Identification of the Molecular Mechanisms for Dedifferentiation at the Invasion Front of Colorectal Cancer by a Gene Expression Analysis

Yoshimasa Oku, Takashi Shimoji, Katsunari Takifuji, Tsukasa Hotta, Shozo Yokoyama, Kenji Matsuda, Takashi Higashiguchi, Toshiji Tominaga, Toru Nasu, Koichi Tamura, Masaaki Matsuura, Satoshi Miyata, Yo Kato, Hiroki Yamaue, and Yoshio Miki

Abstract

Purpose: The aim of this study is to identify gene expression signatures that accompany dedifferentiation at the cancer invasion front in colorectal cancer.

Experimental Design: Two types of colorectal cancer were selected. Both types were well-differentiated adenocarcinomas at the superficial lesion. One type showed a dedifferentiated phenotype at the invasion front (type A, 13 samples); the other showed almost no dedifferentiated cancer cells at the invasion front (type B, 12 samples). Laser microdissection was combined with a cDNA microarray analysis to investigate the superficial lesions and the invasion front in colorectal cancers.

Results: Eighty-three genes were differentially expressed between types A and B in the superficial lesions, and the samples of superficial lesions were divided correctly into two clusters by these genes. Interestingly, the samples of the invasion front were also divided into the two same clusters by these genes. The text mining method selected 10 genes involved in potential mechanisms causing dedifferentiation of cancer cells at the invasion front. The potential mechanisms include the networks of transforming growth factor-β, Wnt, and Hedgehog signals. The expression levels of 10 genes were calculated by quantitative reverse transcription-PCR and 8 genes were confirmed to be significantly differentially expressed between two types (P < 0.05). The gene expression profiles of 8 genes divided 12 test cases into two clusters with one misclassification.

Conclusions: The molecular mechanisms constructed with 8 genes from three networks of transforming growth factor-β, Wnt, and Hedgehog signals were found to correlate with dedifferentiation at the invasion front of colorectal cancer.

The hallmarks of malignant transformation are the capabilities of invasion and metastasis. The dedifferentiation and dissociation of cancer cells has been reported to be the first event of invasion and metastasis in experimental studies (1). Well-differentiated carcinomas often lose epithelial differentiation and dissociate, migrate, and disseminate into the host stroma. The metastatic activity or prognostic outcome in colorectal cancer has been reported to be associated with morphologic features at the invasive front of the primary tumor. These morphologic features are called tumor “budding” (2, 3) or tumor “focal dedifferentiation” (4, 5).

The molecular background of this phenomenon has been shown to be associated with cell-to-cell or epithelial-stromal interactions. Brabletz et al. showed that tumor progression is driven by the tumor environment through the Wnt signal pathway, and they fixed this type of dedifferentiation as “environmental” change but not “genetic” based on stepwise acquisition of genetic alterations (6). Several reports showed that the “environmental” factors and cell-to-cell or cell-to-extracellular matrix interactions regulated the dedifferentiation at the cancer invasion front (7, 8).

Ramaswamy et al. showed a gene expression signature that was differentially expressed in metastatic tumors of diverse origins relative to primary cancer (9). They found that the metastatic signature was also expressed in a subset of the primary tumors analyzed, thus leading to the hypothesis that the signature might represent a metastatic program that is encoded in primary tumors designated for metastasis.

In colorectal cancers, Koehler et al. showed with gene expression profiling that metastatic and primary tumors from the same patients always clustered in the same tumor subgroups.
Translational Relevance

Well-differentiated carcinomas often lose epithelial differentiation and dissociate, migrate, and disseminate into the host stroma. The metastatic activity or prognostic outcome in colorectal cancer has been reported to be associated with these morphologic features at the invasive front of the primary tumor. The identification of molecular mechanisms and markers involved in these morphologic features are needed for improved staging and for better assessment of treatment of colorectal cancer. We aimed to determine gene expression signatures and molecular alterations potentially involved in the mechanisms of the dedifferentiation at the colorectal cancer invasion front based on the gene expression of both a tumor superficial lesion and a tumor invasion front from the same colorectal cancer patients. Our results suggest that the molecular signature that genetically controls the phenotype of the dedifferentiated cells at the invasion front may already be acquired at a relatively early stage of tumorigenesis and seen at the tumor superficial lesion before the differentiation of cancer cells is lost. We selected the candidate genes and constructed molecular networks and these results provide new insight into cancer progression of colorectal cancer, and these candidate genes are potential markers for diagnosis of prognosis and metastasis.

Materials and Methods

Samples. Fresh frozen colorectal cancer tissue specimens who had undergone surgery at the Wakayama Medical University were collected from 2004 to 2005 after the patients gave their informed consent. For some cases, specimens at the largest section containing the tumor “superficial lesion” and “invasion front” of 13 sample specimens of type A and 6 of type B were collected. Seven of 13 type A samples and 6 of 12 type B samples contained both tumor “superficial lesion” and “invasion front.” These cases were designated as “learning cases.” Six samples of type A and 6 of type B contained only the “superficial lesion” without an “invasion front,” and these cases were designated as “test cases” (for validation study; Fig. 1A, Table 1).

Laser microdissection. For laser microdissection, frozen colon cancer tissues were cut into 20 to 50 sections (10 μm thick) and mounted on filmed glass slides. To differentially microdissect “superficial lesion” and “invasion front” of 13 sample specimens of learning cases, the tumors were sliced by using a Cryostat (Sakura) and mounted on foil-coated glass slide, 90 FOIL-SL25 (Leica Microsystems). Before the laser microdissection, sections were fixed in 70% ethanol for 1 min at room temperature and stained with hematoxylin for 1 min and eosin for 1 min. After the sections were air dried, cancer cells from the invasive front of cancer and superficial lesions were then selectively microdissected using the Leica LMD System (Leica Microsystems) following the original definition. We detected the lesion for microdissection as a cancer superficial lesion and an invasion front. A “superficial lesion” was defined as a lesion within 1,000 μm from tumor surface without a deep surface and microdissected only cancer tissue within the superficial lesion for every sample (Fig. 1B). The “invasion front” was defined as tumor peripheral lesion bound to normal stroma, except within a mucosal lesion. For type A tumors, we microdissected only dedifferentiated cancer nests at the invasion front. For type B tumors, we microdissected only the peripheral lesion bound to the stroma that were microdissected (Fig. 1B).

In this study, microdissected samples of superficial lesions of type A tumors were named “zone AS,” superficial lesions of type B tumors “zone BS,” invasion front of type A tumors “zone AI,” and invasion front of type B tumors “zone BI.” For 12 test cases, the cancer tissue specimens were microdissected only from tumor superficial lesion.

For 10 patients, normal epithelial cells were also dissected from the normal mucosal tissue obtained 5 to 10 cm from the cancer to obtain a normal tissue reference mix.

RNA extraction and probe synthesis and labeling. Total RNA was extracted from microdissected tissue specimens with the RNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. Samples were treated with RNase-free DNase (Qiagen) on the RNeasy columns. The RNA quality of these samples was then assessed with a 1,000 Bioanalyzer (Agilent Technologies) and showed optimal quality for oligonucleotide microarray analysis based on the integrity of the observed 185 and 28S ribosomal RNA bands. We used a two-round amplification protocol as described previously (12, 13). In brief, an Agilent recommended protocol was modified for two rounds of amplification using the Agilent Low RNA Input Linear Amplification Kit (Agilent Technologies). First, 4 ng total RNA was used for reverse transcription reaction with 17 promoter primer. The first linear amplification was done with T7 RNA polymerase with NTP mix and unlabeled CTP. For the second round of amplification, the product of the first reaction was divided in half, and the second amplification and labeling were done according to the manufacturer’s standard protocol. Consequently, 4 ng total RNA was amplified approximately to 40 μg cRNA.

Hybridization. For each oligonucleotide microarray measurement of expression ratios, the Cy3- and Cy5-diluted labeled cRNAs of cancer and normal tissue reference, respectively, were hybridized with an Agilent Human Custom Oligonucleotide Microarray in hybridization buffer (Agilent Technologies) for 17 h at 60°C. After hybridization, the slides were washed in 6× SSC/0.005% Triton X-102 for 10 min at room temperature and 0.1× SSC/0.005% Triton X-102 for 5 min on ice. The arrays were scanned using Agilent Microarray Scanner.
hybridizations were done using a reversal fluorescent dye in dye-swap experiments to confirm the reproducibility of the experiments.

**Analysis of microarray data.** The Cy3 and Cy5 fluorescent intensities for each spot were determined with an Agilent DNA Microarray Scanner (Agilent Technologies). After locally weighted regression normalization was done (14), the expression ratios were obtained using the Feature Extraction Software (Agilent Technologies) that comes with the scanner.

The data flagged as being of poor quality according to the software.
findings were removed from the analysis. The expression level of each gene in samples relative to that in the reference pool product from a pool that consisted of an equal amount of cRNA from 10 patients was shown as $E = \log_2(C_{Cy5} / C_{Cy3})$. To choose surely differentially expressed genes, we selected ones that showed difference of at least $z0.2$ in the log-ratio scale between the two groups. The genes differentially expressed in different sample groups were identified by supervised learning test using Mann-Whitney $U$ test ($P < 0.05$).

Hierarchical clustering and visualization was done by GeneSpring software version 6.0 (Silicongenetics).

Quantitative real-time RT-PCR was done using the SYBR Green PCR Master Mix and an ABI Prism 7900 Sequence Detection System (Applied Biosystems). Gene expression levels were determined using a standard calibration curve prepared from gene-specific RT-PCR products. All PCRs were done in triplicate.

### Table 1. Characterization of the tumor sample specimens

<table>
<thead>
<tr>
<th>Patients</th>
<th>Age/Gender</th>
<th>Location*</th>
<th>$T$</th>
<th>$N$</th>
<th>$M$</th>
<th>Type</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>67/M</td>
<td>Rb</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>77/M</td>
<td>A</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>64/F</td>
<td>D</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>65/F</td>
<td>S</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>44/F</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>69/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>55/M</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>30/F</td>
<td>Rs</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>54/M</td>
<td>T</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>83/M</td>
<td>A</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>63/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>50/M</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>66/M</td>
<td>Rs</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Test cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>55/M</td>
<td>Rs</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>67/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>73/M</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>64/F</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>57/M</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>52/F</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>66/F</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>75/M</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>65/F</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>76/F</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>59/M</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>82/F</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>

*Location was defined as follows: C, cecum; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; Rs, rectosigmoid; Ra, rectum (above the peritoneal reflection); Rb, rectum (below the peritoneal reflection).

1 According to tumor-node-metastasis classification: T, primary tumor; N, regional lymph node; M, distant metastasis.

2 Type was defined as follows: type A, tumors showed a dedifferentiated phenotype at the cancer invasion front; type B, tumors did not show dedifferentiated phenotype at the cancer invasion front.

ACGTTCCTGTAGATGGCTGCT (forward) and GGCGTAAGGGCCTATATTTGTGTT (reverse) for COL1A2, TGGCGCCATTGGTTCAGCA (forward) and TGGTCCAAGGGCCTATATTTGTGTT (reverse) for DACH1, GAGAGGGCTGCT (forward) and CCGTTGAAAGAGAGGCTGCT (reverse) for PCNA, GAGAGGGCTGCT (forward) and CCGTTGAAAGAGAGGCTGCT (reverse) for PCNA,

**Fig. 2.** Hierarchical cluster analysis based on 83 differently expressed genes between zones AS and BS of learning cases. (A) 13 samples of “type A” tumor; yellow branches, “type B” tumor; white vertical lines, borderlines of major two clusters. A, samples of zones AS and BS were divided into major two clusters, “type A cluster” and “type B cluster.” B, 13 samples of cancer invasion front (zones AI and BI) also were divided into two major clusters, corresponding to samples at the tumor superficial lesion. C, 10 of 12 test cases (not used for gene extraction) were correctly clustered into “type A cluster” and “type B cluster”.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Age/Gender</th>
<th>Location*</th>
<th>$T$</th>
<th>$N$</th>
<th>$M$</th>
<th>Type</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>67/M</td>
<td>Rb</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>77/M</td>
<td>A</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>64/F</td>
<td>D</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>65/F</td>
<td>S</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>44/F</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>69/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>55/M</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>30/F</td>
<td>Rs</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>54/M</td>
<td>T</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>83/M</td>
<td>A</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>63/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>50/M</td>
<td>Ra</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>66/M</td>
<td>Rs</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Test cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>55/M</td>
<td>Rs</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>67/M</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>73/M</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>64/F</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>57/M</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>52/F</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>66/F</td>
<td>T</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>75/M</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>65/F</td>
<td>S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>76/F</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>59/M</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>82/F</td>
<td>S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>
Analysis of real-time RT-PCR data. The gene expression levels between the samples were normalized based on the expression levels of the β-actin as the housekeeping standard control. The fold difference was determined as the log ratio of the normalized value of each tumor sample to normal colon reference mix. The correlation rate of the expression level of the array data and real-time RT-PCR data was calculated with Pearson’s coefficient. The differences between the groups were estimated using the Mann-Whitney U test.

Results

Extraction of differentially expressed genes between each sample group. A gene expression analysis was done on 13 learning cases of colorectal carcinomas to identify differentially expressed genes between each microdissected samples. Eighty-three differentially expressed genes were identified between zones AS and BS, 189 genes between zones AS and AI, and 352 genes between zones AI and BI (Supplementary Figures).

Hierarchical clustering using 83 differentially expressed genes. The results of a hierarchical cluster analysis of 83 genes for zones AS and BS show two major sample clusters “type A cluster” and “type B cluster,” and 13 samples were correctly divided into two clusters (Fig. 2A). This gene set grouped the 13 samples of cancer invasion front (zones AI and BI) into two major clusters, corresponding to the samples at the tumor superficial lesion (Fig. 2B). This result indicates that the characteristics of cancer cells at the tumor superficial lesion were maintained even if cancer cells deeply invaded and the morphologic phenotype has changed. Moreover, a hierarchical cluster analysis of 12 test cases that were not used for gene extraction was done. Ten of 12 samples were correctly clustered (Fig. 2C).

Construction of molecular networks causing dedifferentiation at the cancer invasion front. From the 83 genes, the genes functionally associated with cancer dedifferentiation were selected by text mining. Based on the functional information of the 83 genes and from the findings of scientific reports of Entrez PubMed, Entrez Genes, Online Mendelian Inheritance in Man, and Kyoto Encyclopedia of Genes and Genomes, the 10 genes involved in the potential mechanisms, including the networks of transforming growth factor-β (TGF-β), Wnt, and Hedgehog signals, were selected (Fig. 3). The 10 genes consisted of 5 TGF-β, 3 Wnt, and 2 Hedgehog signal-related genes (Table 2).

Quantitative analysis of candidate genes involved in the functional networks by using real-time RT-PCR. Three networks were constructed involving 10 genes that cause dedifferentiation at the invasion front of colorectal cancer. Quantitative real-time RT-PCR was used to confirm the expression level of these 10 genes. The expression levels of RT-PCR data were highly correlated with the microarray data. The Pearson’s correlation rates were 0.72 to 0.95 (Table 2). Eight of 10 genes showed significant difference in expression level between types A and B in learning cases (P < 0.05; Table 2). Next, in the test cases, real-time RT-PCR of these 8 genes involved in the three networks was done; then, a cluster analysis of the test cases was done using the real-time RT-PCR data. The results divided the two clusters (type A cluster and type B cluster) with only one misclassification (Fig. 4).

Discussion

Gene expression profiling was applied to matched colorectal superficial lesions and invasion fronts obtained by laser microdissection. Several studies reported the results of global expression analysis of differently located tissues within the same tumor, for example, breast cancer and melanoma,
We calculated that five TGF-β-related genes we selected are related to cancer progression. We specifically focused on the genes. Moreover, there are several reports describing that these genes are involved in dedifferentiation processes, and it is possible that this property may already be acquired at a relatively early stage of tumorigenesis. In this study, we first selected about 80 genes based on microarray data by using the Mann–Whitney U test. In the next process, class-prediction algorithms, such as Support Vector Machine (17) or AdaBoost (18), might be used to select more distinctive genes. However, in this study, we employed a text-mining approach based on the information regarding molecular functions to exclude any uncertain data. In this study, we constructed three networks based on information from known signal pathways and the interactions between the genes. Moreover, there are several reports describing that genes we selected are related to cancer progression. We speculate that five TGF-β-related genes were involved in dedifferentiation at the colorectal cancer invasion front (Fig. 3A). DACH1 bound to endogenous NCoR and Smad4 in cultured cells; Smad4 was required for the DACH1 repression of TGF-β induction of Smad signaling, and the TGF-β induction of Smad signaling was repressed (19, 20). CITED1 interacts with both the DNA-binding Smad proteins and the p300/CBP coactivators through its NH$_2$- and COOH-terminal regions, respectively, while also enhancing the functional link between Smads and p300/ CBP (21). Two transcription products, COL1A2 and CDKN2B were highly expressed in type A tumors. COL1A2 encodes one of the chains for type I collagen, the fibrillar collagen found in most connective tissues. Smad3 and Smad4, the transcription factors of TGF-β signaling, bind to the promoter lesion of COL1A2 and activated the stimulation of COL1A2 (22). CDKN2B encodes a cyclin-dependent kinase inhibitor, which forms a complex with CDK4 or CDK6, and prevents the activation of the CDK kinases; therefore, the encoded protein functions as a cell growth regulator that controls cell cycle G$_1$ progression. PCNA is a protein known to progressively accumulate from stage G$_1$ to M of the cell cycle and disappear at the end of mitosis; it has been postulated to be a cell cycle regulator protein and shows deregulated expression in some neoplasms (23, 24). We speculate that the up-regulation of CDKN2B in type A tumors causes the low-level expression of PCNA. There are several reports that directly or indirectly support this view of the relation between proliferation and cancer dedifferentiation in colorectal cancer.

**Table 2.** List of 10 genes selected from the text mining method with a verification of the array experiments by quantitative real-time PCR in learning cases

<table>
<thead>
<tr>
<th>Network</th>
<th>Symbols</th>
<th>Gene name</th>
<th>CR*</th>
<th>P †</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGF-β</td>
<td>CITED1</td>
<td>CBP/p300-interacting transactivator with Glu/Asp-rich COOH-terminal domain 1</td>
<td>0.91</td>
<td>0.147</td>
</tr>
<tr>
<td>TGF-β</td>
<td>DACH1</td>
<td>Dachshund homologue 1 (Drosophila)</td>
<td>0.74</td>
<td>0.017</td>
</tr>
<tr>
<td>TGF-β</td>
<td>CDKN2B</td>
<td>Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)</td>
<td>0.92</td>
<td>0.014</td>
</tr>
<tr>
<td>TGF-β</td>
<td>COL1A2</td>
<td>Collagen, type I, α2</td>
<td>0.94</td>
<td>0.014</td>
</tr>
<tr>
<td>TGF-β</td>
<td>PCNA</td>
<td>Proliferating cell nuclear antigen</td>
<td>0.72</td>
<td>0.025</td>
</tr>
<tr>
<td>WNT</td>
<td>SFRP4</td>
<td>Secreted frizzled-related protein 4</td>
<td>0.85</td>
<td>0.004</td>
</tr>
<tr>
<td>WNT</td>
<td>NKD1</td>
<td>Naked cuticle homologue 1 (Drosophila)</td>
<td>0.95</td>
<td>0.004</td>
</tr>
<tr>
<td>WNT</td>
<td>VIM</td>
<td>Vimentin</td>
<td>0.78</td>
<td>0.090</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>GLI1</td>
<td>GLI-Kruppel family member GLI1</td>
<td>0.88</td>
<td>0.002</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>GLI2</td>
<td>GLI-Kruppel family member GLI2</td>
<td>0.87</td>
<td>0.017</td>
</tr>
</tbody>
</table>

*Pearson’s correlation rate.
†Mann–Whitney U test.

Microdissection processes, especially for zone A1, were very difficult and time-consuming skill because dedifferentiated cancer cell nests are very small and scattered in the stroma. This great effort produced only 100 to 200 ng total RNA from each sample. As a result, we adopted a two-round amplification method that made the experiment possible with a small amount of total RNA. This protocol amplified 4 ng total RNA approximately to 40 μg cRNA. The reproducibility of the experiment was confirmed by the strong correlation between the microarray data of dye-swap examinations. In addition, we showed that the expression level determined by the microarray correlated strongly with expression level determined by real-time RT-PCR (Table 2), indicating that these microarray analyses were reliable.

Microarray analysis is a powerful tool, but sometimes it can yield uncertain data. In this study, we first selected about 80 genes based on the microarray data by using the Mann-Whitney U test. In the next process, class-prediction algorithms, such as Support Vector Machine (17) or AdaBoost (18), might be used to select more distinctive genes. However, in this study, we employed a text-mining approach based on the information regarding molecular functions to exclude any uncertain data and select the true candidates involved in functional mechanisms that cause dedifferentiation at the cancer invasion front.

We constructed these three networks based on information from known signal pathways and the interactions between the genes. Moreover, there are several reports describing that genes we selected are related to cancer progression. We speculate that five TGF-β-related genes were involved in dedifferentiation at the colorectal cancer invasion front (Fig. 3A). DACH1 bound to endogenous NCoR and Smad4 in cultured cells; Smad4 was required for the DACH1 repression of TGF-β induction of Smad signaling, and the TGF-β induction of Smad signaling was repressed (19, 20). CITED1 interacts with both the DNA-binding Smad proteins and the p300/CBP coactivators through its NH$_2$- and COOH-terminal regions, respectively, while also enhancing the functional link between Smads and p300/CBP (21). Two transcription products, COL1A2 and CDKN2B were highly expressed in type A tumors. COL1A2 encodes one of the chains for type I collagen, the fibrillar collagen found in most connective tissues. Smad3 and Smad4, the transcription factors of TGF-β signaling, bind to the promoter lesion of COL1A2 and activated the stimulation of COL1A2 (22). CDKN2B encodes a cyclin-dependent kinase inhibitor, which forms a complex with CDK4 or CDK6, and prevents the activation of the CDK kinases; therefore, the encoded protein functions as a cell growth regulator that controls cell cycle G$_1$ progression. PCNA is a protein known to progressively accumulate from stage G$_1$ to M of the cell cycle and disappear at the end of mitosis; it has been postulated to be a cell cycle regulator protein and shows deregulated expression in some neoplasms (23, 24). We speculate that the up-regulation of CDKN2B in type A tumors causes the low-level expression of PCNA. There are several reports that directly or indirectly support this view of the relation between proliferation and cancer dedifferentiation in colorectal cancer.

**Fig. 4.** Hierarchical cluster analysis of 12 test cases based on expression levels of 8 genes measured by quantitative real-time RT-PCR. Expression levels of 8 genes divided 12 samples into two major clusters corresponding to “type A cluster” and “type B cluster” with one misclassification.
(25–27). Palmqvist et al. showed the significant association between low tumor cell proliferation at the invasive margin and poor prognosis (27). Taniyama et al. reported that low tumor cell proliferative activity is correlated with areas of low differentiation and that decreased proliferation in diploid tumors is correlated with increased numbers of lymph node metastases (26, 27). The gene expression profile is consistent with these previous reports and suggest that type A tumors had a lower proliferation than type B tumors at the cancer superficial lesion before differentiation of the cancer cells were lost.

We selected 3 genes associated with the WNT signal pathway (Fig. 3B). SFRP4 is a member of the SFRP family that contains a cysteine-rich domain homologous to the putative Wnt-binding site of Frizzled proteins. SFRPs act as soluble modulators of Wnt signaling (28). NKD1 is a Dishevelled-binding protein that functions as a negative regulator of the Wnt-β-catenin-Tcf signaling pathway (29). VIM, transcription product of Wnt-β-catenin-Tcf signaling pathway, was up-regulated in type A tumors in comparison with type B tumors (30). In type A tumors, COL1A2, which is an extracellular matrix component produced through the TGF-β pathway, was highly expressed in comparison with type B tumors. High levels of type 1 collagen in metastatic tumors have also been reported (9, 31). Blabl et al. reported that collagen type 1 is highly expressed at the invasion front of colorectal cancer (32). Collagen type 1 induces invasive phenotypes and changes in the expression pattern of β-catenin and E-cadherin in vitro. This expression pattern indicates the nuclear accumulation of β-catenin, thus leading to the TCF/LEF-1 transcription activity in the dedifferentiated cancer cell at the invasion front (6, 32). Expression levels of VIM, transcription product of Wnt signal pathway, seem to be higher in type A tumors in comparison with type B tumors, although the coefficient of correlation between the array data and RT-PCR tumors in comparison with type B tumors, although the product of Wnt signal pathway, seem to be higher in type A invasion front (6, 32). Expression levels of VIM, transcription product of Wnt-β-catenin-Tcf signaling pathway (29). VIM, transcription product of Wnt signaling (28). NKD1 is a Dishevelled-binding protein that functions as a negative regulator of the Wnt-β-catenin-Tcf signaling pathway (29). VIM, transcription product of Wnt-β-catenin-Tcf signaling pathway, was up-regulated in type A tumors in comparison with type B tumors (30). In type A tumors, COL1A2, which is an extracellular matrix component produced through the TGF-β pathway, was highly expressed in comparison with type B tumors. High levels of type 1 collagen in metastatic tumor have also been reported (9, 31). Blabl et al. reported that collagen type 1 is highly expressed at the invasion front of colorectal cancer (32). Collagen type 1 induces invasive phenotypes and changes in the expression pattern of β-catenin and E-cadherin in vitro. This expression pattern indicates the nuclear accumulation of β-catenin, thus leading to the TCF/LEF-1 transcription activity in the dedifferentiated cancer cell at the invasion front (6, 32). Expression levels of VIM, transcription product of Wnt signal pathway, seem to be higher in type A tumors in comparison with type B tumors, although the coefficient of correlation between the array data and RT-PCR data of VIM was not so high. VIM is type III intermediate filament normally expressed in cells of mesenchymal origin.

VIM has indeed been described in migratory epithelial cells involved in tumor invasion (33, 34). The de novo expression of VIM is also frequently associated with an epithelial-mesenchymal transition and the metastatic conversion of epithelial cells (35). The gene expression profile in this study is consistent with these findings, thus suggesting that type A tumors had a low proliferation but later acquired a high invasive activity; as a result, the cancer cells became dissociated, migrated, and then disseminated into the stroma at the cancer invasion front.

Two Gli genes, GLI1 and GLI2, were up-regulated in type A tumors in comparison with type B tumors. These genes are transcription factor of Hedgehog signal pathway. However, activation of the Hedgehog signal pathway in colorectal cancer is controversial (36, 37).

These networks may help us to obtain a better understanding of the key biological functions that cause dedifferentiation at the cancer invasion front. However, the functional validation of such chosen candidate genes still remains insufficient. Additional experiments are therefore necessary to validate the functional networks that we constructed.

Recently, several text mining tools have been proposed for systemic functional analysis of gene expression signatures. These systems provided us with enormous functional information but were not perfect because they sometimes contained unreliable information and sometimes missed important information. We therefore combined the data by several text mining tools and the information obtained manually by scientific reports from Entrez PubMed, Entrez Genes, Online Mendelian Inheritance in Man, and Kyoto Encyclopedia of Genes and Genomes. Because it seems that the information selected by the text mining method in this study is merely part of the functional mechanisms, more effective methods for selecting functional information from the microarray data should thus be developed and a further research and information should thus be accumulated.

To understand complex biological processes, such as cancer initiation and progression, it is important to consider differential gene expression in the context of complex molecular networks. This is the first report to construct functional networks from a transcriptome analysis associated with dedifferentiation at the cancer invasion front of colorectal cancer. Eight genes from three networks divided two sample groups of types A and B in a hierarchical cluster analysis. This may indicate that three networks, and not a single network, caused dedifferentiation with complex signal cross-talk. For example, collagen type 1, a transcriptional product of the TGF-β signal, caused nuclear accumulation of β-catenin (32). This is only part of the information. Further studies are necessary to more fully elucidate understand the signal cross-talk mechanism.

The aim of this study was to identify the functional mechanism involved in the dedifferentiation at the invasion front of colorectal cancer. The final goal is the understanding of the molecular mechanism of dedifferentiation at the invasion front of colorectal cancer and predicting the dedifferentiation based on the knowledge of the molecular mechanism. Moreover, we indicated the possibility that the genes selected in this study might be clinical markers to predict the patient prognosis and metastasis in colorectal cancer. Further study using a large number of patients is necessary before any definitive conclusions can be made; however, the findings of this study are considered to positively contribute to the elucidation of molecular mechanism of dedifferentiation at the invasion front of colorectal cancer and the establishment of new clinical markers.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank all members of the Genome Center, Japanese Foundation for Cancer Research for helpful discussions and encouragement.

References

4. Ono M, Sakamoto M, Ino Y, et al. Cancer cell morphology at the invasive front and expression of...
cell adhesion-related carbohydrate in the primary le-
sion of patients with colorectal carcinoma with liver
5. Tominaoka N, Nakashiyama Y, Nimura S, Yoshimura Y, 
Sakai Y, Shimoda T. Predictive histopathological factors 
for lymph node metastasis in patients with nonpul-
culated submucosal invasive colorectal cancer. Dis 
expression in colorectal cancers indicates tumor pro-
gression driven by the tumor environment. Proc Natl 
7. Pycke C, Salo S, Raffkiaer E, Romer J, Dano K, 
Ttryggvason K. Laminin-5 is a marker of invading 
cancer cells in some human carcinomas and is coex-
pressed with the receptor for urokinase plasminogen 
avtivator in budding cancer cells in colon adeno-carcinoma. 
bution of CD44 variant 6 and nuclear h-catenin ex-
pression to the formation of budding cancer cells in 
patient with T colorectal cancer. Cancer 2003;92: 
2539–46.
sion profiling of colorectal cancer and metastasis 
drives tumors according to their clinicopathological 
synthesized novel microarray optimized for mouse 
stem cell and early developmental expression profiling. 
pathway characterized by gene expression analysis of 
supraclavicular lymph node metastasis-positive breast 
cDNA microarray data: a robust composite method 
addressing single and multiple slide systematic vari-
15. Roesch A, Vogt T, Stoltz W, Dugas M, Landthaler M, 
Becher V. Discrimination between gene expression 
patterns in the invasive margin and the tumour core 
of malignant melanomas. Melanoma Res 2003;13: 
503–9.
Combination of microdissection and microarray analy-
sis to identify gene expression changes between dif-
ferentially located tumour cells in breast cancer. 
17. Brown MP, Grundy WN, Lin D, et al. Knowledge-
based analysis of microarray gene expression data by 
using support vector machines. Proc Natl Acad Sci 
18. Ben-Dor A, Bruhn L, Friedman N, Nachman I, 
Schummer M, Yakhini Z. Tissue classification with gene 
19. Li X, Perissi V, Liu F, Rose DW, Rosenfeld MG. Tissue 
specific regulation of retinal and pituitary pre-
20. Tahata T, de Ceasterke MP, Lechleider RJ, et al. The 
MSGI non-DNA-binding transactivator binds to the 
p300/CBP coactivators, enhancing their functional 
link to the Smad transcription factors. J Biol Chem 
Synergistic cooperation between SPI and Smad3/
Smad4 mediates transforming growth factor h signal-
ing of a2(1)-collagen (Col1A2) transcription. 
22. Cels JE, Cels A. Cell cycle-dependent variations in 
the distribution of the nuclear protein cyclin/pro-
eliferating cell nuclear antigen in cultured cells. 
cell nuclear antigen (PCNA) immunolocalization in 
paraffin sections: an index of cell proliferation with 
evidence of deregulated expression in some neo-
Williams AC. Cell nuclear antigen (PCNA) immunological 
junction of the nuclear protein cyclin/pro-
eliferating cell nuclear antigen in cultured cells. 
25. Li X, Perissi V, Liu F, Rose DW, Rosenfeld MG. 
Synergistic cooperation between SPI and Smad3/
Smad4 mediates transforming growth factor h signal-
ing of a2(1)-collagen (Col1A2) transcription. 
26. Iacobuzio-Donahue CA, Argani P, Hempen PM, 
Jones J, Kern SE. The desmoplastic response to 
infiltrating breast carcinoma: gene expression at 
the site of primary invasion and implications for compari-
sions between tumor types. Cancer Res 2002;62: 
5351–7.
27. Sommers CL, Byers SW, Thompson EW, Tori JA, 
Gelmann EP. Differentiation state and invasiveness of 
28. Steiner PM, Roop DR. Molecular and cellular biolo-
y of intermediate filaments. Annu Rev Biochem 
1988;57:593–625.
29. Berman DM, Karhadkar SS, Maitra A, et al. Wid-
spread requirement for Hedgehog ligand stimulation 
429:846–51.
30. Qualtrough D, Buda A, Gaffield W, Williams AC, 
Parsakova C. Hedgehog signaling in colorectal tumor 
cells: induction of apoptosis with cycloamine treat-

www.aacrjournals.org
Clinical Cancer Research

Identification of the Molecular Mechanisms for Dedifferentiation at the Invasion Front of Colorectal Cancer by a Gene Expression Analysis

Yoshimasa Oku, Takashi Shimoji, Katsunari Takifuji, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/14/22/7215

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2008/11/14/14.22.7215.DC1

Cited articles
This article cites 37 articles, 13 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/14/22/7215.full.html#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.