INTERNATIONAL JOURNAL OF ONCOLOGY 31: 1029-1037, 2007

Expression profiling of fecal colonocytes for RNA-based screening of colorectal cancer

SATOSHI YAJIMA[1,2], MIE ISHII[3], HISAYUKI MATSUSHITA[4], KAZUHIKO AOYAGI[1], KAZUHIKO YOSHIMATSU[5], HIRONORI KANEKO[2], NOBUKO YAMAMOTO[3], TATSUO TERAMOTO[2], TERUHIKO YOSHIDA[1], YASUHIRO MATSUMURA[4] and HIROKI SASAKI[1]

[1]Genetics Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045; [2]Division of General and Gastroenterological Surgery (Omori), Department of Surgery, School of Medicine, Faculty of Medicine, Toho University, Omori Nishi 6-11-1, Ohta-ku, Tokyo 143-8541; [3]Medical Engineering Development Center, Canon Inc., Shimomaruko 3-30-2, Ohta-ku, Tokyo 146-8501; [4]Investigate Treatment Division, Research Center for Innovative Oncology, National Cancer Center Hospital East, Kashiwanoha 6-5-1, Kashiwa, Chiba 277-8577; [5]Medical Center East, Tokyo Women's Medical University, School of Medicine, Nishiogu 2-1-10, Arakawa-ku, Tokyo 116-8567, Japan

Received May 21, 2007; Accepted July 18, 2007

Abstract. The early detection of colorectal cancer originating from any part of the colorectum is desirable because this cancer can be cured surgically if diagnosed early. We searched for marker genes for a fecal RNA-based colorectal cancer screening method by comparison of genome-wide expression profiles among cancerous and non-cancerous tissues, and healthy volunteer- and cancer patient-derived colonocytes from the feces, and the peripheral blood. Of 14,564 genes, only 3 (PAP, REG1A, and DPEP1) were selectable as final candidates which were expressed frequently at any stage of this cancer and were suppressed in non-cancerous tissues and also in the peripheral blood and colonocytes of healthy volunteers. Next, we directly compared fecal RNA-expression profiles between colorectal cancer patients and healthy volunteers, and found that most of the genes (92%) expressed in the colonocytes of the cancer patients were not expressed in those of the healthy volunteers. Six genes (SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11) selected randomly from 85 cancer patient-derived colonocyte-specific genes were evaluated. In total, reverse transcription-polymerase chain reaction or focused microarray of all those 9 genes detected 18 (78%) of 23 curable colorectal cancers (Dukes stages A-C), 9 or 10 (64% or 71%) of 14 early cancers with no lymph node metastasis (Dukes stage A or B) and 4 (80%) of 5 right-sided cancers. Our extensive gene list provides other markers for fecal RNA-based colorectal cancer screening.

Correspondence to: Dr Hiroki Sasaki, Genetics Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, Japan
E-mail: hksasaki@gan2.res.ncc.go.jp

Key words: expression profiling, colonocyte, colorectal cancer screening, microarray

Introduction

Colorectal cancer is a common malignancy which is curable by surgical resection if diagnosed at a sufficiently early stage (stage I/Dukes stage A or stage II/Dukes stage B). Five-year survival rates on surgical resection, for example, at Dukes stage A, Dukes stage B and stage III/Dukes stage C are 95%, 80% and 50-60%, respectively. For stage IV/Dukes stage D, curative resection is impossible. Therefore, early detection of this cancer originating from any part of the colorectum is desired. For mass cancer screenings, a simple, economic, and noninvasive method of cancer detection is required. The Hemoccult test is currently used in many countries for this purpose (1-5). However, this test is nonspecific and is not sufficiently sensitive to detect early-stage cancer, although a higher sensitivity has been reported for the advanced stage (6).

For fecal DNA-based colorectal cancer screening, in 1992, Sidransky first reported Ras oncogene mutations in the fecal DNA of patients with curable colorectal cancer (7). To date, many screening methods based on mutated DNA detection in the feces have been reported (8-19). These methods, however, are time-consuming and are not sufficiently sensitive. The major reason for this inaccuracy is the fact that fecal DNAs are derived from an enormous number and variety of bacteria and normal living cells including normal colorectal mucus cells, lymphocytes, red blood cells and anal squamous cells. Immunocytochemical analysis provides a simple method; however, this method is insensitive because only the surface portion of the feces can be assayed. On the other hand, Tarin and colleagues first reported that cancer-specific CD44 splicing variants are useful for fecal RNA-based colorectal cancer screening (20,21). By the use of the repetition of the Percoll centrifugation method for isolating the colonocytes from feces, we have also demonstrated that unusual CD44 variants could be targets for cancer-detection using feces (22). However, the method is found to distort the morphology of

colonocytes and to have a low retrieval rate. Accordingly, the sensitivity of this mRNA-based method also appears to be insensitive.

In any method for colorectal cancer detection using feces, an effective method which allows the simple isolation of the colonocytes from not only the surface but also the central portion of the feces while maintaining the initial morphology is needed. Recently, we successfully developed a new, very effective method that is based on the filtration of the homogenates of feces and magnetic cell sorting (MACS) with an epithelial cell-specific antibody, which we here abbreviated to FMCI (filtration and MACS-based colonocyte isolation) (23). It has been shown that this method can provide a high quality of colonocyte DNA or RNA for molecular biological analysis and also provide the colonocyte with its original morphology for cytology. Considering the advantage in the use of FMCI, we here report expression profiling of colonocytes for fecal RNA-based detection of curable colorectal cancer and a sensitive focused microarray assay that uses multiple marker genes for detecting minimal cancer cells in the feces of patients with colorectal cancer.

Materials and methods

Clinical materials. This study protocol was reviewed and approved by the Institutional Review Board of the National Cancer Center, Tokyo. Written informed consent was obtained from all the patients and healthy volunteers. Before surgical resection, stool samples were obtained from 23 patients with colorectal cancer (Dukes stages A-C), for which curable resection is possible, and from 15 healthy volunteers a few weeks after they had undergone a total colonoscopy. Naturally evacuated feces from subjects who had not taken laxatives were used as stool samples. Each patient was instructed to evacuate into a polystyrene disposable tray (AS one, Osaka, Japan) measuring 5x10 cm in size. Preparation of the stool samples for examination was conducted within 1-6 h after the evacuation. Tissue samples were obtained from the surgically resected specimens of colorectal cancer patients, and were snap frozen in liquid nitrogen and stored at -80˚C until use. RNA of the tissues was extracted by using an Isogen kit (Nippon Gene, Toyama, Japan). The peripheral blood samples of 58 healthy volunteers and their RNA were prepared as in our previous report (24).

Isolation of exfoliated cells from feces. The procedure is detailed in our previous report (23). In brief, approximately 5-10 g of feces was used to isolate exfoliated cells. Feces were collected in Stomacher Lab Blender bags (Seward, Thetford, UK). The stool samples were homogenized with a buffer (200 ml) consisting of Hank's solution, 25 mM Hepes (pH 7.35), and 10% fetal bovine serum at 200 rpm for 1 min using a Stomacher (Seward). The homogenates were then filtered through a nylon filter (pore size, 512 μm), followed by division into 5 portions (40 ml each). Subsequently, 40 μl of magnetic beads coated with a mouse IgG1 monoclonal antibody (mAb Ber-EP4) specific for the glycopolypeptide membrane antigen Ep-CAM, which is expressed on most normal and neoplastic human epithelial cells (Dynal, Oslo, Norway), was added to each portion, and the mixtures were incubated for 30 min under gentle rolling in a mixer at room temperature. After 15-min shaking, the colonocytes were recovered from 5 tubes. The colonocytes from a single tube were stored at -80˚C for RNA extraction. The colonocyte RNA was extracted by using an Isogen kit (Nippon Gene, Toyama, Japan).

Microarray analysis. We used human U133A Gene Chip (Affymetrix, Santa Clara, CA) for genome-wide expression profiling of mRNAs corresponding to 14,564 genes, 18,445 transcripts including splicing variants, and 22,215 probe sets. The procedures were conducted according to the supplier's protocols. Briefly, 10 μg of fragmented cRNA was hybridized to the microarrays in 200 μl of a hybridization cocktail at 45˚C for 16 h in a rotisserie oven set at 60 rpm. The arrays were then washed with a nonstringent wash buffer (6X SSPE) at 25˚C, followed by a stringent wash buffer [100 mM MES (pH 6.7), 0.1 M NaCl, and 0.01% Tween-20] at 50˚C, stained with streptavidin phycoerythrin (Invitrogen, Carlsbad, CA), washed again with 6X SSPE, stained with biotinylated anti-streptavidin IgG, followed by a second staining with streptavidin phycoerythrin and a third wash with 6X SSPE. The arrays were scanned using a GeneArray scanner (Affymetrix) at 3-μm resolution, and the expression value (average difference: AD) of each gene was calculated using GeneChip Analysis Suite version 5.0 software (Affymetrix). The mean of AD values in each experiment was 1000 to reliably compare variable multiple arrays.

Reverse transcription-polymerase chain reaction (RT-PCR). RT-PCR on colonocyte RNA was carried out using primer sets designed for detecting the 3' side of cDNA of each gene. Primers were 5'-ACCAGTGTGAGGACTCACCC-3' and 5'-TGCTCTTTAAAGCCTTAGGCC-3' for PAP; 5'-AGCAAT TACAACGGAGTCAA-3' and 5'-TCCAAAGACTGGGGT AGGT-3' for REGIA; 5'-TCTCTCCTGTGAAACCTGGG-3' and 5'-AAGGGGTGTTGCTTTTATTGC-3' for DPEP1; 5'-ATTAGCAGTTTAGAATGGAGG-3' and 5'-CTGTATCCA ATTCTGTACTGC-3' for SEPP1; 5'-TGGGCTGCCAACAT GCCATC-3' and 5'-TGTAGTAGCCCGATCGCACC-3' for RPL27A; 5'-GGCAAGCGAGATGAAGATAAGG-3' and 5'-AGGTCCCATACGTATGACAG-3' for ATP1B1; 5'-AGAC TATCCACCTTTGGGTCG-3' and 5'-GATGCATTGTTATC ATTAACCAGTC-3' for EEF1A; 5'-TTGAGCGCACCTAA CCACTGGT-3' and 5'-GAGAGGAAACATGGTCACACC CA-3' for SFN, and 5'-ACATTCAGACTGAGCGTGCCTA-3' and 5'-GATCTGGACGTCCCTGAAGCA-3' for RPS11. PCR was performed under conditions of 30-35 cycles of 3 steps of temperature, 95˚C for 1 min, 55˚C for 1 min, and 72˚C for 1 min, using the AccuPrime TaqDNA polymerase system (Invitrogen).

Marker gene detection using focused microarray. A focused microarray was constructed by fixing 50-60 mer of oligonucleotide probes on a slide glass using our previously developed Bubble Jet Technology (25). The microarray contained a single spot for each sequence of 9 marker genes (PAP, REG1A, DPEP1, SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11) and a control artificial DNA sequence. Each probe sequence used for the microarray is listed in Table I.

Table I. Sequences of primers and probes for focused microarray analysis.

Gene	Forward primers Reverse primers	Probes
PAP	GAGAAGCACAGCATTTCTGAG TGCTCTTTAAAGCCTTAGGCC	TTCCCCCAACCTGACCACCTCATTCTTATCTTTCTTCTGT TTCTTCCTCCCCGCTGTCAT
REG1A	AATCCTGGCTACTGTGTGAG TCCAAAGACTGGGGTAGGT	GACCATCTCTCCAACTCAACTCAACCTGGACACTCTCTT CTCTGCTGAGTTTGCCTTGTT
DPEP1	ACCCATTACGGCTACTCCTC AAGGGGTGTTGCTTTTATTGC	CAGATGCCAGGAGCCCTGCTGCCCACATGCAAGGACCA GCATCTCCTGAGAG
SEPP1	AATTAGCAGTTTAGAATGGAGG CTGTATCCAATTCTGTACTGC	CCATAGTCAATGATGGTTTAATAGGTAAACCAAACCCTA TAAACCTGACCTCCTTTATGG
RPL27A	TGGGCTGCCAACATGCCATC TGTAGTAGCCCGATCGCACC	CCAACTGTCAACCTTGACAAATTGTGGACTTTGGTCAGT GAACAGACACGGGTGAATGCT
ATP1B1	GGCAAGCGAGATGAAGATAAGG AGGTCCCATACGTATGACAG	GAGTGTAAGGCGTACGGTGAGAACATTGGGTACAGTGA GAAAGACCGTTTTCAGGGACGT
EEF1A1	AGACTATCCACCTTTGGGTCG GATGCATTGTTATCATTAACCAGTC	CCACCCCACTCTTAATCAGTGGTGGAAGAACGGTCTCAG AACTGTTTGTTTCAATTGGCC
SFN	TTGAGCGCACCTAACCACTGGT GAGAGGAAACATGGTCACACCCA	CTCTGATCGTAGGAATTGAGGAGTGTCCCGCCTTGTGGC TGAGAACTGGACAGTGG
RPS11	ACATTCAGACTGAGCGTGCCTA GATCTGGACGTCCCTGAAGCA	TCATCCGCCGAGACTATCTGCACTACATCCGCAAGTACA ACCGCTTCGAGAAGCG

Focused microarray analysis consists of 3 steps: i) Cy3-dUTP labeling by multiplex-RT-PCR; ii) hybridization Cy3-labeled cDNA to microarray, and iii) fluorescence scanning (Fig. 3). Using 0.5 to 1 μg of total RNA prepared from colonocytes, reverse transcription was performed with Superscript II (Invitrogen) with T7-oligo dT 24 primer in a total volume of 20 μl according to the manufacturer's protocol. To obtain 5-10 μg of cRNA, T7 transcription was performed. Using 5-10 μg of the cRNA, reverse transcription was performed with Superscript II with random hexamer in a total volume of 20 μl. Multiplex-RT-PCR was performed in two tubes at different PCR cycles: 35 cycles for PAP, REG1A and DPEP1, and 25 cycles for SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11. PCR primer sequences are also listed in supplementary Table II. Twenty-five μl of the PCR solution in each tube consisted of 1 μl of template cDNA, primers (6.25 pmol each), 50 μM Cy3-dUTP, 2.5 μl of AccuPrime 10X buffer1 (2 mM dNTP, 15 mM $MgCl_2$) and 1.0 μl of AccuPrime Taq polymerase (Invitrogen). A thermal cycler was set with initial heating at 95˚C for 5 min followed by an amplification cycle heated at 95˚C for 30 sec, 58˚C for 30 sec and 72˚C for 40 sec, followed by heating at 72˚C for 10 min. The two PCR solutions were mixed and purified with a QIAquick PCR purification kit (Qiagen, Tokyo, Japan). The entire Cy3-labeled cDNA solution (50 μl) was mixed in 120 μl of a hybridization cocktail (6X SSPE containing 900 mM NaCl, 60 mM NaH_2PO_4, and H_2O, and 6 mM EDTA, pH 7.4/ 10% formamide/0.05% SDS) including 0.1 nM Cy3-labeled oligonucleotide which hybridizes the control artificial DNA sequence. By using a hybridization apparatus, HybStation (Genomic Solutions, Ann Arbor, MI), an array was pre-heated to 65˚C for 3 min, filled with the hybridization cocktail, and incubated at 92˚C for 2 min and then at 55˚C for 4 h. Subsequently, the array was washed with 2X SSC, 0.1% SDS at 25˚C and then with 2X SSC at 20˚C, and rinsed with 0.1X SSC in accordance with a conventional manual, and finally dried by a spin drier. The array was scanned by an apparatus for DNA microarrays, Genepix 4000B (Axon Instruments, Union City, CA) and the fluorescence intensity from each probe spot was obtained after subtracting the background level. This focused microarray assay belongs in a negative or positive assay. However, it is required for determining the cutoff values. In this study, the maximum value of each gene plus 2- or 3-times standard deviation in 7 healthy volunteers was used as the cutoff-value.

Results

Marker gene selection through genome-wide expression profiles of cancer tissues, non-cancerous tissues, and the peripheral blood. In the feces of colorectal cancer patients, living cells other than bacteria include a small amount of cancer cells and normal colorectal mucus cells, lymphocytes, red blood cells and anal squamous cells. It is noted that the content of lymphocytes and red blood cells is increased in the feces of people with hemorrhoids. Therefore, genes that are expressed in almost all cases of early and advanced colorectal cancer and that are not expressed in normal colorectal mucosas,

Table II. Eighty-five genes expressed in the cancer patient-derived colonocytes but not in the healthy volunteer-derived colonocytes.

No.	Gene symbol	Gene title	Entrez gene ID	No expression in the PB
1	JUND	jun D proto-oncogene	3727	
2	TPT1	tumor protein, translationally-controlled 1	7178	*
3	RPL41	ribosomal protein L41	6171	*
4	RPS11	ribosomal protein S11	6205	*
5	RPS29	ribosomal protein S29	6235	*
6	RPL38	ribosomal protein L38	6169	*
7	SEPP1	selenoprotein P, plasma, 1	6414	*
8	RPL23	ribosomal protein L23	9349	*
9	B2M	ß-2-microglobulin	567	*
10	CFL1	cofilin 1 (non-muscle)	1072	*
11	RPL31	ribosomal protein L31	6160	*
12	RPS3A	ribosomal protein S3A	6189	
13	TMSB10	thymosin, ß 10	9168	*
14	RPL39	ribosomal protein L39	6170	*
15	HMGB1	high-mobility group box 1	3146	*
16	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	4680	
17	ATP1B1	ATPase, Na^+/K^+ transporting, ß 1 polypeptide	481	*
18	RPS20	ribosomal protein S20	6224	*
19	ARF6	ADP-ribosylation factor 6	382	*
20	RPS21	ribosomal protein S21	6227	*
21	EIF5A	Eukaryotic translation initiation factor 5A	1984	*
22	RPL30	ribosomal protein L30	6156	*
23	EEF1A1	eukaryotic translation elongation factor 1 α 1	1915	*
24	RPL23A	ribosomal protein L23a	6147	*
25	LOC56902	putatative 28 kDa protein	56902	
26	RPL27	ribosomal protein L27	6155	*
27	SFN	stratifin	2810	*
28	CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	1048	*
29	RPS24	ribosomal protein S24 /// ribosomal protein S24	6229	*
30	MARCKS	Myristoylated alanine-rich protein kinase C substrate	4082	*
31	PDE4C	phosphodiesterase 4C, cAMP-specific (phosphodiesterase E1 dunce homolog, *Drosophila*)	5143	
32	LOC651423	similar to mitogen-activated protein kinase kinase 3 isoform A	651423	
33	RPS10	ribosomal protein S10	6204	
34	CEP27	centrosomal protein 27 kDa	55142	
35	IL1RN	interleukin 1 receptor antagonist	3557	*
36	SLC35E1	solute carrier family 35, member E1	79939	
37	RPS27	ribosomal protein S27 (metallopanstimulin 1)	6232	*
38	RPS19	ribosomal protein S19	6223	*
39	RPS16	ribosomal protein S16	6217	*
40	MORF4L2	mortality factor 4 like 2	9643	*
41	RPL22	ribosomal protein L22	6146	*
42	RPS2	ribosomal protein S2	6187	*
43	RPLP2	ribosomal protein, large, P2	6181	*
44	RPL7A	ribosomal protein L7a	6130	
45	RPL7	ribosomal protein L7	6129	
46	RPS18	ribosomal protein S18	6222	*
47	HNRPH1	Heterogeneous nuclear ribonucleoprotein H1 (H)	3187	*
48	ZNF160	zinc finger protein 160	90338	

Table II. Continued.

No.	Gene symbol	Gene title	Entrez gene ID	No expression in the PB
49	RPS25	ribosomal protein S25	6230	*
50	PGF	Placental growth factor, vascular endothelial growth factor-related protein	5228	
51	SPG21	spastic paraplegia 21 (autosomal recessive, Mast syndrome)	51324	
52	RPL9	ribosomal protein L9	6133	*
53	PLEKHA5	Pleckstrin homology domain containing, family A member 5	54477	
54	PRR11	proline rich 11	55771	
55	CTNNB1	catenin (cadherin-associated protein), ß 1, 88 kDa	1499	*
56	NFKBIA	nuclear factor of κ light polypeptide gene enhancer in B-cells inhibitor, α	4792	*
57	GTSE1	G-2 and S-phase expressed 1	51512	
58	ATP8B1	ATPase, Class I, type 8B, member 1	5205	
59	TMED2	transmembrane emp24 domain trafficking protein 2	10959	*
60	RPS4X	ribosomal protein S4, X-linked	6191	
61	MUC3B	mucin 3B, cell surface associated	57876	
62	TTLL12	tubulin tyrosine ligase-like family, member 12	23170	
63	FTL	ferritin, light polypeptide	2512	*
64	TSPAN13	Tetraspanin 13	27075	*
65	PTP4A2	protein tyrosine phosphatase type IVA, member 2	8073	*
66	EGLN3	egl nine homolog 3 (C. elegans)	112399	*
67	ROCK2	Rho-associated, coiled-coil containing protein kinase 2	9475	
68	NDRG1	N-myc downstream regulated gene 1	10397	*
69	GTPBP1	GTP binding protein 1	9567	*
70	CAPZA1	capping protein (actin filament) muscle Z-line, α 1	829	*
71	RPL13	ribosomal protein L13	6137	*
72	CIDEC	cell death-inducing DFFA-like effector c	63924	*
73	SIRT3	sirtuin (silent mating type information regulation 2 homolog) 3 (S. cerevisiae)	23410	
74	LAPTM4A	lysosomal-associated protein transmembrane 4 α	9741	*
75	NOS1	nitric oxide synthase 1 (neuronal)	4842	*
76	COQ10B	coenzyme Q10 homolog B (S. cerevisiae)	80219	*
77	SAT	spermidine/spermine N1-acetyltransferase	6303	*
78	C1orf107	chromosome 1 open reading frame 107	27042	
79	TXN	thioredoxin	7295	*
80	SLC7A1	solute carrier family 7 (cationic amino acid transporter, y^+ system), member 1	6541	
81	SLC1A7	solute carrier family 1 (glutamate transporter), member 7	6512	
82	VIL2	villin 2 (ezrin)	7430	*
83	NTRK2	neurotrophic tyrosine kinase, receptor, type 2	4915	
84	GSTA1	Glutathione S-transferase A1	2938	
85	PTP4A3	protein tyrosine phosphatase type IVA, member 3	11156	

(*) Genes that were expressed in the cancer patient-derived colonocytes but not in either the healthy volunteer-derived colonocytes or the peripheral blood (PB) mixture.

peripheral blood, and squamous cells are potentially good markers for the screening of colorectal cancer from the feces. To identify effective genes for fecal RNA-based screening, we first compared 10 gene expression profiles of 6 early colorectal cancer tissues (2 Dukes stage A and 4 Dukes stage B cases), 3 advanced cancer RNA mixtures (6, 6, and 7 Dukes stage C or D cases), and a normal colorectal mucosa RNA mixture (6 cases). Of 14,564 genes, 2,926 were identified as genes which were not detected in the normal mucosa but detected in at least one of the above 9 cancer samples. Among these 2,926 cancer-specific genes, 205 genes, which were expressed in all of the 3 advanced cancer mixtures, were identified; however, only 3 genes were found to be expressed in all of the 6 early cancers. The cause of these results may

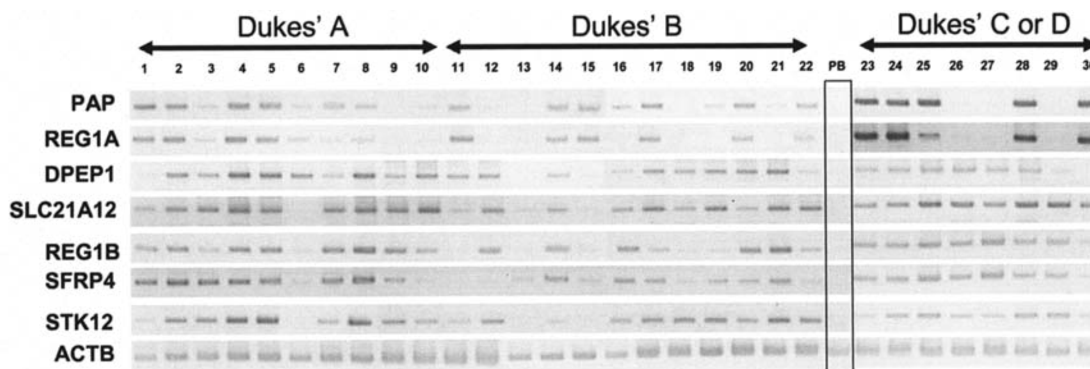

Figure 1. Results of RT-PCR of 7 genes (PAP, REG1A, DPEEP1, SLC21A12, REG1B, SFRP4, and STK12) selected by microarray analysis on 30 colorectal cancer tissues and on a peripheral blood mixture (PB). By microarray analyses, we first identified 15 genes which were expressed not in the normal colorectal mucosa mixture or in a peripheral blood mixture but in more than 4 of 6 early cancers (Dukes stage A or B) and in all of 3 advanced cancer mixtures (Dukes stage C or D). Next, we examined the frequency of the expression of these 15 genes in 30 colorectal cancer tissues (10 Dukes stage A, 12 Dukes stage B, and 8 Dukes stage C or D cancers), and selected, by RT-PCR, 7 genes (PAP, REG1A, DPEP1, SLC21A12, REG1B, SFRP4, and STK12) as the frequently expressed genes at any stage of colorectal cancer.

be that the expression profile of early cancer varies from case to case. Of 14,564 genes, we were able to select 65 genes which were expressed not in the normal colorectal mucosa mixture but in more than 4 of the 6 early cancers and in all of the 3 advanced cancer mixtures.

Considering bleeding by nonmalignant diseases such as hemorrhoids, which often give false positives in fecal colorectal cancer screening, we selected 15 genes from the 65 genes, because the expression of all the other 50 genes was detectable in a peripheral blood mixture of 58 healthy volunteers when a highly sensitive nested RT-PCR analysis with outer and inner primer sets was performed (data not shown). Next, we examined the frequency of the expression of these 15 genes in 30 colorectal cancer tissues (10 Dukes stage A, 12 Dukes stage B, and 8 Dukes stage C or D cancers), and selected, by RT-PCR, 7 genes (PAP, REG1A, DPEP1, SLC21A12, REG1B, SFRP4, and STK12) as the frequently expressed genes at any stage of colorectal cancer (Fig. 1). By RT-PCR, we lastly checked the expression of these 7 genes in RNA samples of the colonocytes of 15 healthy volunteers, which were isolated from the feces by FMCI (23). No mRNA expression of 3 genes (PAP, REG1A, and DPEP1) was detected in the colonocyte samples of all the 15 healthy volunteers; however, the other 4 genes (SLC21A12, REG1B, SFRP4, and STK12) were found to be expressed in some samples (data not shown). This fact may be due to the contamination of anal squamous cells, which were dissociated from the anus and survived in the feces, because our gene selection process can minimize the effect on the contamination of lymphocytes, red blood cells and dissociated normal colorectal epithelium. Under the above criteria, only 3 genes were selected as the final candidates for the fecal RNA-based early detection of colorectal cancer.

Marker gene selection by comparison of expression profiles between healthy volunteer- and cancer patient-derived colonocytes from the feces. Next, we obtained and compared 5 gene expression profiles of 4 colonocyte RNA samples (CF15, CF17, CF18, and CF25), which were isolated from the feces of 4 colorectal cancer patients by FMCI, and a colonocyte RNA mixture (HVF) of 7 healthy volunteers. Of 14,564 genes, the number of detectable genes in 5 colonocyte samples, CF15, CF17, CF18, CF25, and HVF is 768, 603, 772, 459, and 326, respectively. The number of detectable genes in the colonocyte is approximately 6.5% of that (11,343) in the colorectal cancer tissue. The major reason seems to be that most colonocytes are not in the cell division cycle but are resting, because the detectable gene number (1,535) in the peripheral blood composing such resting cells was also small. Unexpectedly, 716 (93%), 553 (92%), 712 (92%), and 424 (92%) of the above detectable genes (768, 603, 772, and 459) in the the colonocytes of the cancer patients were not expressed in those of the healthy volunteers. The huge difference of the colonocyte expression profiles between the colorectal cancer patients and the healthy volunteers might lead to success in gene selection for fecal RNA-based early detection of colorectal cancer. Eighty-five genes, whose expression was found in 3 or 4 of the 4 colorectal cancer patient samples (CF15, CF17, CF18, and CF25) but not in the HVF, were identified (Table II). Of these 85 genes, 29 (34%) were found to encode ribosomal proteins (RPLs or RPSs). In the course of a series of studies, it is predicted that normal mucous cells will die and be exfoliated during turnover and that the colorectal cancer cells will survive for a long time in the isolation processes as well as in the feces (22,23). Therefore, protein synthesis in the cells would be maintained actively for cancer cell survival under these conditions. The FMCI can minimize the contamination of both lymphocytes and red blood cells because the FMCI contains the enrichment process of epithelial cells such as colorectal cancer cells, the contaminated anal squamous cells, and a few living cells dissociated from the normal colorectal mucosa by MACS (23). Therefore, expression status in the peripheral blood is not needed for the gene selection process for fecal RNA-based screening; all of the 85 genes are expected to be good markers if the colonocytes are isolated by FMCI.

RT-PCR and focused microarray analyses of 9 selected genes in 30 colonocyte RNA samples. Next, we performed RT-PCR of the first 3 identified genes (PAP, REG1A, and DPEP1) in the colonocyte RNA samples which were prepared from 23 curable colorectal cancer patients (Dukes stages A-C) and 7 healthy volunteers. The 23 colorectal cancer patients were 8

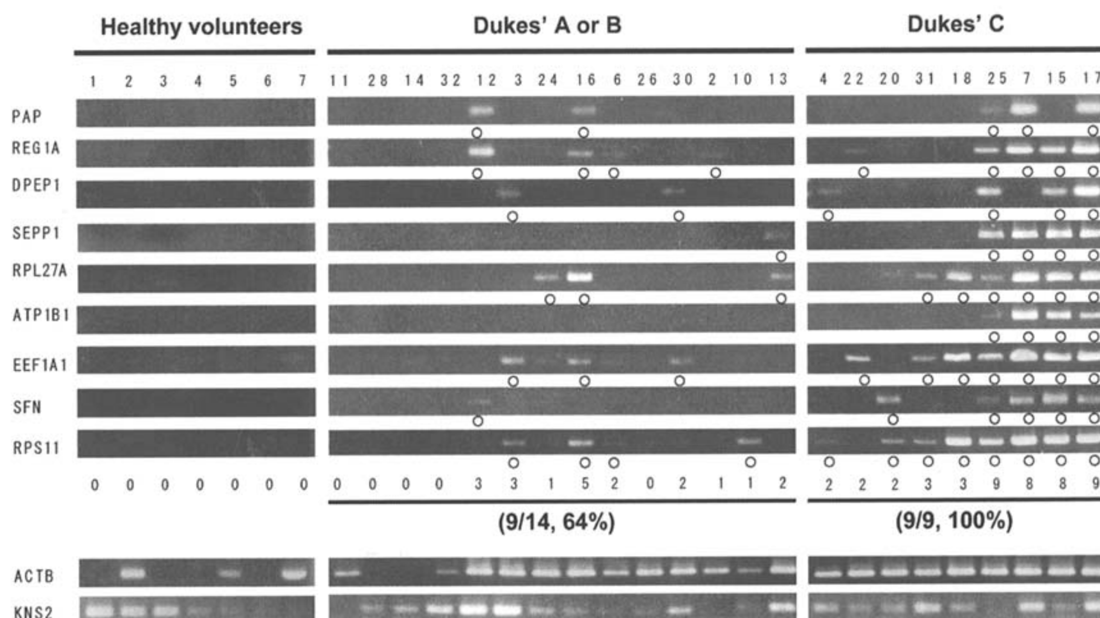

Figure 2. Results of RT-PCR of 9 genes (PAP, REG1A, DPEEP1, SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11) in the colonocyte RNA samples from 23 curable colorectal cancer patients and 7 healthy volunteers. We performed RT-PCR of the first 3 identified genes (PAP, REG1A, and DPEP1) in the colonocyte RNA samples which were prepared from 23 curable colorectal cancer patients (8 Dukes stage A, 6 Dukes stage B, and 9 Dukes stage C) and 7 healthy volunteers. Next, to test the power of the 85 genes, which were identified by colonocyte gene expression profiling, we randomly selected 6 (SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11), and performed RT-PCR on the same samples. In total, RT-PCR of those 9 genes detected 18 (78%) of the 23 cancer patients, and 9 (64%) of the 14 early cancers (Dukes stage A or B) were detected; however, the expression of all of the 7 genes was hardly detected in the 7 healthy volunteers (upper panel). The expression level of housekeeping genes such as ACTB (ß-actin) was highly varied from sample to sample (lower panel). KNS2 encoding kinesin 2 was selected, by microarray analyses, as a gene expressed constantly in any colonocyte RNA sample; however, the expression level was also varied. Open circles indicate positive RT-PCR product, and numbers indicate the number of the positive genes in each sample.

with Dukes stage A, 6 with Dukes stage B, and 9 with Dukes stage C cancers; 5 were right-sided and 18 were left-sided cancers. Twelve (52%) of the 23 cancers were positive by RT-PCR in at least one of the 3 genes whereas no positive gene was found in any of the healthy volunteers (Fig. 2). To test the power of the 85 genes, which were identified by colonocyte gene expression profiling, we randomly selected 6 (SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11). RT-PCR of these 6 genes detected 16 (70%) of the 23 cancers as at least positive for 1 gene whereas no positive gene was found in any of the healthy volunteers (Fig. 2). No or a quite low signal of all the 9 genes was found in another RT-PCR experiment with 8 healthy volunteers (data not shown). In total, RT-PCR of those 9 genes detected 18 (78%) of the 23 cancer patients (Fig. 2). The 18 patients detected were 4 with Dukes stage A, 5 with Dukes stage B, and 9 with Dukes stage C cancers; 4 were right-sided and 14 were left-sided cancers. Therefore, 9 (64%) of the 14 early cancers (Dukes stage A or B), which have no lymph node metastasis, and show a good prognosis, were able to be detected. Importantly, 4/5 (80%) of the right-sided colorectal cancers were detected, which have been reported to be very difficult to detect by any feces-based molecular biological method, because most right-sided cancer-derived colonocytes are severely damaged from remaining for a long time in the feces.

For fecal RNA-based detection of early colorectal cancer, quantitative real-time RT-PCR is thought to hardly apply in the colonocyte because the expression level of housekeeping genes was highly varied from sample to sample (Fig. 2). The expressional variation could be explained by the difference of the physiological condition of colorectal cancer cells and anal squamous cells isolated from the feces by FMCI. All of the 9 genes were selected as cancer cell- or cancer patient-derived colonocyte-specific genes. Therefore, a negative or positive assay was thought to be sufficient for fecal RNA-based colorectal cancer detection. Accordingly, we developed a multiplex RT-PCR-based microarray assay for evaluating the above RT-PCR results and for providing an effective imaging tool for mass cancer screenings (Fig. 3). The Cy3-labeled cDNAs prepared by multiplex RT-PCR in a single tube were hybridized with 9 gene sequences on a focused microarray, which was manufactured by our previously developed Bubble Jet Technology with a small modification (25). Hybridization signals and the number of positive genes in the above 23 cancer patient-derived colonocyte RNA samples and 7 healthy volunteer-derived colonocyte RNA samples are shown in Fig 4. In total, a high concordance was observed between the focused microarray and RT-PCR. The focused microarray detected 18 (78%) of the 23 cancer patients. Ten (71%) of the 14 early cancers (Dukes stage A or B) and 4 (80%) of the 5 right-sided cancers were detected.

Discussion

Although the number of samples examined in this study is considered to be small, the evidence suggests that these successful results could be obtained from the high-quality of the RNA of the colonocytes, which were isolated by FMCI. From a practical point of view for mass cancer screenings, it is noted that the same number of colonocytes from fecal

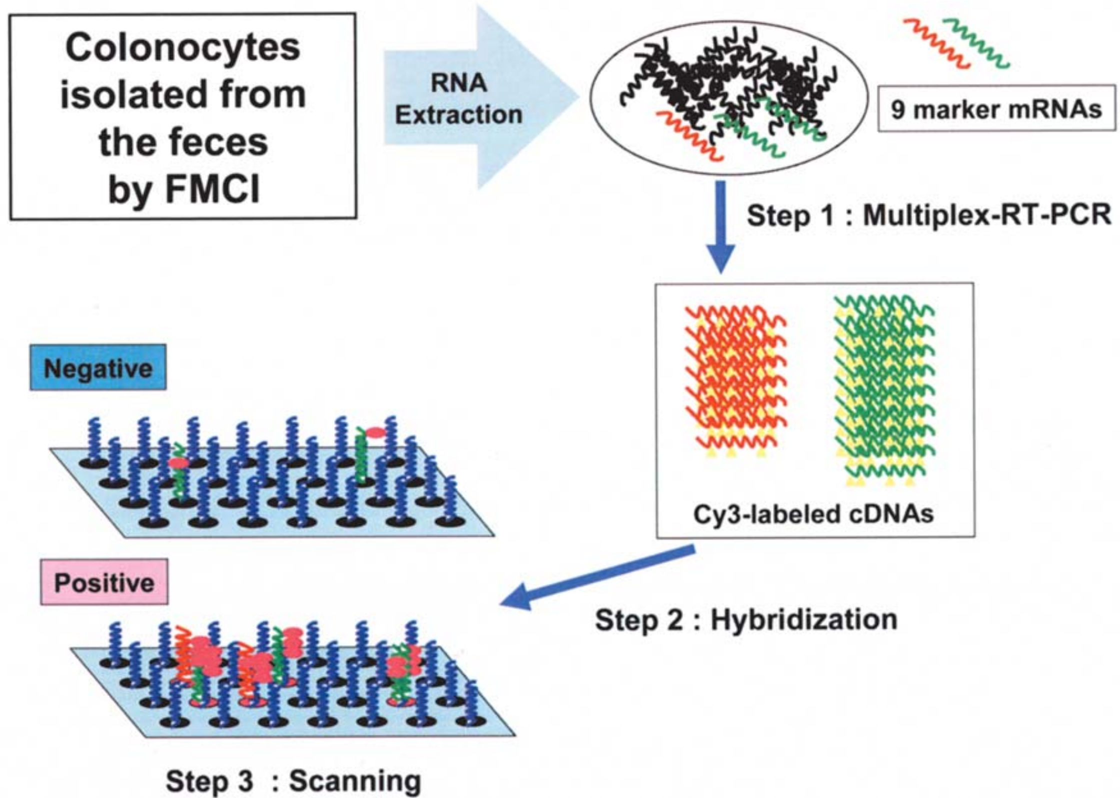

Figure 3. Schematic flow diagram of a focused micoarray assay. Marker mRNAs (PAP, REG1A, DPEP1, SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11 mRNAs) were amplified and labeled with Cy3-dUTP by multiplex-RT-PCR among total RNAs from colonocytes isolated by FMCI (step 1) and hybridized to focused micoarray (step 2), followed by fluorescence intensity scanning (step 3).

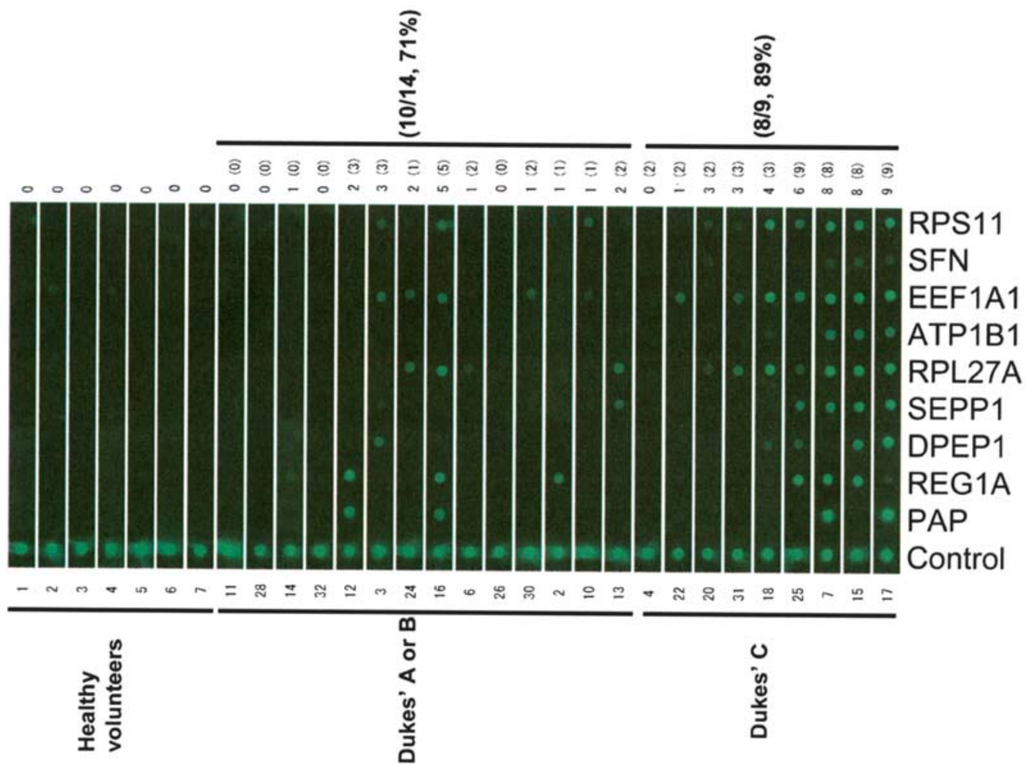

Figure 4. Hybridization image of focused microarray analysis and the number of positive genes in 30 colonocyte RNA samples. The Cy3-labeled cDNAs prepared by multiplex RT-PCR in two tubes were hybridized with 9 gene sequences (PAP, REG1A, DPEP1, SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11) on a focused microarray, which was manufactured by our previously developed Bubble Jet Technology with a small modification (25). Hybridization signals and the number of positive genes (right) in the above 23 cancer patient-derived colonocyte RNA samples and 7 healthy volunteer-derived colonocyte RNA samples are shown. In total, a high concordance was observed between focused microarray and RT-PCR. Ten (71%) of the 14 early cancers (Dukes stage A or B) and 8 (89%) of 9 Dukes stage C cancers were detected by the focused microarray analysis. The number of positive genes in RT-PCR are in parentheses (Fig. 2).

materials 6 h to 3 days after evacuation can be obtained if the feces are kept at 4˚C (23). However, if conventional fecal RNA preparation methods without the epithelial cell enrichment process are used for colorectal cancer screening, we have to consider the contamination of blood in the feces, which derives from nonmalignant diseases. Considering the use of such methods, to this end we further provided 56 genes, which were expressed in the cancer patient-derived colonocytes but not in either the healthy volunteer-derived colonocytes or the peripheral blood mixture (Table II). This study suggests that the fecal RNA-based method could be a promising procedure for the detection of early or right-sided colorectal cancers. We recently developed a very effective focused microarray assay for detecting minimal gastric cancer cells in peritoneal washings, demonstrating a specificity and sensitivity equal to or better than cytology in two large specialist hospitals with trained cytologists (26). Therefore, the focused microarray assay could provide an effective imaging tool for mass screening, and our extensive gene list provides useful markers.

Acknowledgments

This study was supported in part by the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio); in part by a Grant-in-Aid for the Third Comprehensive 10-Year Strategy for Cancer Control and for Cancer Research (16-15) from the Ministry of Health, Labour and Welfare of Japan, and in part by a research and development project of the Industrial Science and Technology Program supported by the New Energy and Industrial Technology Development Organization (NEDO) of Japan. We thank Ayako Nomura and Megumi Sato for their excellent assistance with primer design for RT-PCR.

References

1. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schman LM and Ederer F: Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 328: 1365-1371, 1993.
2. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, James PD and Mangham CM: Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 348: 1472-1477, 1996.
3. Kronborg O, Fenger C, Olson J, Jorgensen OD and Sondergaard O: Randomized study of screening for colorectal cancer with faecal-occult-blood test. Lancet 348: 1467-1471, 1996.
4. Towler B, Irwig L, Glasziou P, Kewenter J, Weller D and Silagy C: A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, hemoccult. BMJ 317: 559-565, 1998.
5. Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferucci J, Ganiats T, Levin T, Woolf S, Johnson D, Kirk L, Litin S and Simmang C: Colorectal cancer screening and surveillance: clinical guidelines and rationale-update based on new evidence. Gastroenterology 124: 544-560, 2003.
6. Mandel JS, Church TR, Bond JH, Ederer F, Geisser MS, Mongin SJ, Snover DC and Schuman LM: The effect of fecal occult-blood screening on the incidence of colorectal cancer. N Engl J Med 343: 1603-1607, 2000.
7. Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P and Vogelstein B: Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. Science 256: 102-105, 1992.
8. Hasegawa Y, Takeda S, Ichii S, Koizumi K, Maruyama M, Fujii A, Ohta H, Nakajima T, Okuda M, Baba S, *et al*: Detection of K-ras mutations in DNAs isolated from feces of patients with colorectal tumors by mutant-allele-specific amplification (MASA). Oncogene 10: 1441-1445, 1995.
9. Smith-Ravin J, England J, Talbot IC and Bodmer W: Detection of c-Ki-ras mutations in faecal samples from sporadic colorectal cancer patients. Gut 36: 81-86, 1995.
10. Eguchi S, Kohara N, Komuta K and Kanematsu T: Mutations of the p53 gene in the stool of patients with resectable colorectal cancer. Cancer 77: 1707-1710, 1996.
11. Nollau P, Moser C, Weinland G and Wagener C: Detection of K-ras mutations in stools of patients with colorectal cancer by mutant-enriched PCR. Int J Cancer 66: 332-336, 1996.
12. Ratto C, Flamini G, Sofo L, Nusera P, Ippliti M, Curigliano G, Ferreti G, Sgambato A, Merico M, Doglietto GB, Cittadini A and Crucitti F: Detection of oncogene mutation from neoplastic colonic cells exfoliated in feces. Dis Colon Rectum 39: 1238-1244, 1996.
13. Deuter R and Muller O: Detection of APC mutations in stool DNA of patients with colorectal cancer by HD-PCR. Hum Mutat 11: 84-89, 1998.
14. Ahlquist DA, Skoletsky JE, Boynton KA, Harrington JJ, Mahoney DW, Pierceall WE, Thibodeau SN and Shuber AP: Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. Gastroenterology 119: 1219-1227, 2000.
15. Dong SM, Traverso G, Johnson C, Geng L, Favis R, Boynton K, Hibi K, Goodman SN, D'Allessio M, Paty P, Hamilton SR, Sidransky D, Barany F, Levin B, Shuber A, Kinzler KW, Vogelstein B and Jen J: Detecting colorectal cancer in stool with the use of multiple genetic targets. J Natl Cancer Inst 93: 858-865, 2001.
16. Rengucci C, Maiolo P, Saragoni L, Zoli W, Amadori D and Calistri D: Multiple detection of genetic alterations in tumors and stool. Clin Cancer Res 7: 590-593, 2001.
17. Traverso G, Shuber A, Olsson L, Levin B, Johnson C, Hamilton SR, Boynton K, Kinzler KW and Vogelstein B: Detection of proximal colorectal cancers through analysis of faecal DNA. Lancet 359: 403-404, 2002.
18. Traverso G, Shuber A, Levin B, Levin B, Johnson C, Hamilton SR, Boynton K, Kinzler KW and Vogelstein B: Detection of APC mutations in fecal DNA from patients with colorectal tumors. N Engl J Med 346: 311-320, 2002.
19. Boynton KA, Summerhayes IC, Ahlquist DA and Shuber AP: DNA integrity as a potential marker for stool-based detection of colorectal cancer. Clin Chem 49: 1058-1065, 2003.
20. Matsumura Y and Tarin D: Significance of CD44 gene products for cancer diagnosis and disease evaluation. Lancet 340: 1053-1058, 1992.
21. Mastumura Y, Hanbury D, Smith J and Tarin D: Non-invasive detection of malignancy by identification of unusual CD44 gene activity in exfoliated cancer cells. BMJ 308: 619-624, 1994.
22. Yamao T, Matsumura Y, Shimada Y, Moriya Y, Sugihara K, Akasu T, Fujita S and Kakizoe T: Abnormal expression of CD44 variants in the exfoliated cells in the feces of patients with colorectal cancer. Gastroenterology 114: 1196-1205, 1998.
23. Matsushita H, Matsumura Y, Moriya Y, Akasu T, Fujita S, Yamamoto S, Onouchi S, Saito N, Sugito M, Ito M, Kozu T, Minowa T, Nomura S, Tsunoda H and Kakizoe T: A new method for isolating colonocytes from naturally evacuated feces and its clinical application to colorectal cancer diagnosis. Gastroenterology 129: 1918-1927, 2005.
24. Mori K, Aoyagi K, Ueda T, Danjoh I, Tsubosa Y, Yanagihara K, Matsuno Y, Sasako M, Sakamoto H, Mafune K, Kaminishi M, Yoshida T, Terada M and Sasaki H: Highly specific marker genes for detecting minimal gastric cancer cells in cytology negative peritoneal washings. Biochem Biophys Res Commun 313: 931-937, 2004.
25. Okamoto T, Suzuki T and Yamamoto N: Microarray fabrication with covalent attachment of DNA using Bubble Jet Technology. Nat Biotech 18: 438-441, 2000.
26. Mori K, Suzuki T, Uozaki H, Nakanishi H, Ueda T, Matsuno Y, Kodera Y, Sakamoto H, Yamamoto N, Sasako M, Kaminishi M and Sasaki H: Detection of minimal gastric cancer cells in peritoneal washings by focused microarray analysis with multiple markers: clinical implications. Ann Surg Oncol 14: 1694-1702, 2007.