Gene expression of colorectal cancer: Preoperative genetic diagnosis using endoscopic biopsies

TAKAMICHI KOMORI[1], ICHIRO TAKEMASA[1], MAKOTO YAMASAKI[1], MASAAKI MOTOORI[1],
TAKESHI KATO[2], NOBUTERU KIKKAWA[2], NAOMASA KAWAGUCHI[3],
MASATAKA IKEDA[1], HIROFUMI YAMAMOTO[1], MITSUGU SEKIMOTO[1],
KENICHI MATSUBARA[4], NARIAKI MATSUURA[3] and MORITO MONDEN[1]

[1]Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871;
[2]Department of Surgery, Minoh City Hospital, 5-7-1 Sugano, Minoh, Osaka 562-0014; [3]Department of Molecular Pathology,
Graduate School of Medicine and Health Science, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871;
[4]DNA Chip Research Inc., 1-1-43 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Abstract. In colorectal cancer, to predict the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively, a more competent diagnostic system is required, in addition to conventional diagnosis based on morphology and pathology. The application of gene expression profiling to preoperative cancer diagnosis using endoscopic biopsies could enable the selection of a more appropriate therapy for patients. In this study, we evaluated the feasibility of gene expression profiling using preoperative biopsies of colorectal tumors in a clinical setting, by investigating the influence of intra-tumor heterogeneity on the profiles and testing the prediction ability of tumor malignancy. Under endoscopic examination, two biopsies were sampled from each of 10 colorectal cancers and 10 adenomas, and their gene expression data were obtained using cDNA microarrays. The intra- and inter-tumor heterogeneities of the profiles were compared with unsupervised clustering analysis. Molecular prediction of tumor malignancy using biopsies was performed with the supervised classification algorithm. In clustering analysis, almost all paired biopsies from the same tumors joined each other. Pearson's correlation coefficients of the profiles between biopsies from the same tumors (mean, 0.83) were significantly greater than those of the profiles between biopsies from other cancers (mean, 0.58) (p<0.0001). In the supervised classification method, malignancy was correctly predicted in 39 out of 40 biopsies with 8-71 informative genes. Gene expression profiling using endoscopic biopsies of colorectal tumors revealed that the intra-tumor heterogeneity was smaller than the inter-tumor heterogeneity and tumor malignancy was correctly predicted. Our findings suggest that the technique of gene expression profiling accurately represents the biological properties of colorectal cancer and could help the preoperative diagnosis of this disease.

Introduction

The incidence of colorectal cancer (CRC) is increasing and it is one of the leading causes of cancer death in Japan (1). Conventional diagnosis based on morphology and pathology, such as Dukes' classification and the tumor-node-metastasis (TNM) staging system, has played an important role in the clinical decision-making and evaluation of prognosis for CRC (2-4). However, it is difficult to differentiate the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively by conventional diagnosis. To predict such individual heterogeneous cancers' characteristics preoperatively, a more competent diagnostic system is required.

Comprehensive gene expression assay using microarray technology has provided insights into cancer pathogenesis and is expected to help to fulfil the clinical demands for individualized medicine (5). The discovery of a set of new molecular markers, which can classify cancers according to their properties using surgical specimens, has been reported in various cancers (6). In CRC, this technology has been used to elucidate the mechanisms involved in carcinogenesis (7-16) as well as to predict various clinicopathological aspects, such as recurrence after surgery (17-22). However, in clinical practice, this fruitful molecular prediction using surgical specimens will be limited to the selection of postoperative medicine such as adjuvant therapy and follow-up schedules. On the other hand, analysis using preoperative endoscopic biopsies, instead of

Table I. Patients and tumor characteristics.

Case	Age (years)	Sex	Tumor location	Tumor size (cm)	Tumor type	TNM staging		
						T	N	Stage
Ca1	67	M	D	2.0	Well differentiated adenocarcinoma	T1	N0	I
Ca2	55	F	R	1.5	Well differentiated adenocarcinoma	T1	N0	I
Ca3	69	F	R	5.6	Well differentiated adenocarcinoma	T2	N0	I
Ca4	59	F	R	3.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca5	66	M	R	5.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca6	48	M	S	2.8	Moderately differentiated adenocarcinoma	T2	N0	I
Ca7	73	F	S	5.7	Well differentiated adenocarcinoma	T3	N0	II
Ca8	55	M	R	3.8	Moderately differentiated adenocarcinoma	T3	N0	II
Ca9	67	F	R	3.4	Moderately differentiated adenocarcinoma	T3	N0	II
Ca10	60	F	R	3.0	Well differentiated adenocarcinoma	T2	N1	III
Ad1	60	M	S	4.0	Tubulovillous adenoma			
Ad2	74	M	S	2.5	Tubular adenoma			
Ad3	77	M	A	2.5	Tubulovillous adenoma			
Ad4	68	M	T	1.1	Tubular adenoma			
Ad5	62	F	A	1.0	Tubulovillous adenoma			
Ad6	60	M	D	2.0	Tubulovillous adenoma			
Ad7	72	M	R	0.8	Tubulovillous adenoma			
Ad8	61	M	R	0.8	Tubular adenoma			
Ad9	64	F	S	1.2	Tubular adenoma			
Ad10	78	F	S	1.8	Tubular adenoma			

M, male; F, female; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum.

surgically resected samples, would widen the utility of microarray technology.

In rectal cancer, preoperative rather than postoperative chemo- and/or radio-therapy reduces local recurrence after surgery (23). Using preoperative biopsies under colonoscopic examination, prediction of the response to preoperative chemo- and/or radio-therapy would be useful for the selection of patients who would most benefit from preoperative therapies aimed at a better prognosis or improved chances of sphincter preservation (24,25). The prediction of lymph node metastasis could contribute to the avoidance of unnecessary surgery for early invasive CRC. This is clinically important since lymph node metastasis is found in only approximately 10% of early invasive CRCs (26); the remaining 90% without metastasis are more suited to undergo local excision such as colonoscopic resection or transanal endoscopic microsurgery.

To determine the clinical importance of diagnosis based on gene expression profiles using preoperative endoscopic biopsies in CRC, we investigated the data quality from low volume samples and the influence of intra-tumor hetero-geneity on the profiles. Wide differences in profiles within a tumor would hinder the adoption of this technique. There are no reports that compared intra- and inter-tumor heterogeneity in expression profiles in colorectal tumors. In the present study, we sampled two biopsy specimens from each of 10 cancers and 10 adenomas obtained under colonoscopic examinations, and determined their gene expression profiles using cDNA

microarrays. By comparing their profiles, we assessed the intra-tumor heterogeneity in colorectal tumors. Moreover, as a first step for clinical applicability, by testing the molecular prediction ability of tumor malignancy, we investigated whether this technique, employing preoperative endoscopic biopsies, allows characterization of the biological properties of colorectal cancer.

Materials and methods

Patients and sample collection. From each of the 10 cancers and 10 adenomas, two biopsy specimens were obtained under colonoscopic examination. The clinicopathological data of patients and their tumors are summarized in Table I. Six bulky samples of the same cancers and 8 normal colorectal epithelium tissues were obtained from surgically resected specimens. None of the adenomas contained a cancerous component. Samples were stored in RNAlater (Qiagen, Valencia, CA) at -20°C after sampling until RNA extraction. This study was approved by the Institutional Review Board of Osaka University and Minoh City Hospital, and informed consent was obtained from all patients.

RNA extraction. Total RNA was extracted from biopsy samples using an RNeasy kit (Qiagen). The average volume of extracted total RNA from one biopsy was 21.1 μg in cancers and 22.6 μg in adenomas. From bulky samples, total RNA was extracted

using Trizol reagent (Invitrogen, Carlsbad, CA) and cleaned using the RNeasy kit. The quality of extracted total RNA was checked by 0.8% agarose-gel electrophoresis and a LabChip kit (Agilent Technologies, Palo Alto, CA), and the quantity was determined with a spectrophotometer. The total RNA solution was stored at -80˚C until use.

Hybridization to cDNA microarray. We used cDNA micro-arrays containing 4608 clones which were derived from 30,000 clones expressed in CRC tissues (10,16,22). The Pearson's correlation coefficient of the data from replicated samples using our microarrays was 0.95. Gene expression data were obtained using a previously described method (10), with some improvements. As a standard normal control reference, a mixture of total RNA extracted from 40 normal colorectal epithelia was used. Labeled cDNA targets for hybridization were synthesized by reverse transcription from standard- and sample-total RNA respectively, with the indirect labeling method.

For each reverse transcription, 12 μg of total RNA was mixed with 1 μg of oligo-dT primer (Invitrogen) in a total volume of 15.5 μl, heated to 70˚C for 10 min and cooled for 5 min. The mixture consisted of 6 μl of 5X first strand buffer, 3 μl of 0.1 M DTT, 3 μl of nucleotide cocktail (5 mM each dATP, dCTP, and dGTP, 3 mM dTTP and 2 mM aminoallyl-dUTP) and 0.5 μl of 40 U/μl RNase inhibitor was added. After incubation for 2 min at 42˚C, 2 μl of 200 U/μl Superscript II reverse transcriptase (Invitrogen) was added. After incubation for 1 h at 42˚C, 3.3 μl of 0.5 M EDTA was added and the RNA strand was degraded with 3.3 μl of 2 N NaOH and incubation at 70˚C for 20 min. The mixture was neutralized with 3.3 μl of 2 N HCL and 60 μl of distilled water. This cDNA was purified using a QIAquick kit (Qiagen) and dried cDNA was dissolved in 9 μl of 0.2 M $NaHCO_3$-Na_2CO_3 (pH 9.0). Cy-dye solution (1 μl) (Amersham, Piscataway, NJ), Cy3 for the standard target, and Cy5 for the tested target, were added and incubated under shade at room temperature for 1 h, respectively. After purification using Micro Bio-Spin Columns P-30 Tris (Bio-Rad, Hercules, CA), the separately synthesized Cy3- and Cy5-labeled targets were combined and purified using Microcon YM-30 (Millipore, Billerica, MA), concentrated to a volume of 16.5 μl. To the concentrated target, 3.5 μl of human COT-1 DNA (Invitrogen), 7 μl of 20X SSC, 7 μl of 20X Denhart's solution, and 1 μl of 10% SDS were added. The 35 μl of target mixture was denatured by heating for 2 min at 95˚C and cooled on ice. After the incubation at 50˚C, the target was placed on the array. The array was incubated at 50˚C for 14 h in a humid chamber. After hybridization, the slides were washed in 2X SSC with 0.1% SDS for 10 min, 0.1X SSC with 0.1% SDS for 10 min, and 0.1X SSC for 5 min at 30˚C.

Scan and data processing. The array was scanned with ScanArray Lite (Perkin-Elmer, Wellesley, MA). The images were analyzed with QuantArray software (GSI Lumonics, Billerica, MA), converting signal intensities of each spot into numerical data. Data was processed after background subtraction, as described previously (10,16,22). Cy5/Cy3 ratios were log-transformed, and the global normalization was performed. Genes with >15% missing values in each

group of cancers and adenomas were excluded from further analysis.

Statistical analysis. First, gene expression profiles of 20 cancer biopsies and 8 normal epithelia were compared. Hierarchical clustering analysis (HCA) was performed with the software GeneMaths version 2.0 (Applied Maths, Inc., Austin, TX) using all 1966 genes after data processing. Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using arithmetic average as the clustering algorithm. In HCA, similarity in the expression pattern of paired biopsies from the same cancer was compared with those from other cancers, to define the difference between intra- and inter-tumor heterogeneity of profiles in cancers. Statistical significance of the difference was determined using t-test. To verify the validity of the list of differentially expressed genes involved in colorectal carcinogenesis using biopsy samples, examined genes in cancer biopsies and normal epithelia were ranked according to the signal-to-noise ratio (SNR) (5) and they were compared with previous reports using tissue samples. The p-value of SNR was calculated by performing the random permutation test 1,000 times (5).

Next, using 20 cancer biopsies and 20 adenoma biopsies, the prediction of malignancy in colorectal tumors by their gene expression profiles was tested. Genes differentially expressed in cancer and adenoma biopsies were ranked according to SNR, and differential diagnosis was performed using the weighted-votes method (5). The prediction accuracy was determined with the leave-one-out cross validation (5). Positive prediction strength was judged to be a cancer and negative to be an adenoma. Using the gene set with the highest accuracy, HCA was performed to identify the difference between the intra- and the inter-tumor heterogeneity of the profiles in cancers and adenomas.

Immunohistochemical staining. Immunohistochemical staining (IHC) was performed to investigate the translation of mRNA of differentially expressed genes to each coding protein and to examine the intra-tumor heterogeneity of expression patterns in cancers at the protein level. Among differentially expressed genes between cancer biopsies and normal epithelia in our transcriptional analyses, 7 genes whose antibodies were commercially available were selected. Buffered formalin-fixed (10%), paraffin-embedded sections were prepared from 10 surgically resected cancers, whose biopsy specimens were included in the microarray analyses. The streptavidin-biotin immunoperoxidase complex method (27) was used for IHC. Primary antibodies used were as follows; polyclonal antibody (pAb) to human peroxiredoxin 1 (PRDX1, Alexis Biochemicals, Lausen, Switzerland), pAb to high-mobility group box 1 (HMGB1, Santa Cruz Biotechnology, Santa Cruz, CA), pAb to DEK oncogene (DEK, Santa Cruz Biotechnology), pAb to poly(A) binding protein, cytoplasmic 1 (PABPC1, Santa Cruz Biotechnology), monoclonal antibody (mAb) to heat shock 60-kDa protein 1 (HSPD1, Sigma Aldrich, St. Louis, MO), mAb to nucleolin (NCL, Santa Cruz Biotechnology) and mAb to carbonic anhydrase II (CA2, Rockland, Gilbertsville, PA). Sections for negative control were tested by using normal mouse serum instead of primary
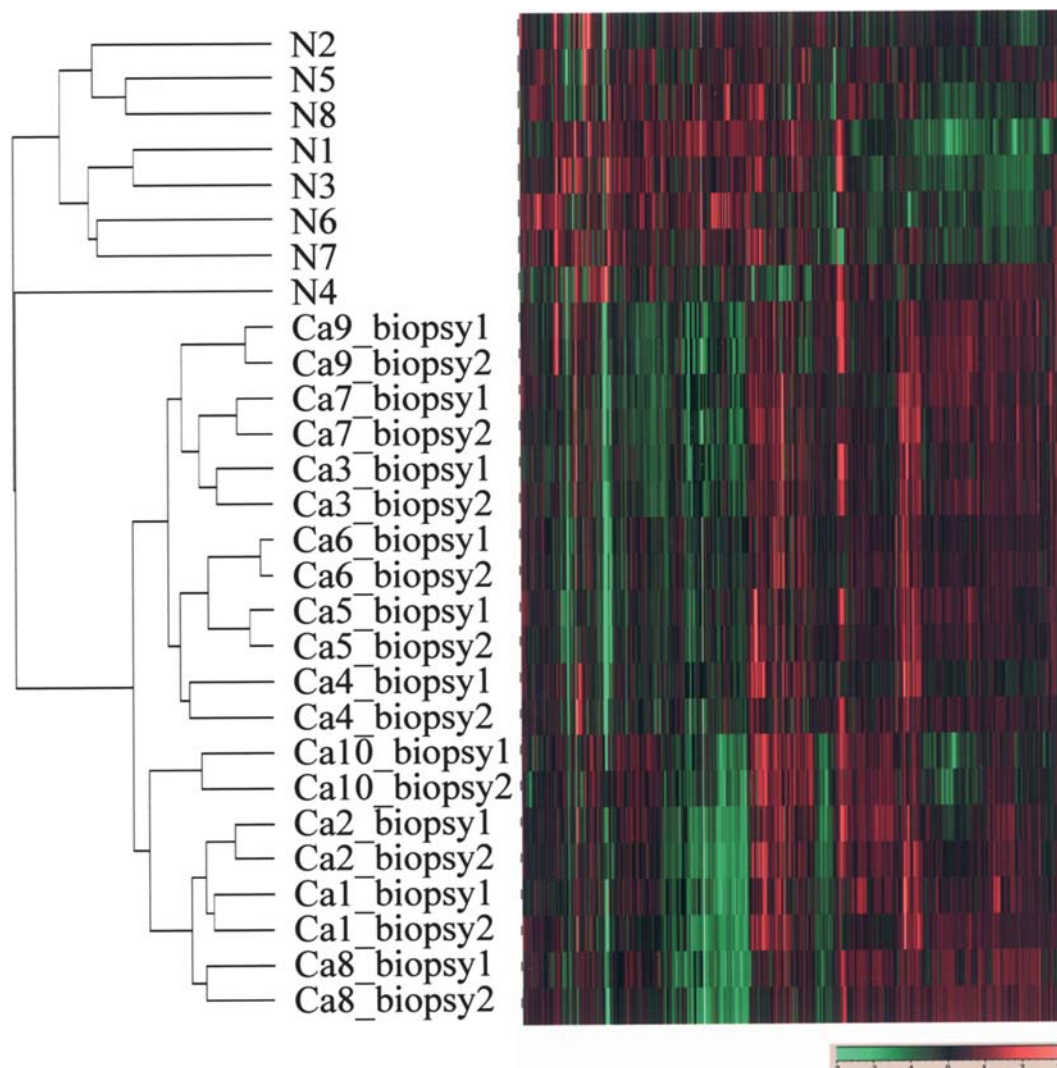
Figure 1. Hierarchical clustering analysis using full genes. Samples consisted of cancer biopsies (n=20; 10 pairs) and normal epithelia (n=8). Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using the arithmetic average as the clustering algorithm. Red indicates overexpression, and green indicates underexpression. The respective paired biopsies from different areas of the same tumors are joined together.

antibody. Tissue sections of thyroid adenoma (for PRDX1), poorly differentiated gastric adenocarcinoma (for HMG1 and DEK), uterine cervical carcinoma (for HSPD1 and NCL), normal testis (for PABPC1) and normal colon (for CA2) were prepared as positive controls according to the recommendations of the manufacturer and previous publications on IHC. All slides were evaluated by a pathologist who was blinded to the microarray data. For each immunohistochemical analysis, the mean intensity in epithelial or tumor cells was evaluated in comparison with the positive controls as follows: weak, 1+; moderate, 2+; and strong, 3+.

Results

Intra-tumor heterogeneity in colorectal cancers. Comparison of gene expression profiles between cancers and normal epithelia by HCA showed clear separation of the 20 cancer biopsies from the 8 normal tissues using all genes (Fig. 1). Moreover, the respective paired biopsies from different areas of the same cancers were joined together. Pearson's correlation coefficients of paired biopsies from the same cancers (mean,

0.83; SD, 0.08) were significantly greater than those of unpaired biopsies from other cancers (mean, 0.58; SD, 0.12; p<0.0001). Pearson's correlation coefficients between biopsies and their parent surgical bulky samples (mean, 0.60; SD, 0.15) were significantly greater than those between biopsies and surgical bulky samples from other cancers (mean, 0.37; SD, 0.16; p<0.0001).

Differentially expressed genes in cancer biopsies. Differentially expressed genes in cancer biopsies and normal epithelia were ranked according to SNR. Among them, 692 up-regulated and 219 down-regulated genes in cancer biopsies had SNR (p<0.001). The top 20 up-regulated and 20 down-regulated genes, excluding 4 down-regulated genes with no definition, are listed in Table II. PABPC1 and HSPD1 in the up-regulated genes and CA2, carboxylesterase 2 (CES2) and one EST in the down-regulated genes in cancer biopsies were also included in differentially expressed genes involved in colorectal carcinogenesis in our previous reports (10,16). Nine of the 40 differentially expressed genes were reported previously with respect to colorectal carcinogenesis with DNA microarray

Table II. Differentially expressed genes in colorectal cancer biopsies and normal colorectal epithelia.

Accession no.	Symbol	Gene definition	Previous report[a]	SNR
Up-regulated genes in cancer biopsies				
NM_181697	**PRDX1**	**peroxiredoxin 1, transcript variant 3**		**2.645**
NM_002128	**HMGB1**	**high-mobility group box 1**	**8,11,15**	**2.364**
NM_003472	**DEK**	**DEK oncogene (DNA binding)**	**11**	**2.341**
NM_001009	RPS5	ribosomal protein S5		2.052
NM_002568	**PABPC1**	**poly(A) binding protein, cytoplasmic 1**	**11,16**	**1.994**
NM_199440	**HSPD1**	**heat shock 60-kDa protein 1 (chaperonin), nuclear gene encoding mitochondrial protein, transcript variant 2**	**10,11,16,19**	**1.985**
NM_021130	**PPIA**	**peptidylprolyl isomerase A (cyclophilin A), transcript variant 1**		**1.957**
BE564899		EST		1.930
NM_001011	RPS7	ribosomal protein S7		1.894
NM_000978	RPL23	ribosomal protein L23		1.878
NM_002954	RPS27A	ribosomal protein S27a		1.860
AK090536		EST		1.826
NM_198829	**RAC1**	**ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1), transcript variant Rac1c**		**1.814**
NM_000971	RPL7	ribosomal protein L7		1.795
NM_003908	EIF2S2	eukaryotic translation initiation factor 2, subunit 2 ß, 38 kDa		1.772
NM_003756	**EIF3S3**	**eukaryotic translation initiation factor 3, subunit 3 γ, 40 kDa**		**1.763**
NM_000998	RPL37A	ribosomal protein L37a		1.732
NM_021034	**IFITM3**	**interferon induced transmembrane protein 3 (1-8U)**		**1.730**
NM_005381	**NCL**	**nucleolin**	**9**	**1.727**
NM_002106	H2AFZ	H2A histone family, member Z		1.721
Down-regulated genes in cancer biopsies				
NM_174977	SEC14L4	SEC14-like 4 (S. cerevisiae)		-2.518
NM_000067	**CA2**	**carbonic anhydrase II**	**8-11,15,16**	**-2.280**
NM_017958	PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2		-2.272
NM_005578	**LPP**	**LIM domain containing preferred translocation partner in lipoma**	**18**	**-2.236**
BF826364		EST		-1.926
BC051280		EST		-1.921
NM_003869	**CES2**	**carboxylesterase 2 (intestine, liver), transcript variant 1**	**10,14,16**	**-1.897**
BX647478		EST		-1.834
XM_375330		EST		-1.811
NM_002518	NPAS2	neuronal PAS domain protein 2		-1.789
CD652660		EST		-1.766
NM_020746	KIAA127	KIAA1271 protein		-1.742
NM_000355	TCN2	transcobalamin II; macrocytic anemia		-1.731
NM_005242	F2RL1	coagulation factor II (thrombin) receptor-like 1		-1.728
BM987276		EST		-1.698
XM_370781		**EST**	**16**	**-1.684**
AK129631		FLJ26120		-1.657
NM_002096	GTF2F1	general transcription factor IIF, polypeptide 1, 74 kDa		-1.623
NM_000014	A2M	α-2-macroglobulin		-1.617
NM_020675	Spc25	kinetochore protein Spc25		-1.606

Top 20 up-regulated and 20 down-regulated genes excluding 4 down-regulated genes with no definition, among 692 up-regulated and 219 down-regulated genes in cancer biopsies with SNR (p<0.001). Bold: genes reportedly involved in colorectal carcinogenesis based on microarray analysis or genes reportedly involved in carcinogenesis of other cancer types. [a]Previous report on colorectal carcinogenesis using microarray analysis (ref. no.). SNR, signal-to-noise ratio.

Table III. Immunohistochemical staining.

Gene name	Symbol		Expression grade[a]				Staining pattern[b]
			-	1+	2+	3+	
Peroxiredoxin 1	PRDX1	Normal	10				
		Cancer		6	4		Homogeneous
High-mobility group box 1	HMGB1	Normal		10			
		Cancer		1	7	2	Homogeneous
DEK oncogene	DEK	Normal			7	3	
		Cancer		4	6		Heterogeneous
Poly(A) binding protein, cytoplasmic 1	PABPC1	Normal		10			
		Cancer		3	7		Heterogeneous
Heat shock 60-kDa protein 1	HSPD1	Normal	10				
		Cancer			2	8	Homogeneous
Nucleolin	NCL	Normal	4	6			
		Cancer			5	5	Homogeneous
Carbonic anhydrase II	CA2	Normal			4	6	
		Cancer		8	2		Homogeneous

Microarray analyses revealed up-regulation of PRDX1, HMGB1, DEK, PABPC1, HSPD1 and NCL and down-regulation of CA2 in cancers. [a]The mean intensity of immunohistochemical staining in the epithelial or tumor cells evaluated relative to the positive controls as follows: weak, 1+; moderate, 2+; strong, 3+. [b]Intra-tumoral heterogeneity in immunohistochemical staining pattern.

assay. Among the 20 up-regulated genes, 6 genes of ribosomal proteins were included; this finding was in agreement with previous observations (7,9,11,18). In addition, relationships with carcinogenesis of other cancers have been reported regarding PRDX1 (28), peptidylprolyl isomerase A (PPIA) (29), ras-related C3 botulinum toxin substrate 1 (RAC1) (30), eukaryotic translation initiation factor 3, subunit 3 γ (EIF3S3) (31) and interferon induced transmembrane protein 3 (IFITM3) (32).

Immunohistochemical staining. The expression status of encoding proteins from 7 genes in IHC is summarized in Table III. The up- and down-regulation of PRDX1, HMGB1, PABPC1, HSPD1, NCL, and CA2 at the protein level was not in conflict with observations at the transcription level (Table III, Fig. 2). PRDX1, HMGB1, HSPD1, NCL and CA2 showed a homogeneous staining pattern in cancer tissues regardless of the region in the tumors, while the others showed a heterogeneous pattern.

Molecular prediction of tumor malignancy. In the differential diagnosis between 20 cancer biopsies and 20 adenoma biopsies by their gene expression profiles using the supervised classification method, the highest prediction accuracy was 97.5% when 8-71 genes were used (Fig. 3). Comparison between intra- and inter-tumor heterogeneity of the profiles using HCA showed that respective paired biopsies from the same tumors tended to join each other. When the selected 71-gene set with the highest accuracy was used, cancer biopsies and adenoma biopsies were clearly separated, and 18 of 20

paired biopsies were clustered side by side though a small number of genes was used (Fig. 4). Among the 71-gene set, COL1A1, COL1A2, and EIF2S2 were also reported in other studies as useful discriminating genes between cancers and adenomas (11).

Discussion

In the present study, we applied comprehensive expression analysis and found that intra-tumor heterogeneity of the gene expression profiles was smaller than inter-tumor heterogeneity, using preoperative endoscopic biopsies of colorectal tumors. We also showed that tumor malignancy could be accurately diagnosed with the profile in a single biopsy. Such accuracy is a promising first step in the clinical application of this technique for various settings such as prediction of the response to preoperative chemo- and/or radio-therapy or the existence of lymph node metastasis. Our findings suggest that this technique can be potentially used to define accurately the biological properties of colorectal tumors.

In CRC, changes in gene expression profiles that occurred during chemotherapy were detected using rectal cancer biopsies (33). The possible prediction of response to preoperative chemo- and/or radio-therapy for rectal cancers by using gene expression profiling of a single biopsy has also been reported (24,25). However, in the application of diagnosis in the clinical field based on gene expression profiling using preoperative endoscopic biopsies, it is not favorable that the profiles of biopsies from a tumor are widely different from each other. In this regard, our results added support to those
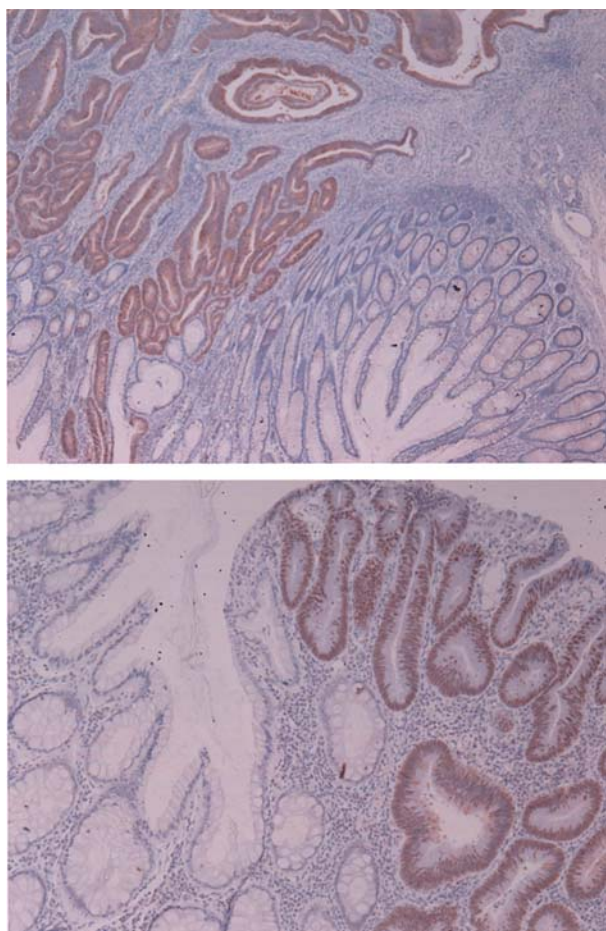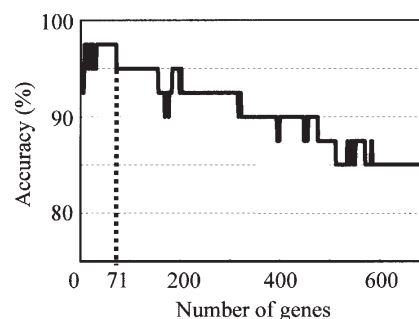
Figure 3. The accuracy curve of the differential diagnosis of 20 cancer biopsies and 20 adenoma biopsies using the supervised classification algorithm. Genes differentially expressed in cancer and adenoma biopsies were ranked according to the signal-to-noise ratio, and the accuracy was calculated with the weighted-votes method and the leave-one-out cross validation. The highest prediction accuracy was 97.5% when 8-71 genes were used.

Figure 2. Immunohistochemical staining of HSPD1 (A) and NCL (B) in colorectal cancer tissues. (A) HSPD1 was expressed strongly in cancer cells while no HSPD1 expression was observed in adjacent normal epithelial cells (x4). (B) NCL was expressed strongly in cancer cells and weakly in normal epithelial cells in the lower portion of the colonic gland (x10).

of previous investigators (24,25), thus further promoting the technique.

Through using small-volume samples of a single biopsy of colorectal tumor, a gene expression profile was successfully obtained and a differentially expressed gene set related to colorectal carcinogenesis was detectable, as when using whole tissue samples. The feasibility of microarray-based study using fine-needle aspiration biopsy (FNAB) samples in some human tumors and endoscopically obtained tissues from precancerous lesions such as Barrett's esophagus was also reported (34,35). These results suggested that low-volume tissue samples such as endoscopic biopsy might give an accurate picture of gene expression in the whole tumor.

The histopathological features are not always homogeneous within a solid tumor. In CRC, the surface and invasive front of the tumor are sometimes histopathologically different. The influence of such morphological intra-tumor heterogeneity on gene expression profiles is not clear, although heterogeneity was detected in individual genes (36). Heterogeneity based on the superficial area of the tumor would have an unfavorable impact on molecular diagnosis when biopsy samples are used. In our study, the gene expression profiles of paired biopsies from the same tumors were identical in almost all tumors. At the protein level, IHC also showed a homogeneous staining
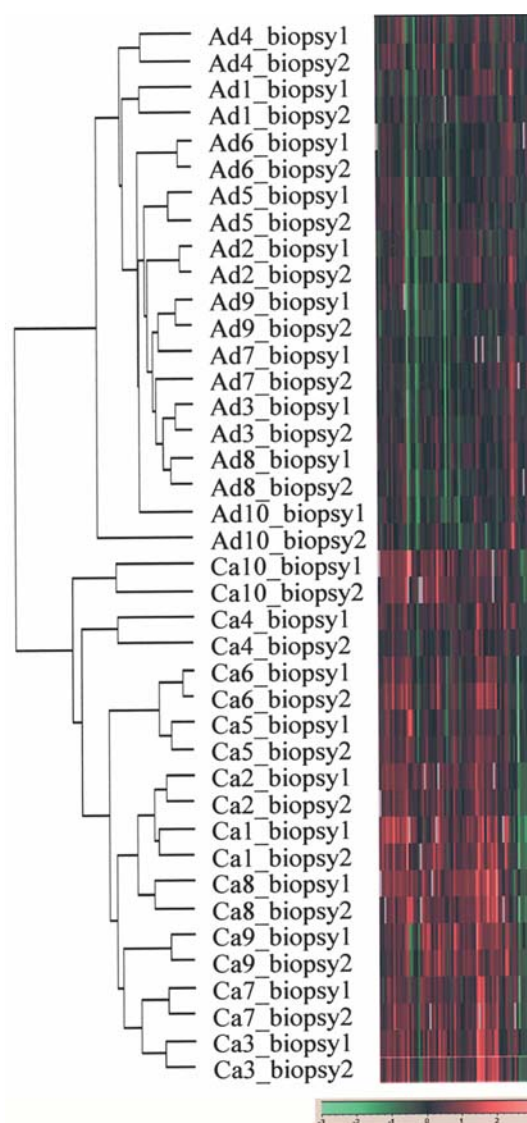

Figure 4. Hierarchical clustering analysis using 71 differentially expressed genes between cancer and adenoma biopsies. Samples consisted of cancer biopsies (n=20; 10 pairs) and adenoma biopsies (n=20; 10 pairs). Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using arithmetic average as the clustering algorithm. Red indicates overexpression, and green indicates underexpression. Cancer biopsies and adenoma biopsies were clearly separated, and 18 of 20 paired biopsies were joined together.

pattern in cancer cells at the superficial position in 5 out of 7 genes. The heterogeneity in some varied genes at the protein level did not affect the intra-tumor homogeneity of the profiles at the transcriptional level. This may be because the use of tens to thousands of genes in profiling compensated for the difference due to heterogeneity in some varied genes or because the data at the transcriptional level do not always parallel those at the protein level (37).

Concerning the intra-tumor heterogeneity of the profiles in other cancers, differences in gene expression between cancer cells in the periphery and those in the center of breast cancers were detected using microdissected samples (38). However, only in a few candidate genes was the difference in the expression levels more than two-fold. In breast cancer, the profile of FNAB was reported to resemble closely that of the corresponding source tumor (39). In a comparison between FNABs and core biopsies, the difference in profiles within the same tumor was not greater than that between other tumors (40). Also in soft tissue sarcomas, the intra-tumor heterogeneity of the profiles was smaller than the inter-tumor heterogeneity (41).

It is not clear how many biopsies are sufficient to represent the biological properties of a tumor with gene expression profiles. The first part of our study indicated that the profile of one biopsy could distinguish cancers from adenomas, beyond the intra-tumor heterogeneity. However, our next pursuit is the differential diagnosis of more delicate and important differences, such as the potential of metastasis and sensitivity to chemo- or radio-therapy. Any unfavorable variation based on the sampling skill and the experiment may be a drawback on the actual clinical application of this method. Multiple biopsies from a tumor may be required to balance the differences within a tumor. The number of biopsies for microarray analysis was studied using biopsies from rectal epithelia and two biopsies per person were recommended, based on the equality of the expression data within a person (42), though similar analyses using biopsies from colorectal tumors have not reported.

Sampling techniques are important. Every biopsy was sampled by a specialist in the colonoscopic unit and very reliable sampling was carried out in this study. Inaccurate sampling would give confusing results showing a wide gap of profiles even between biopsies from one tumor. We took care to avoid contamination of the normal epithelium or adenomatous component on the periphery of the cancer tissue and confirmed the histology with pathological diagnosis of simultaneously obtained biopsies and surgical samples.

Our results suggest that gene expression profiling using endoscopic biopsies can accurately describe the biological properties of colorectal cancer. Further studies of gene expression profiling using preoperative endoscopic biopsies may allow the development of new diagnostic systems for the selection of neoadjuvant therapy or the method most appropriate for tumor resection. Ultimately, it may lead to individualized therapy for colorectal cancer.

References

1. Yiu HY, Whittemore AS and Shibata A: Increasing colorectal cancer incidence rates in Japan. Int J Cancer 109: 777-781, 2004.
2. Dukes CE: The classification of cancer of the rectum. J Pathol Bacteriol 35: 323-332, 1932.
3. Astler VB and Coller FA: The prognostic significance of direct extension of carcinoma of the colon and rectum. Ann Surg 139: 846-852, 1954.
4. Sobin LH and Witedkind CH (eds): TNM Classification of Malignant Tumors (5th edition). Wiley-Liss, New York, 1997.
5. Golub TR, Slonim DK, Tamayo P, *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
6. Ramaswamy S and Golub TR: DNA microarrays in clinical oncology. J Clin Oncol 20: 1932-1941, 2002.
7. Alon U, Barkai N, Notterman DA, *et al*: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA 96: 6745-6750, 1999.
8. Notterman DA, Alon U, Sierk AJ, *et al*: Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 61: 3124-3130, 2001.
9. Kitahara O, Furukawa Y, Tanaka T, *et al*: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 61: 3544-3549, 2001.
10. Takemasa I, Higuchi H, Yamamoto H, *et al*: Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. Biochem Biophys Res Commun 285: 1244-1249, 2001.
11. Lin YM, Furukawa Y, Tsunoda T, *et al*: Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 21: 4120-4128, 2002.
12. Zou TT, Selaru FM, Xu Y, *et al*: Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 21: 4855-4862, 2002.
13. Ichikawa Y, Ishikawa T, Takahashi S, *et al*: Identification of genes regulating colorectal carcinogenesis by using the algorithm for diagnosing malignant state method. Biochem Biophys Res Commun 296: 497-506, 2002.
14. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, *et al*: Gene expression in colorectal cancer. Cancer Res 62: 4352-4363, 2002.
15. Williams NS, Gaynor RB, Scoggin S, *et al*: Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. Clin Cancer Res 9: 931-946, 2003.
16. Komori T, Takemasa I, Higuchi H, *et al*: Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. J Exp Clin Cancer Res 23: 521-527, 2004.
17. Wang Y, Jatkoe T, Zhang Y, *et al*: Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 22: 1564-1571, 2004.
18. 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.
19. Kwon HC, Kim SH, Roh MS, *et al*: Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. Dis Colon Rectum 47: 141-152, 2004.
20. Li M, Lin YM, Hasegawa S, *et al*: Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. Int J Oncol 24: 305-312, 2004.
21. Arango D, Laiho P, Kokko A, *et al*: Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 129: 874-884, 2005.
22. Yamasaki M, Takemasa I, Komori T, *et al*: The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. Int J Oncol 30: 129-138, 2007.
23. Sauer R, Becker H, Hohenberger W, *et al*: Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med 351: 1731-1740, 2004.
24. Ghadimi BM, Grade M, Difilippantonio MJ, *et al*: Effectiveness of gene expression profiling for response prediction of rectal adenocarcinoma to preoperative chemoradiotherapy. J Clin Oncol 23: 1826-1838, 2005.
25. Watanabe T, Komuro Y, Kiyomatsu T, *et al*: Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. Cancer Res 66: 3370-3374, 2006.

26. Okabe S, Shia J, Nash G, *et al*: Lymph node metastasis in T1 adenocarcinoma of the colon and rectum. J Gastrointest Surg 8: 1032-1040, 2004.
27. Ito Y, Yoshida H, Matsuzuka F, *et al*: Expression of the components of the Cip/Kip family in malignant lymphoma of the thyroid. Pathobiology 71: 164-170, 2004.
28. Chang JW, Jeon HB, Lee JH, *et al*: Augmented expression of peroxiredoxin I in lung cancer. Biochem Biophys Res Commun 289: 507-512, 2001.
29. Campa MJ, Wang MZ, Howard B, *et al*: Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin a as potential molecular targets in non-small cell lung cancer. Cancer Res 63: 1652-1656, 2003.
30. Pan Y, Bi F, Liu N, *et al*: Expression of seven main Rho family members in gastric carcinoma. Biochem Biophys Res Commun 315: 686-691, 2004.
31. Saramaki O, Willi N, Bratt O, *et al*: Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. Am J Pathol 159: 2089-2094, 2001.
32. Hisamatsu T, Watanabe M, Ogata H, *et al*: Interferon-inducible gene family 1-8U expression in colitis-associated colon cancer and severely inflamed mucosa in ulcerative colitis. Cancer Res 59: 5927-5931, 1999.
33. Clarke PA, George ML, Easdale S, *et al*: Molecular pharmacology of cancer therapy in human colorectal cancer by gene expression profiling. Cancer Res 63: 6855-6863, 2003.
34. Centeno BA, Enkemann SA, Coppola D, *et al*: Classification of human tumors using gene expression profiles obtained after microarray analysis of fine-needle aspiration biopsy samples. Cancer 105: 101-109, 2005.
35. EL-Seraq HB, Nurqalieva Z, Souza RF, *et al*: Is genomic evaluation feasible in endoscopic studies of Barrett's esophagus? A pilot study. Gastrointest Endosc 64: 17-26, 2006.
36. Kaio E, Tanaka S, Kitadai Y, *et al*: Clinical significance of angiogenic factor expression at the deepest invasive site of advanced colorectal carcinoma. Oncology 64: 61-73, 2003.
37. Nishizaki S, Charboneau L, Young L, *et al*: Proteomic profiling of the NCI-60 cancer cell line using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci USA 100: 14229-14234, 2003.
38. Zhu G, Reynolds L, Crnogorac-Jurcevic T, *et al*: Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer. Oncogene 22: 3742-3748, 2003.
39. Assersohn L, Gangi L, Zhao Y, *et al*: The feasibility of using fine needle aspiration from primary breast cancers for cDNA microarray analyses. Clin Cancer Res 8: 794-801, 2002.
40. Symmans WF, Ayers M, Clark EA, *et al*: Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. Cancer 97: 2960-2971, 2003.
41. Francis P, Fernebro J, Eden P, *et al*: Intratumor versus intertumor heterogeneity in gene expression profiles of soft-tissue sarcomas. Gene Chromosome Cancer 43: 302-308, 2005.
42. Pellis L, Franssen-van Hal NL, Burema J, *et al*: The intraclass correlation coefficient applied for evaluation of data correction, labeling methods, and rectal biopsy sampling in DNA microarray experiments. Physiol Genomics 16: 99-106, 2003.