Identification of natural antisense transcripts involved in human colorectal cancer development

KEISUKE KOHNO[1*], MITSURU CHIBA[1*], SOICHIRO MURATA[1], SUGIRU PAK[1], KENTARO NAGAI[1], MASAYOSHI YAMAMOTO[1], KAZUHIKO YANAGISAWA[1], AKIHIKO KOBAYASHI[1], HIROSHI YASUE[2] and NOBUHIRO OHKOHCHI[1]

[1]Department of Surgery, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575; [2]Animal Genome Research Unit, National Institute of Agrobiological Sciences, 2 Ikenodai, Tsukuba, Ibaraki 305-0901, Japan

Received July 12, 2010; Accepted August 30, 2010

DOI: 10.3892/ijo_00000794

Abstract. Natural antisense transcripts (NATs) constitute a class of non-coding RNAs that have emerged as important regulators of gene expression. However, involvement of NATs in colorectal cancer (CRC) development has not been reported to date. In the present study, the up- and down-regulation of NATs were investigated in human CRC for their possible involvement in CRC development. Total RNAs isolated from 51 CRC tissues, 9 corresponding non-cancerous tissues and 19 liver metastatic tissues from surgically resected samples were subjected to expression analysis using a custom-microarray containing human sense/antisense probes for *ca*. 21,000 genes. Comparing CRC tissues with non-cancerous tissues, we identified 415 NATs differentially expressed in CRC and non-cancerous tissues to a significant degree ($p<0.001$, fold change >4.0 or ≤4.0). When a hierarchical clustering was performed on CRC and non-cancerous samples using these 415 NATs, the samples were separately clustered. Principal component analysis with the same NATs showed clear separation of CRC and non-cancerous samples using the first two principal components (PC1, 80%; PC2, 10%). To validate the expression results obtained from the microarray, the expressions of the 3 selected NATs were examined by strand-specific RT-qPCR, revealing that these expression profiles were consistent with those obtained from microarray analysis. In addition, the NAT expression patterns were found to be different between primary tumors with liver metastasis and those without liver metastasis. In conclusion, these findings taken together indicated that NATs indentified in the present study would be involved in CRC development as well as possibly in its metastasis.

Correspondence to: Dr Nobuhiro Ohkohchi, Department of Surgery, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan
E-mail: nokochi3@md.tsukuba.ac.jp

*Contributed equally

Key words: natural antisense transcripts, human colorectal cancer, microarray, gene expression profile, strand-specific RT-qPCR

Introduction

Colorectal cancer (CRC) is one of the most frequent cancers in the world. The American Cancer Society estimates that CRC was the third leading cause of cancer deaths in both men and women in 2009 (1). In Japan, the prevalence of CRC patients has doubled in the past two decades, and CRC has been the second cause of death in neoplastic diseases (2).

CRC is a heterogeneous disease arising from a complex series of molecular events. The evolution of normal colonic mucosa to a potentially invasive cancer via benign adenoma has been reported to be associated with a series of genetic events (3). Molecular detection methods based on gene mutation for *APC, p53* and *K-ras*, have been developed within the past two decades (4). Despite the advent of these molecular markers, their usage is still limited for diagnosis of CRC, due to the fact that the CRC detection rate is not high enough for practical usage, indicating that additional factors should be involved in CRC development. Therefore, identification of additional molecular events involved in CRC development is essential for more accurate diagnosis of CRC including the precancerous state.

Since microarray technology has provided information on expression levels of thousands of genes in a single analysis, this technology is considered to provide new potential tool in finding diagnostic biomarkers and molecular targets (5,6). Several early studies succeeded in identifying genes expressed specifically in CRC, and, subsequently, many researchers have focused on investigating the expression of messenger RNAs, which encode proteins (5,7-10).

In recent years, a large number of non-coding RNAs have been discovered. Although non-coding RNAs do not directly participate in protein synthesis, these RNAs have been demonstrated to be involved in gene regulation. Currently, non-coding RNAs are classified as various RNA species such

as microRNA, Piwi-interacting RNA and antisense RNAs (11-13). Among the non-coding RNA species, natural antisense transcripts (NATs) have been systematically identified in across mammalian species (14), and global transcriptome analysis shows that up to 70% of transcripts have antisense partners and that perturbation of NATs can alter the expression of the sense gene (15). Recently, NATs of *p15* have been discovered to regulate the expression of *p15* in leukemia cells through heterochromatin formation (16). Furthermore, the differential expression between normal and malignant breast tissues was observed for many sense and antisense pairs (17). However, a comprehensive NAT analysis using CRC samples has not been reported to date.

In the present study, the up- and down-regulation of NATs were investigated in human CRC for their possible involvement in CRC development. The expression profiles of NATs were determined using a custom microarray containing human sense/antisense probes for *ca*. 21,000 genes. Our objective here was to identify the up- and down-regulation of NATs in human CRC for their possible involvement in CRC development, and to explore biomarkers for CRC.

Materials and methods

Patients and samples. Surgical samples of 51 primary tumors, 9 corresponding adjacent non-cancerous colorectal tissues and 19 liver metastasis tumors were obtained from 68 CRC patients who underwent surgical resection from April 2006 to March 2009 at Tsukuba University Hospital (Tsukuba, Japan). None of the patients received radiation and/or chemotherapy before colorectal surgery. The main characteristics of the CRC cases are listed in Table I. Informed consent was obtained from all patients for the collection of specimens and the study protocol was approved by the hospital ethics committee. All samples were frozen in liquid nitrogen immediately after surgical resection and were stored at -80˚C until RNA extraction.

Total RNA extraction. Total RNA was isolated from frozen samples using Isogen reagent (Nippon Gene, Tokyo, Japan) according to the manufacturer's instructions. Quality and concentration of the RNA were assessed with the NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE) according to the manufacturer's instructions. All RNA samples indicated 260/280 nm absorbance ratios of 1.8-2.0. The integrity of the RNA was monitored by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and an RNA 6000 Nano LabChip kit (Agilent Technologies). Based on the instruction of the Bioanalyzer, total RNAs thus obtained were selected for further analysis; i.e., microarray analysis and strand-specific RT-qPCR.

Probe design of custom-microarray. One of the downstream research steps of microarray gene expression analysis is to investigate gene expression sites in tissues using *in situ* hybridization, for which probe size should be empirically ~120 nucleotides (nt) to obtain a satisfactory hybridization signal/noise ratio. Since the copy numbers of gene transcripts were shown to be different depending on the region of genes (18), probe sequences for microarray and for *in situ* hybridization should be selected in the same region of genes in order to interpret the results of microarray in the combination of those of *in situ* hybridization. Therefore, 120 nt sequences were first selected from human ORF sequences (Build35) for probe of *in situ* hybridization (Genetyx, Tokyo, Japan). The selected sequences were confirmed to be unique in the human genomic sequence by blast analysis, and were then submitted to Agilent server (Agilent Technologies) to design 60 nt sequences from 120 nt sequences for microarray probes. The sense and antisense sequences of 60 nt sequences thus designed were arranged in an Agilent 44 K x 4 system (20882 ORFs: Agilent eArray Design ID = 19052 produced by Tsukuba GeneTech Lab., Tsukuba, Japan) (Agilent Technologies).

Microarray analysis. Cyanine 3 (Cy3)-labeled cDNA was synthesized from 10 μg total RNA of CRC and non-cancerous samples using a LabelStar Array kit (Qiagen, Valencia, CA), Cy3-dUTP (GE Healthcare, Fairfield, CT), and random nonamer primer. Agilent 44 K x 4 human sense/antisense custom microarray slides described above were hybridized with the Cy3-labeled cDNA (2 μg) in a hybridization solution prepared with an *In Situ* Hybridization Kit Plus (Agilent Technologies), following the manufacturer's instructions. The Cy3 fluorescence signal images on the slides were obtained by a DNA microarray scanner (Agilent Technologies), and processed using the Feature Extraction version 8.1 software based on the instruction from Agilent Technologies.

Gene expression profiles of the samples were analyzed using GeneSpring GX10 software (Agilent Technologies). The expression data were normalized to the 75 percentile of all values on that microarray, followed by normalization of the median expression level of all samples. Gene expression data, when classified as either flag-'Present' or flag-'Marginal' in >70% of all samples, were loaded into the software.

The expression profiles of the samples were compared using unpaired t-tests (with Bonferroni FWER correction for unequal variances) as described in Results. Two-dimensional hierarchical clustering was performed for the log-transformed data using centroid-linkage and with euclidean correlation as the similarity measure. Variation in multigene expression was compared by principal component analysis (PCA).

Strand-specific RT-qPCR. Total RNAs were used for RT-qPCR of antisense RNAs. In order to normalize the values of antisense RNAs among samples, EGFP RNA was mixed in total RNAs as described below.

An aliquot of each RNA sample was mixed with an amount of RNA fragment (218 nucleotides) synthesized from pEGFP-C1 vector (Invitrogen, Carlsbad, CA) to attain a final amount of $5x10^{-5}$ pmol/10 μg total RNA (19). These RNA mixtures were subjected to synthesis of the first-strand cDNA only from antisense RNAs using forward primer (Table II), EGFP reverse primer, and Reverse Transcriptase (Promega, Madison, WI), according to the procedure recommended by Promega (20). Then, the mixtures were incubated at 55˚C for 60 min. The resulting cDNAs were incubated at 99˚C for 5 min and at 37˚C for 60 min with RNase A to digest RNA.

To confirm whether the fragments amplified in the qPCR were derived from the target sequences, the fragments amplified with primer pairs for qPCR were electrophoresed

Table I. Characteristics of human CRC cases analyzed by microarray analysis.

	Primary tumor			Metastasized liver tumor	
Sample group	PT-Hep(-)			PT-Hep(+)	Hep
AJCC stage	1	2	3	4	4
Number	8	14	16	13	19
Gender					
Male	5	4	11	7	14
Female	3	10	5	6	5
Median age, years (range)	66 (50-75)	65 (48-79)	66 (53-75)	65 (43-80)	65 (53-72)
Organ of primary tumor					
Colon	2	8	8	9	15
Rectum	6	6	8	4	4
pT UICC					
1	3	0	0	0	
2	5	0	3	1	
3	0	14	12	11	
4	0	0	1	1	
pN UICC					
0	8	14	0	4	
1	0	0	11	4	
2	0	0	5	5	
3	0	0	0	0	
pM UICC					
0	8	14	16	0	
1	0	0	0	13	
Liver metastasis					
(+)	0	0	0	13	
(-)	8	14	16	0	

AJCC, American Joint Committee on Cancer; UICC, International Union Against Cancer. pT, pathological staging of primary tumor; pN, pathological staging of regional lymph nodes; pM, pathological staging of distant metastasis; PT-Hep(-), primary tumor without liver metastasis; PT-Hep(+), primary tumor with liver metastasis; Hep, metastasized liver tumor.

Table II. Primer list for RT-qPCR.

Primer name	Sequence (5'-3')	Size (nt)	PCR products size (bp)
a primer (SLC26A3 forward primer)	TTCTGACGAAGAGCTGGACAAC	22	60
b primer (SLC26A3 reverse primer)	GTGGTATTGATTGGCTGGTCC	21	
c primer (IGJ forward primer)	AAATGTAAGTGTGCCCGGATTAC	23	60
d primer (IGJ reverse primer)	CTCATTAGGATCTTCGGAAGAAC	23	
e primer (LRRC24 forward primer)	TACGTTCGCACAGCTAGAGG	20	60
f primer (LRRC24 reverse primer)	TTGATGACGAACATCTCGTGGC	22	
EGFP forward primer	CAGCAGAACACCCCCATC	18	120
EGFP reverse primer	GAACTCCAGCAGGACCATGT	21	

Table III. The top 20 NATs differentially up- and down-regulated in CRC vs. non-cancerous tissues.

Accession no.	Gene symbol	Gene name	Fold change
NM_001005166.1	OR52E5	Olfactory receptor, family 52, subfamily E, member 5	13.521
NM_020408.3	C6orf149	Chromosome 6 open reading frame 149	12.726
NM_153010.3	C18orf16	Chromosome 18 open reading frame 16	12.162
NM_199350.2	C9orf50	Chromosome 9 open reading frame 50	10.797
NM_006920.3	SCN1A	Sodium channel, voltage-gated, type I, alpha	9.596
NM_001037234.1	TMEM75	Transmembrane protein 75	8.404
NM_144962.1	PEBP4	Phosphatidylethanolamine-binding protein 4	7.685
A_24_P922378	ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)	7.213
NM_153021.3	PLB1	Phospholipase B1	6.760
NM_199046.1	TEPP	Testis/prostate/placenta-expressed protein, isoform 2	6.411
NM_001037290.1	BCDO2	Beta-carotene dioxygenase 2	6.266
XR_018371.1	LOC648282	Similar to tropomyosin 3 isoform 2	6.227
NM_001029996.1	MGC33657	Similar to hypothetical protein	6.156
NM_181532.2	ERAS	ES cell expressed Ras	6.045
XM_001127227.1	LOC728331	Hypothetical protein LOC728331	5.974
NM_004413.1	DPEP1	Dipeptidase 1 (renal)	5.913
XM_001130346.1	LOC729468	Similar to phosphoglucomutase 5	5.888
XM_001133978.1	LOC732458	Hypothetical protein LOC732458	5.840
NM_033337.1	CAV3	Caveolin 3	5.836
NM_001024678.2	LRRC24	Leucine rich repeat containing 24	5.827
NM_000111.1	SLC26A3	Solute carrier family 26, member 3	-63.524
NM_144646.2	IGJ	Immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	-56.711
NM_001738.1	CA1	Carbonic anhydrase I	-41.308
NM_000669.3	ADH1C	Alcohol dehydrogenase 1C (class I), gamma polypeptide	-21.598
NM_019010.1	KRT20	Keratin 20	-21.428
XR_017975.1	C14orf139	Chromosome 14 open reading frame 139	-19.238
NM_000667.2	ADH1A	Alcohol dehydrogenase 1A (class I), alpha polypeptide	-18.712
NM_014479.2	ADAMDEC1	ADAM-like, decysin 1	-16.997
NM_032608.5	MYO18B	Myosin XVIIIB	-16.326
NM_014056.1	HIGD1A	HIG1 domain family, member 1A	-16.090
NM_001004124.1	OR4P4	Olfactory receptor, family 4, subfamily P, member 4	-16.032
XM_001129176.1	LOC731315	Hypothetical protein LOC731315	-15.746
NM_001040441.1	ZBTB8	Zinc finger and BTB domain containing 8	-15.650
NM_000067.1	CA2	Carbonic anhydrase II	-15.580
NM_014836.3	RHOBTB1	Rho-related BTB domain containing 1	-15.165
XM_001129054.1	LOC728362	Hypothetical protein LOC728362	-14.658
NM_152315.1	FAM55A	Family with sequence similarity 55, member A	-14.500
NM_000336.1	SCNN1B	Sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	-14.449
NM_022131.1	CLSTN2	Calsyntenin 2	-14.226
NM_138639.1	BCL2L12	BCL2-like 12 (proline rich)	-14.188

though 4% agarose gel to obtain their fragment size, and then subjected to direct sequence analysis using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA).

The first-strand cDNAs derived from antisense RNAs were then used as a template for quantitative PCR (qPCR) using SYBR Green Real-time PCR Master Mix Plus (Toyobo,

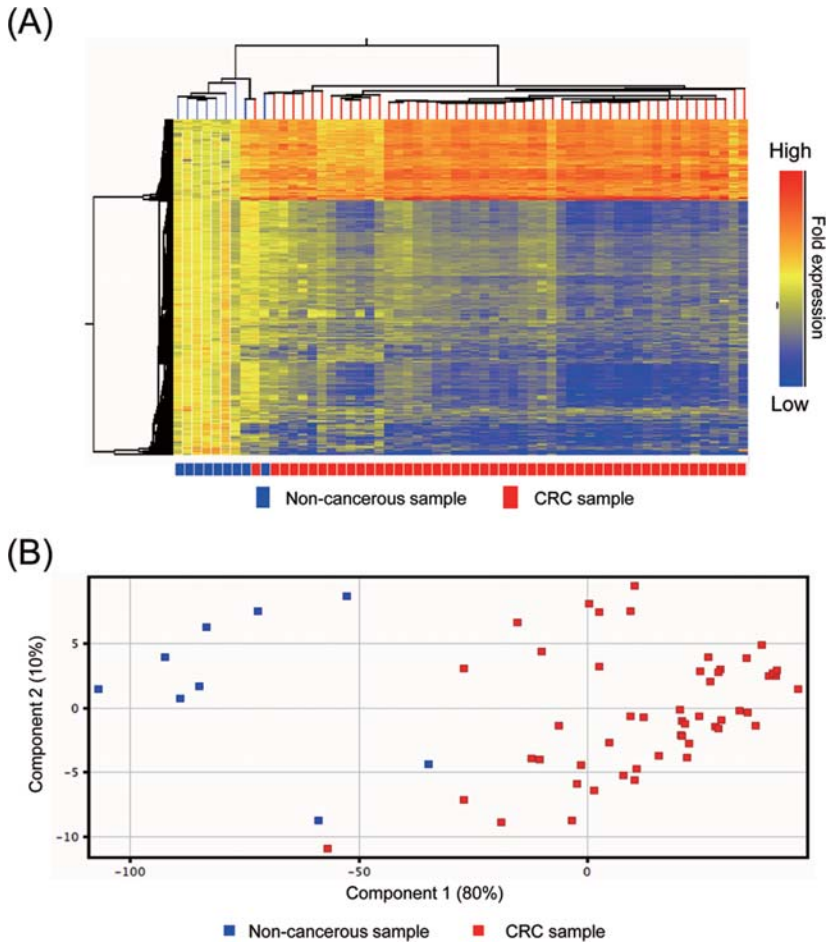

Figure 1. (A) Hierarchical cluster analysis with 415 up- and down-regulated NATs of CRC and non-cancerous tissue samples. Warm color gradient represents up-regulated NATs, and blue color gradient represents down-regulated NATs. A blue box at the bottom of the figure represents a non-cancerous sample; and a red box, a CRC sample. (B) Graphical presentation of principal component analysis (PCA) for 9 non-cancerous samples and 51 CRC samples using these 415 NATs based on PC1 (contribution ratio: 80%) and PC2 (contribution ratio: 10%). A blue box represents a non-cancerous sample; and a red box, a CRC sample.

Osaka, Japan) and primer pair sets described in Table II. The qPCRs were performed using an Applied Biosystem 7500 Real-Time PCR system (Applied Biosystems) under the conditions of 1 min at 95˚C followed by 40 cycles each of 95˚C for 15 sec and 60˚C for 60 sec, following the procedure recommended by the manufacturer. In order to compare the PCR results, the values for antisense RNAs were normalized based on values of EGFP. The results of qPCR are presented as means ± standard errors of means (SEM) of the samples.

Results

Gene expression profiling of up- and down-regulated NATs between the CRC and non-cancerous groups. Total RNAs isolated from 51 CRC tissues and 9 corresponding non-cancerous tissues (Table I) were subjected to custom microarray containing human sense/antisense probes. Microarray analysis was performed as described in Materials and methods. When gene expression profiles in the CRC and non-cancerous groups were compared, 863 sense/antisense transcripts were identified as those with amounts of which showed a >4-fold difference between the two groups using an unpaired t-test ($p<0.001$). Among these transcripts, sense transcripts numbered 448, and NATs numbered 415. Among the sense transcripts, 181 showed a higher expression (FC>4.0) and 267 showed a lower expression (FC<-4.0) in the CRC group than in the non-cancerous group (data not shown). These findings for sense transcripts are not inconsistent with several earlier studies (5,7-10). However, since NATs have not yet been reported in CRC development, the up- and down-regulated NATs were further investigated in the present study. In NATs, 101 showed a higher expression (FC >4.0) and 314 showed a lower expression (FC<-4.0) in the CRC group than in the non-cancerous group (data not shown, available on request). The top 20 up- and down-regulated NATs are listed in Table III.

A hierarchical cluster analysis was performed using individual values of the respective transcripts for samples. The results of the analysis are presented in Fig. 1A. The CRC and non-cancerous samples were clustered into two groups; i.e., 'CRC' and 'non-cancerous' groups except for two cases. When PCA with 415 NATs was performed, the CRC and non-cancerous samples were clearly separated using the first two principal components (contribution ratios of PC1 and PC2 were 80 and 10%, respectively) (Fig. 1B). These results demonstrate that there is a significant difference in 415 NATs

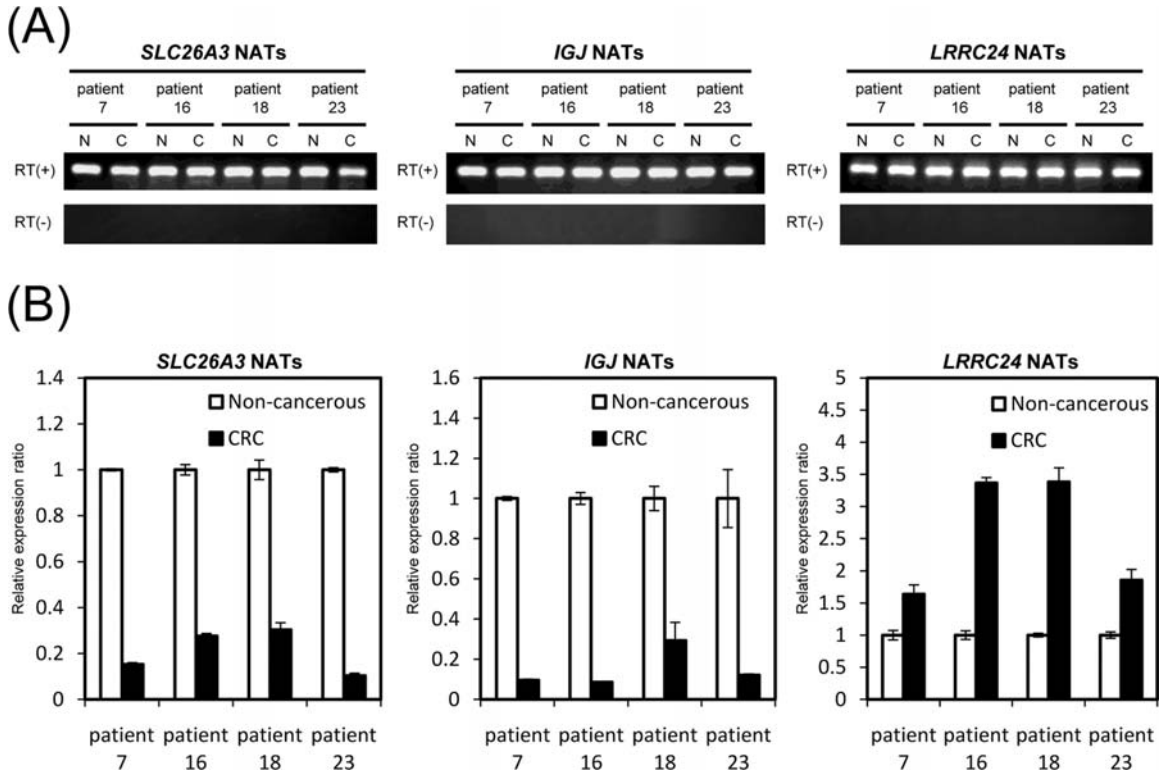

Figure 2. (A) Agarose gel electrophoresis of *SLC26A3, IGJ* and *LRRC24* DNA fragments amplified by RT-qPCR. (B) Determination of NAT amounts of *SLC26A3, IGJ* and *LRRC24*. The NATs expression levels of *SLC26A3, IGJ* and *LRRC24* were determined using a strand-specific RT-qPCR analysis of samples of patients 7, 16, 18, and 23.

between the CRC and non-cancerous groups, and infer that a small number of transcripts would contribute to differentiate the cancerous and non-cancerous state.

RT-qPCR analysis of NATs in the CRC and non-cancerous samples. OR52E5, C6orf149, ERG, PLB1, LRRC24, SLC26A3, and *IGJ* were randomly selected from the genes shown in Table III, and subjected to validation of the NATs amounts obtained in the microarray with RT-qPCR. Studies on NATs, including those for mice, have shown that NATs have been characterized with various initiation sites and termination sites. Therefore, NATs of the above selected genes were first examined as to whether the primer pairs designed produced unique-sized fragments in the RT-qPCR using total RNA of the one patient (patient 7). The examination revealed that *LRRC24, SLC26A3*, and *IGJ* produced unique-sized fragments; while *OR52E5, C6orf149, ERG*, and *PLB1* produced multiple-sized fragments (data not shown). The DNA fragments produced for NATs of *LRRC24, SLC26A3*, and *IGJ* were sequenced to confirm that the DNA fragments were derived from the respective NATs. The NATs of *LRRC24, SLC26A3*, and *IGJ* were then subjected to RT-qPCR to determine their amounts in the total RNA samples of randomly selected four patients (patients 7, 16, 18, and 23). As shown in the results of the RT-qPCR analysis in Fig. 2, NAT amounts of *SLC26A3*, and *IGJ* were smaller in the CRC than in the non-cancerous tissues of the patients, whereas the NAT amount of *LRRC24* was larger in CRC than in the non-cancerous tissues. Although the NAT amount ratios of the CRC to non-cancerous tissues obtained by microarray analysis was consistently higher than those obtained by RT-qPCR, the observations obtained in the RT-qPCR were consistent with those in the microarray. Based on the fact that NATs were characterized with various initiation sites and termination sites, it might be inferred that some portion of the NAT did not encompass both primer sequences used for RT-qPCR.

Consequently, although the number of NATs and patients applied for comparison between the microarray and RT-qPCR analyses was limited, it was judged that the results of the microarray analysis would reflect the NAT expression profiles in CRC and non-cancerous tissues.

Identification of NATs that distinguish primary tumors with liver metastasis from primary tumors without liver metastasis. In order to examine NATs profiles in liver metastasis, we investigated NATs expression profiles between primary tumors without liver metastasis [PT-Hep(-)] and primary tumors with liver metastasis [PT-Hep(+)]. Two hundred and fifty-six NATs were identified as the transcripts, amounts of which showed a >2-fold difference between the two groups using an unpaired t-test ($p<0.05$). Of these NATs, 226 showed a higher expression (FC >2.0), and 30 showed a lower expression (FC <-2.0) in PT-Hep(+) group than in PT-Hep(-) group (data not shown, available on request). The top 10 up- and down-regulated NATs are listed in Table IV. When a hierarchical cluster analysis was performed using individual values of the respective NATs for samples, no significant clustering was obtained in the viewpoint of metastasis of CRC (data not shown). When PCA with 226 NATs was performed, the CRC and non-cancerous samples were not clearly separated using

Table IV. The top 10 NATs differentially up- and down-regulated in primary tumors with liver metastasis vs. primary tumors without liver metastasis.

Accession no.	Gene symbol	Gene name	Fold change
NM_133266.1	SHANK2	SH3 and multiple ankyrin repeat domains 2	4.841
NM_006821.3	ACOT2	Acyl-CoA thioesterase 2	3.685
NM_006030.2	CACNA2D2	Calcium channel, voltage-dependent, alpha 2/delta subunit 2	3.452
NM_016200.3	LSM8	LSM8 homolog, U6 small nuclear RNA associated (*S. cerevisiae*)	3.338
NM_001004480.1	OR11H6	Olfactory receptor, family 11, subfamily H, member 6	3.222
NM_007036.3	ESM1	Endothelial cell-specific molecule 1	3.199
NM_001040441.1	ZBTB8	Zinc finger and BTB domain containing 8	2.997
XM_936719.2	LOC647662	Similar to WW domain binding protein 11	2.973
NM_138700.2	TRIM40	Tripartite motif-containing 40	2.949
NM_005542.3	INSIG1	Insulin induced gene 1	2.908
NM_033661.3	WDR4	WD repeat domain 4	-3.854
NM_023935.1	C20orf116	Chromosome 20 open reading frame 116	-3.145
NM_005023.2	PGGT1B	Protein geranylgeranyltransferase type I, beta subunit	-2.847
XM_001133434.1	LOC732422	Hypothetical protein LOC732422	-2.743
NM_021946.2	BCORL1	BCL6 co-repressor-like 1	-2.708
XR_019353.1	LOC650267	Similar to double homeobox, 4	-2.700
NM_000413.1	HSD17B1	Hydroxysteroid (17-beta) dehydrogenase 1	-2.627
NM_015015.1	JMJD2B	Jumonji domain containing 2B	-2.397
NM_001009812.1	LBX2	Ladybird homeobox homolog 2 (*Drosophila*)	-2.391
NM_006442.2	DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	-2.361

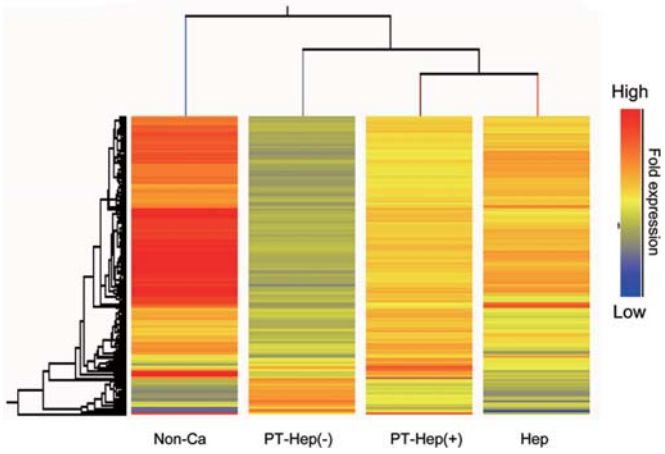

Figure 3. Hierarchical cluster analysis using mean values of the individual NATs that distinguished PT-Hep(+) from PT-Hep(-). Warm color gradient represents up-regulated NATs, and blue color gradient represents down-regulated NATs.

the first two principal components (contribution ratios of PC1 and PC2 were 68 and 10%, respectively) (data not shown). Then, additional hierarchical cluster analysis was performed using mean values of the individual transcripts in the respective groups, revealing two major nodes (Fig. 3): non-cancerous and cancerous. The cancerous node was branched off into the node of PT-Hep(-) and PT-Hep(+) clustered with its metastasized liver tumor (Hep).

Discussion

A large number of studies have been performed on the difference of gene expressions between cancerous and non-cancerous tissues to obtain biomarkers for diagnosis of cancer as well as to understand the cancer development mechanism. However, the vast majority of these studies has focused on the mRNA of genes (sense transcripts), with only a few focusing on antisense transcript in human clinical cancer tissues (17,21). In the present study, we performed, for the first time, a comprehensive analysis of NATs in the human CRC tissues and non-cancerous tissues, seeking biomarkers for CRC and a clue to understand CRC development.

By comparing gene expression profiles of CRC tissues with non-cancerous tissues, we identified 415 NATs differentially expressed between the CRC and non-cancerous tissues to a significant degree. Hierarchical clustering analysis and PCA clearly distinguished the gene profiles of CRC from those of non-cancerous tissues using these 415 NATs. Strand-specific RT-qPCRs of 3 selected NATs (*SLC26A3, IGJ, LRRC24*) have validated the results of microarray. As described in Results, 4 of the 7 NATs selected for the validation failed to give single fragments in the RT-qPCR. The reason for this failure could be interpreted as follows. The annealing

temperature of the primers to RNA in RT was not able to be raised to the optimum due to the nature of the reverse transcriptase, which allowed the amplification of fragments from RNA species other than the target species. The absolute NAT amount ratios of CRC to non-cancerous tissues for the 3 genes examined in the microarray were consistently larger than those in RT-qPCR. It might be inferred that since NATs have been demonstrated to have various transcription initiation and termination sites (20,22), a limited number of NAT molecules were reverse transcribed to cDNA using the primer pairs listed in Table II. To verify this inference, a large number of RNA molecules should be randomly sequenced to identify of the NATs related to the 3 genes using a next-gen sequencer.

We also identified 256 NATs whose amounts differed significantly between PT-Hep(+) and PT-Hep(-). However, only the hierarchical clustering using mean values of the individual transcripts in the respective groups could separate PT-Hep(+) and PT-Hep(-) as different groups, but other analyses including PCA provided unclear separation. These results may indicate that events leading to metastasis of the liver may not significantly affect the amounts. Since, in the present study, NAT amounts were measured for one region per gene, and since NAT amounts were indicated to be different depending on the region of a gene (18), NATs of regions other than those examined might clearly distinguish the CRC with respect to metastasis.

Asymmetric strand specific analysis of gene expression was applied on 5 different human cell types to reveal the expression of NATs from 2900 to 6400 genes, suggesting that they are a fundamental component of gene regulation (23). Interestingly, knockdown or blockade of NATs can have multiple outcomes, with the corresponding sense transcript concentration showing either an increase (discordant regulation) or a decrease (concordant regulation) (24). These variable intrinsic properties indicate that antisense-mediated regulation of gene expression must operate through a variety of mechanisms, and further suggest that NATs are a heterogeneous group of regulatory RNAs (25). Yu *et al* found that leukemia cells had larger amounts of *p15* NAT and smaller amounts of *p15* mRNA than normal lymphocytes. Furthermore, they demonstrated that an antisense of *p15* expression construct induced *p15* silencing, which has been frequently observed in leukemia (16). In addition, Grigoriadis *et al* (17) reported that differential expressions between normal and malignant breast samples were observed for many sense and antisense pairs. Taking these observations together with our findings that 415 NATs were expressed differentially in CRC and non-cancerous tissues to a significant degree, NATs are strongly indicated to be involved in various types of cancer, including CRC. In the future, a comprehensive investigation should be performed for all types of cancer to obtain NAT catalogs for various cancers and stages of cancers. Such a catalog could be the basis for finding molecular markers of precancerous stages as well as for understanding cancer development.

Acknowledgements

This study was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J and Thun MJ: Cancer statistics, 2009. CA Cancer J Clin 59: 225-249, 2009.
2. Tsukuma H, Ajiki W and Oshima A: [Cancer incidence in Japan]. Gan To Kagaku Ryoho 31: 840-846, 2004.
3. Koehler A, Bataille F, Schmid C, *et al*: Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. J Pathol 204: 65-74, 2004.
4. Conlin A, Smith G, Carey FA, Wolf CR and Steele RJ: The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. Gut 54: 1283-1286, 2005.
5. Croner RS, Foertsch T, Brueckl WM, *et al*: Common denominator genes that distinguish colorectal carcinoma from normal mucosa. Int J Colorectal Dis 20: 353-362, 2005.
6. Raetz EA and Moos PJ: Impact of microarray technology in clinical oncology. Cancer Invest 22: 312-320, 2004.
7. Ohmachi T, Tanaka F, Mimori K, Inoue H, Yanaga K and Mori M: Clinical significance of TROP2 expression in colorectal cancer. Clin Cancer Res 12: 3057-3063, 2006.
8. Bertucci F, Salas S, Eysteries S, *et al*: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 23: 1377-1391, 2004.
9. Bianchini M, Levy E, Zucchini C, *et al*: Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. Int J Oncol 29: 83-94, 2006.
10. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, *et al*: Gene expression in colorectal cancer. Cancer Res 62: 4352-4363, 2002.
11. Hannon GJ: RNA interference. Nature 418: 244-251, 2002.
12. Novina CD and Sharp PA: The RNAi revolution. Nature 430: 161-164, 2004.
13. Willingham AT and Gingeras TR: TUF love for 'junk' DNA. Cell 125: 1215-1220, 2006.
14. Rosok O and Sioud M: Systematic identification of sense-antisense transcripts in mammalian cells. Nat Biotechnol 22: 104-108, 2004.
15. Katayama S, Tomaru Y, Kasukawa T, *et al*: Antisense transcription in the mammalian transcriptome. Science 309: 1564-1566, 2005.
16. Yu W, Gius D, Onyango P, *et al*: Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 451: 202-206, 2008.
17. Grigoriadis A, Oliver GR, Tanney A, *et al*: Identification of differentially expressed sense and antisense transcript pairs in breast epithelial tissues. BMC Genomics 10: 324, 2009.
18. Clark TA, Schweitzer AC, Chen TX, *et al*: Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol 8: R64, 2007.
19. Kiuchi S, Yamada T, Kiyokawa N, Saito T, Fujimoto J and Yasue H: Genomic structure of swine taste receptor family 1 member 3, TAS1R3, and its expression in tissues. Cytogenet Genome Res 115: 51-61, 2006.
20. Chiba M, Kiyosawa H, Hiraiwa N, Ohkohchi N and Yasue H: Existence of Pink1 antisense RNAs in mouse and their localization. Cytogenet Genome Res 126: 259-270, 2009.
21. Reis EM, Nakaya HI, Louro R, *et al*: Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. Oncogene 23: 6684-6692, 2004.
22. Kiyosawa H, Mise N, Iwase S, Hayashizaki Y and Abe K: Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res 15: 463-474, 2005.
23. He Y, Vogelstein B, Velculescu VE, Papadopoulos N and Kinzler KW: The antisense transcriptomes of human cells. Science 322: 1855-1857, 2008.
24. Wahlestedt C: Natural antisense and noncoding RNA transcripts as potential drug targets. Drug Discov Today 11: 503-508, 2006.
25. Faghihi MA and Wahlestedt C: Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10: 637-643, 2009.