGGC
TTTCGAGGTG AGG

Figure 3. First three exons of hMLH1 variant types 1 and 3. The second exon is shown in bold. ATG is underlined.

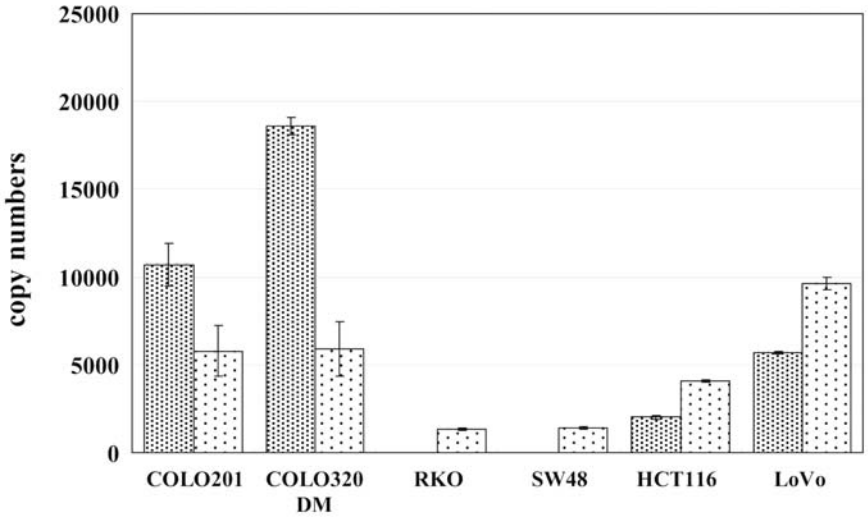

Figure 4. hMLH1 mRNA expression of variant types 1 (left column) and 3 (right column) in the cell lines LoVo, HCT116, RKO, SW48, COLO320 DM, COLO201. Total RNA isolated from cells was reverse-transcribed, and cDNA was amplified by PCR. Copy numbers were measured using real-time PCR.

the expression of variant type 2 was negligibly small compared with that of variant type 1, we examined the difference between variant types 1 and 3 of hMLH1 whose 5' most extended sequences are shown in Fig. 3.

As a result, we categorized the cell lines into three groups in terms of the expressional change of variant types 1 and 3 by quantitative real-time PCR (Fig. 4). Thus, the expressional patterns of variant type 3, whose existence was shown in the present investigation for the first time, were completely different from those of variant type 1. In the COLO201 and COLO320 DM group, variant type 1 expressed at a higher rate than variant type 3. In contrast, even though the expression of variant type 1 of RKO and SW48 was completely suppressed by methylation in the promoter region, their expression of variant type 3 was observed. In addition, in the HCT116 and LoVo group, although the expression of variant type 1 was shown to be positive by qualitative real-time PCR (13), it was down-regulated compared with the cases of COLO201 and COLO320 DM, while the expression of variant type 3 showed almost no change. In other words, MSI-positive cell lines expressed variant type 3 more than variant type 1, including a complete suppression of variant type 1 cell lines RKO and SW48.

Discussion

Role of variant type 3 of hMLH1 gene. The domain structure of the hMLH1 protein is shown in Fig. 1(b) (12,21). hMLH1 encodes a protein of 756 amino acids including a highly conserved N-terminal ATPase domain (ATPase domain) as well as a variant C-terminal domain (PMS interaction domain) (21). The translation start site (ATG) of the hMLH1 protein

is in the first exon of variant type 1. As the second ATG in that reading frame is also in the first exon of variant type 1, the most plausible coding sequence for variant types 2 and 3 starts from the third ATG of the genuine hMLH1 coding sequence, which starts 723 bp downstream of the first ATG. This causes a 241 amino acid truncation at the N-terminal end of the hMLH1 protein. The predicted protein sequence coded by variant type 3 mRNA lacks most of its ATPase domain including conserved loops L45 that are thought to contain the MutS-MutL interaction interface (21). In spite of this, we found that variant type 3 is transcribed to a large extent.

Although the mechanism for variant type 3 expression and the specific effect of the expression on carcinogenesis are unknown, we considered from these results that the role of variant type 3 has at least three possible roles to play. First, variant type 3 is 'junk'. Second, the translatable protein from variant type 3 lacks ~30% of the amino acids at the N-terminal end of the intact hMLH1 protein. This results in the loss of most of the ATPase domain whose original function was that of mismatch repair, and thus exhibits the biological functions of a shorter protein. Third, it is a so-called functional non-coding RNA. Recently, Martianov *et al* reported that non-coding RNA from a minor promoter, which is different from a major one (also known as a normal promoter) for complete mRNA transcription, causes transcriptional repression (22). The first case seems to be a waste of energy and biologically meaningless. However, it is plausible that a certain amount of 'junk' mRNAs are transcribed as many splicing variants and alternative promoters have been identified (1,2). If the transcript of variant type 3 codes a functional shorter protein or is a functional non-coding RNA, it can affect the expression of hMLH1, resulting in MSI and carcinogenesis. It may also have a novel function. However, revealing the role of variant types 2 and 3 requires further investigation.

Expressional regulation of variant types 1 and 3 of hMLH1. Human MSH2 (hMSH2, human MutS homologue 2) is a significant responsible gene that induces MSI as well as hMLH1. Germ-line mutations or genomic deletions of hMSH2 or hMLH1 cause HNPCC syndrome (23-25). One group that measured protein expression by immunohistochemistry has reported that ~90% of families that express only hMLH1 (variant type 1) or hMSH2 proteins are MSI-positive, while families expressing hMLH1 and hMSH2 are microsatellite stable (26).

We showed that LoVo and HCT116 express variant type 3 more than variant type 1 and they show MSI to be positive. HCT116 has a mutation in exon 9 of hMLH1 that introduces a translational stop codon (TCA-TAA), and LoVo has truncated RNA (exon 4-8 deletion) for the hMSH2 gene (13). These mutations produce defective hMLH1 and hMSH2 proteins, respectively, which probably causes MSI in each cell line. Accordingly, it is notable that defects in the hMLH1 and hMSH2 proteins down-regulate the expression of the intact mRNA of the hMLH1. There is also a common mechanism for expressing the two mRNAs that require the existence of both intact proteins. Using a publicly available database, we investigated whether or not there is a correlation between the expression of hMLH1 and hMSH2 mRNA. The Genomics Institute of the Novartis Research Foundation, has designed custom arrays that interrogate the expression of the vast majority of protein-encoding human and mouse genes and has used them to profile a panel of 79 human and 61 mouse tissues, and the results are available at http://symatlas.gnf.org/SymAtlas/ (27). The correlation analysis on that site showed a close correlation between hMLH1 and hMSH2 gene expression. The array system used the array analysis protocols gcRMA (28) and Affymetrix's default MAS5.0 (29), for which correlation data was provided. According to these analyses, hMSH2 was positioned 8th by gcRMA and 14th by MAS5.0 to hMLH1 for >20,000 target sequences and hMLH1 was positioned 6th by gcRMA and 11th by MAS5.0 to hMSH2 (data not shown). Furthermore, among the genes whose expression correlates with the hMSH2 gene, there is only one common gene that ranks ahead of hMLH1. Results obtained from the two analysis methods also showed that among the genes whose expression correlates with the hMLH1 gene, there is also only one common gene that ranks ahead of hMSH2. This shows a close correlation between the expression of the two genes. Thus, there may be a common mechanism to transcribe them in the same manner as that regulated by the two proteins.

Applicability of expressional changes between variant types to cancer diagnosis. Colorectal cancer is one of the most serious malignant tumors that cause mortality, but is curable if it is surgically treated in the early stages. The hemoccult fecal blood test is currently used for mass screening, but it is non-specific and not sufficiently sensitive to detect early stage colorectal cancer (30). Screening methods for colorectal cancer that make use of the detection of DNA mutations in the cancer-related genes have been reported. Among these, Matsumura's group reported a method that isolates colonocytes from several grams of feces, detects DNA mutations for APC, K-ras and p53 genes, and analyzes MSI (31). These methods mainly depend on the detection of the mutation of the multi-step process model-related genes of colorectal cancer, and attained sensitivities are at most ~70%. Research indicates that 10-15% of sporadic colorectal tumors show MSI (10). Thus, the effective detection of MSI or defects of mismatch repair will improve those sensitivities. We propose a novel method that measures the expression ratio of variant types 1-3 of the hMLH1 gene to improve sensitivity. Even though a method described in the literature (10) uses BAT26 analysis, BAT26 alone is not sufficient to detect MSI or the defection of mismatch repair gene(s) as a whole. MSI is a result of a mismatch repair defect, making the detection of the latter more direct and therefore more effective. Matsumura's group also demonstrated recovery mRNA from exfoliated colonocytes isolated from feces (32). Consequently, gene expression measurement in a single cell has become feasible (33). Quantitative measurement has always been considered to have difficulties in maintaining quantitativeness, but the proposed method has resolved this issue by adding an ideal inner standard: variant type 3. We anticipate that the present findings will contribute to the development of an early stage method of screening for colorectal cancer.

SPANDIDOS PUBLICATIONS **edgements**

We thank Dr Hiroyuki Tsunoda for his technical advice and helpful discussion.

References

1. Caceres JF and Kornblihtt AR: Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet 18: 186-193, 2002.
2. Landry JR, Mager DL and Wilhelm BT: Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet 19: 640-648, 2003.
3. Kimura K, Wakamatsu A, Suzuki Y, *et al*: Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 16: 55-65, 2006.
4. Brinkman BM: Splice variants as cancer biomarkers. Clin Biochem 37: 584-594, 2004.
5. Venables JP: Aberrant and alternative splicing in cancer. Cancer Res 64: 7647-7654, 2004.
6. Thibodeau SN, Bren G and Schaid D: Microsatellite instability in cancer of the proximal colon. Science 260: 816-819, 1993.
7. Parsons R, Li GM, Longley M, *et al*: Mismatch repair deficiency in phenotypically normal human cells. Science 268: 738-740, 1995.
8. Boland CR, Thibodeau SN, Hamilton SR, *et al*: A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58: 5248-5257, 1998.
9. Weber JL and May PE: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 388-396, 1989.
10. Loukola A, Eklin K, Laiho P, *et al*: Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). Cancer Res 61: 4545-4549, 2001.
11. Kinzler KW, Nilbert MC, Su LK, *et al*: Identification of FAP locus genes from chromosome 5q21. Science 253: 661-665, 1991.
12. Bronner CE, Baker SM, Morrison PT, *et al*: Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 368: 258-261, 1994.
13. Deng G, Chen A, Hong J, *et al*: Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res 59: 2029-2033, 1999.
14. Fishel R, Lescoe MK, Rao MR, *et al*: The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75: 1027-1038, 1993.
15. Tanko Q, Franklin B, Lynch H, *et al*: A hMLH1 genomic mutation and associated novel mRNA defects in a hereditary nonpolyposis colorectal cancer family. Mutat Res 503: 37-42, 2002.
16. Wheeler JM, Loukola A, Aaltonen LA, *et al*: The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. J Med Genet 37: 588-592, 2000.
17. Suzuki Y and Sugano S: Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. Methods Mol Biol 221: 73-91, 2003.
18. Altschul SF, Gish W, Miller W, *et al*: Basic local alignment search tool. J Mol Biol 215: 403-410, 1990.
19. Han HJ, Maruyama M, Baba S, *et al*: Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Hum Mol Genet 4: 237-242, 1995.
20. Ito E, Yanagisawa Y, Iwahashi Y, *et al*: A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. Biochem Biophys Res Commun 256: 488-494, 1999.
21. Plotz G, Raedle J, Brieger A, *et al*: N-terminus of hMLH1 confers interaction of hMutLalpha and hMutLbeta with hMutSalpha. Nucleic Acids Res 31: 3217-3226, 2003.
22. Martianov I, Ramadass A, Serra Barros A, *et al*: Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445: 666-670, 2007.
23. Wijnen J, van der Klift H, Vasen H, *et al*: MSH2 genomic deletions are a frequent cause of HNPCC. Nat Genet 20: 326-328, 1998.
24. Charbonnier F, Olschwang S, Wang Q, *et al*: MSH2 in contrast to MLH1 and MSH6 is frequently inactivated by exonic and promoter rearrangements in hereditary nonpolyposis colorectal cancer. Cancer Res 62: 848-853, 2002.
25. Wahlberg SS, Schmeits J, Thomas G, *et al*: Evaluation of microsatellite instability and immunohistochemistry for the prediction of germ-line MSH2 and MLH1 mutations in hereditary nonpolyposis colon cancer families. Cancer Res 62: 3485-3492, 2002.
26. Renkonen E, Zhang Y, Lohi H, *et al*: Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. J Clin Oncol 21: 3629-3637, 2003.
27. Su AI, Wiltshire T, Batalov S, *et al*: A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 101: 6062-6067, 2004.
28. Wu Z and Irizarry RA: Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol 12: 882-893, 2005.
29. Hubbell E, Liu WM and Mei R: Robust estimators for expression analysis. Bioinformatics 18: 1585-1592, 2002.
30. Mandel JS, Church TR, Bond JH, *et al*: The effect of fecal occult-blood screening on the incidence of colorectal cancer. N Engl J Med 343: 1603-1607, 2000.
31. Matsushita H, Matsumura Y, Moriya Y, *et al*: A new method for isolating colonocytes from naturally evacuated feces and its clinical application to colorectal cancer diagnosis. Gastroenterology 129: 1918-1927, 2005.
32. Yamao T, Matsumura Y, Shimada Y, *et al*: Abnormal expression of CD44 variants in the exfoliated cells in the feces of patients with colorectal cancer. Gastroenterology 114: 1196-1205, 1998.
33. Levsky JM, Shenoy SM, Pezo RC, *et al*: Single-cell gene expression profiling. Science 297: 836-840, 2002.