Gene Expression Profile of Gastric Carcinoma: Identification of Genes and Tags Potentially Involved in Invasion, Metastasis, and Carcinogenesis by Serial Analysis of Gene Expression

Naohide Oue, Yoichi Hamai, Yoshitsugu Mitani, Junji Matsumura, Yasuhiro Oshimo, Phyu Phyu Aung, Kazuya Kuraoka, Hirofumi Nakayama, and Wataru Yasui

Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan

ABSTRACT

Gastric carcinoma (GC) is one of the most common malignancies worldwide. To better understand the genetic basis of this disease, we performed serial analysis of gene expression (SAGE) on four primary GC samples and one associated lymph node metastasis. We obtained a total of 137,706 expressed tags (Gene Expression Omnibus accession number GSE545, SAGE Hiroshima gastric cancer tissue), including 38,903 that were unique. Comparing tags from our GC libraries containing different stages and different histologies, we found several genes and tags that are potentially involved in invasion, metastasis, and carcinogenesis. Among these, we selected 27 genes and measured mRNA expression levels in an additional 46 GC samples by quantitative reverse transcription-PCR. Frequently overexpressed genes (tumor/normal ratio > 2) were COL1A1 (percentage of cases with overexpression, 78.3%), CDH17 (73.9%), APOC1 (67.4%), COL1A2 (58.7%), YFL13H12 (52.2%), CEACAM6 (50.0%), APOE (50.0%), REGIV (47.8%), S100A11 (41.3%), and FUS (41.3%). Among these genes, mRNA expression levels of CDH17 and APOE were associated with depth of tumor invasion (P = 0.0060 and P = 0.0139, respectively), and those of FUS and APOE were associated with degree of lymph node metastasis (P = 0.0416 and P = 0.0066, respectively). In addition, mRNA expression levels of SAGE were associated with stage (P = 0.0156, P = 0.0395, and P = 0.0125, respectively). Quantitative reverse transcription-PCR analysis also showed a high level of REGIV expression (>100 arbitrary units) in 14 of 46 GC samples (30.4%) but not in noncancerous tissues. We detected V5-tagged RegIV protein in the culture media of cells transfected with pcDNA-RegIV-V5 by Western blot. Our results provide evidence that REGIV may serve as a specific biomarker for GC.

INTRODUCTION

Gastric carcinoma (GC) is one of the most common human cancers. Despite improvements in cancer therapy, ~650,000 patients with GC die/year (1). A variety of genetic and epigenetic alterations are associated with GC (2—4). However, the underlying mechanism of gastric carcinogenesis is still poorly understood. To identify potential molecular markers for GC and to better understand the development of GC at the molecular level, comprehensive gene expression analysis may be useful. Although several large-scale gene expression studies with cDNA or oligonucleotide arrays have been performed in GC (5—8), they have used different platforms that varied in the number and identity of genes printed on them. On the other hand, serial analysis of gene expression (SAGE) analyzes 14-bp tags derived from defined positions of cDNAs without a priori knowledge of the sequence of the genes expressed (9). Thus, SAGE offers an unbiased, comprehensive gene expression profiling approach. Recently, three SAGE studies of GC were reported, and several up-regulated and down-regulated genes were identified (10—12). However, only one (10) or two samples (11, 12) were examined, and the relation to invasion and metastasis was not analyzed. In the present study, we performed SAGE analysis on four samples of GC of different stages and different histologies. In addition, we performed SAGE analysis on one lymph node metastasis of GC. We report here the identification of several genes and tags potentially involved in invasion, metastasis, and carcinogenesis of GC. Among these, we focused on the REGIV gene because this gene is frequently overexpressed in GC, and REGIV expression is narrowly restricted in noncancerous tissues. In addition, the amino acid sequence of the RegIV protein suggests that it may be secreted.

MATERIALS AND METHODS

Tissue Samples. For SAGE analysis, four primary GC samples and 1 associated lymph node metastasis were used (Table 1). We confirmed microscopically that the tumor specimens consisted mainly (>80%) of carcinoma tissue with the exception of S219T. For quantitative reverse transcription-PCR (RT-PCR), 46 GC samples and corresponding nonneoplastic mucosa samples were used. Of the 46 GC samples, lymph node metastasis samples were available for 9. The samples were obtained at surgery at Hiroshima University Hospital and affiliated hospitals. Noncancerous samples of the heart, aorta, lung, tongue, esophagus, stomach, duodenum, ileum, colon, liver, gallbladder, pancreas, kidney, urinary bladder, thyroid gland, adrenal gland, spleen, skin, endometrium, and lymph node were obtained from a 28-year-old woman diagnosed with multiple sclerosis. Samples were frozen immediately in liquid nitrogen and stored at −80°C until use. Histological classification of GC was performed according to the Lauren classification system (13). In addition, diffuse-type GC samples were additionally classified into diffuse-adherent and diffuse-scattered subtypes (14). Tumor staging was carried out according to the Tumor-Node-Metastasis Stage Grouping (15).

SAGE. SAGE was performed according to SAGE protocol version 1.0c, June 23, 2000. Tags were extracted from the raw sequence data with SAGE2000 analysis software version 4.12 kindly provided by Dr. Kenneth W. Kinzler (The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine). Clinicopathological details of the 5 samples are shown in Table 1. To identify genes involved in tumor progression, we analyzed 2 GC samples (W226T and W246T). Both samples were classified as intestinal type GC. However, W226T was early, and W246T was advanced. Early GC is limited to the mucosa or the mucosa and submucosa, regardless of nodal status (16). We confirmed microscopically that these 2 samples showed similar histological features (Fig. 1). To identify genes involved in tumor metastasis, we analyzed 1 GC sample (P208T) and its lymph node metastasis (P208L). Histologically, these samples were classified as diffuse-adherent type, and we confirmed microscopically that both the primary tumor (P208T) and the metastatic tumor (P208L) contained few stromal cells and lymphocytes (Fig. 1). Scirrhous-type GC belongs to the diffuse-scattered type, often occurs in young women, and is characterized by extensive fibrous stroma, infiltrative and rapid growth, and poor prognosis (17). Sample S219T was a scirrhous-type GC showing scattering growth in an abundant fibrous stroma (Fig. 1). To
permit direct comparison, each library was normalized to a total of 1,000,000 tags.

Cluster Analysis. The Cluster and TreeView computer programs were obtained from online resources.1 We compared SAGE tags from 4 primary GC samples with those from samples of normal gastric epithelia [GSM784, SAGE normal gastric body epithelial (10)], available from SAGEmap (18).2 We also compared SAGE tags from 2 primary GC samples, also available from SAGEmap [GSM757, SAGE gastric cancer-G234 (10) and GSM2385, SAGE gastric cancer-G189] with those from normal gastric epithelia (GSM784) and obtained the 20 most up-regulated and 20 most down-regulated tags. This produced a dataset of 128 tags. These data were imported into the Cluster program and were log-transformed, and complete linkage clustering was performed.

Quantitative RT-PCR Analysis. Total RNA was extracted with an RNeasy Mini kit (Qiagen, Hilden, Germany), and 1 µg of total RNA was converted to cDNA with a first-strand cDNA synthesis kit (Amersham Pharmacia Biotech, Uppsala, Sweden). PCR was performed with a SYBR Green PCR Core Reagents kit (Applied Biosystems, Foster City, CA). Real-time detection of the emission intensity of SYBR green bound to double-stranded DNA was performed with an ABI PRISM 7700 Sequence Detection System (Applied Biosystems) as described previously (19). The sequences primer are listed in Supplementary Table 1. We calculated the ratio of target gene mRNA expression levels between GC tissue (T) and corresponding nonneoplastic mucosa (N). T/N ratios > 2-fold were considered to represent overexpression. Genes with T/N ratios > 2 in >40% of the samples examined were defined as frequently up-regulated genes.

Cell Lines, Expression Vector, and Western Blot. Two cell lines derived from human GC were used. MKN-28 was kindly provided by Dr. Toshimitsu Suzuki. HSC-39 was kindly provided by Dr. Kazuyoshi Yanagihara (20). All cell lines were maintained in RPMI 1640 (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) containing 10% fetal bovine serum (BioWhittaker, Walkersville, MD) in a humidified atmosphere of 5% CO2 and 95% air at 37°C. Culture media was concentrated with a PROTEIN concentrate kit (Takara Bio, Indianapolis, IN) according to the manufacturer's instructions. For Western blot detection system (Amersham Pharmacia Biotech).

Statistical Methods. Statistical analyses were performed with the Mann-Whitney U test. P of <0.05 was regarded as statistically significant.

RESULTS

Generation of SAGE Data. A total of 137,706 tags was generated, including 38,903 that were unique. The numbers of tags and unique tags are shown in Table 1. Sequence data from our SAGE libraries are publicly available at SAGEmap (GEO accession number GSE 545, SAGE Hiroshima gastric cancer tissue).

Comparison of Expression Patterns in GCs and Normal Stomach. We compared SAGE tags from 4 primary GC samples with those from normal gastric epithelia (GSM784). The 20 most up-regulated tags and 20 most down-regulated tags in each GC are shown in Supplementary Table 2. Among the up-regulated tags, 12 were commonly up-regulated in both W226T (intestinal type) and W246T (intestinal type). These tags included lysozyme (LYZ), trefoil factor 3 (TFF3), aldolase A (ALDOA), and S100 calcium-binding protein, which may participate in the genesis of intestinal type GC. P208T (diffuse-adherent type) and S219T (diffuse-scattered type) showed many different tags from those of W226T and W246T. The down-regulated tags were similar in all 4 GC samples and included lipase (LIPF), pepsinogen (PGA5), and antrum mucosa protein (AMP18), which are expressed physiologically in normal gastric glands.

The SAGE data were also analyzed by a clustering algorithm to delineate patterns in the expression of 128 tags among all four libraries (our four GC libraries, our one lymph node metastasis library, and three libraries available from SAGEmap2; Fig. 2). These tags were selected as described in “Materials and Methods.” Clusters of coexpressed tags suggested that the two intestinal type GC libraries (W226T and W246T), despite being derived from 2 different patients at different stages, were the most similar to each other. The primary GC (P208T) and its lymph node metastasis (P208L) appeared not to be similar to each other. To identify ideal biomarkers for GC, we focused on a cluster of 14 tags, the expression of which was up-regulated in 6 GC samples (our four GC libraries plus two GC libraries available from SAGEmap) and 1 lymph node metastasis. Because some genes share the same SAGE tag, these 14 tags represented 22 genes (Fig. 2 and Table 2). To validate the SAGE data, the expression of 12 known genes was analyzed by quantitative RT-PCR of an additional 46 GC samples and corresponding nonneoplastic mucosa samples. Frequently overexpressed genes were APOC1 (percentage of samples with overexpression: T/N ratio > 2; 67.4%), YF13H12 (52.5%), and CEACAM6 (50.0%; Fig. 3A, see also Supplementary Fig. 1). Other genes were less frequently overexpressed. The expression levels of all 12 genes were not associated with T grade (depth of tumor invasion), N grade (degree of lymph node metastasis), or tumor stage.

Comparison of Expression Patterns in Early and Advanced GC. To identify genes involved in tumor progression, we compared tags from early GC (W226T) and advanced GC (W246T). The 10
most up-regulated tags are shown in Table 3, and the 10 most
down-regulated tags are shown in Supplementary Table 3. Because
some genes share the same SAGE tag, these up-regulated 10 tags
represented 12 genes (Table 3). To validate the SAGE data, the
expression of 9 known genes was analyzed by quantitative RT-PCR of
an additional 46 GC samples and corresponding nonneoplastic mu-
cosa samples. Genes frequently overexpressed in GC compared with
nonneoplastic mucosa were COL1A1 (78.3%), CDH17 (73.9%),
COL1A2 (58.7%), and FUS (41.3%; Fig. 3B, see also Supplementary
Fig. 2). Other genes were less frequently overexpressed. The mRNA
expression levels of CDH17 were associated with T grade
\( P = 0.0060 \). The mRNA expression levels of FUS were associated
with N grade \( P = 0.0416 \). The mRNA expression levels of FUS,
COL1A1, and COL1A2 were associated with tumor stage
\( P = 0.0414, P = 0.0156, \) and \( P = 0.0395 \), respectively; Table 4).

Comparison of Expression Patterns in Primary GC and Asso-
ciated Lymph Node Metastasis. To identify genes involved in tumor
metastasis, we compared tags from primary GC (P208T) and its
lymph node metastasis (P208L). The 10 most up-regulated tags are
shown in Table 5, and the 10 most down-regulated tags are shown in
Supplementary Table 4. The up-regulated tags represented 12 genes
(Table 5). To validate the SAGE data, the expression of 5 known
genes was analyzed by quantitative RT-PCR of an additional 46 GC
samples and their lymph node metastases in 9 samples. A frequently
overexpressed gene in lymph node metastasis compared with primary
GC was not found (Fig. 3C, see also Supplementary Fig. 3). APOE
mRNA expression in lymph node metastasis tended to be higher than
that in primary GC. Other genes were less frequently overexpressed.
Genes frequently overexpressed in GC compared with nonneoplastic

Fig. 1. Histological features of gastric carcino-
ma (GC) samples analyzed by serial analysis of
gene expression. Formalin-fixed, paraffin-embed-
ded sections were stained with H&E. Both W226T
(A) and W246T (B) were to intestinal type GC, and
histological features were similar. Both P208T (C)
and P208L (D) were diffuse-adherent type GC.
P208L was a lymph-node metastasis of P208T.
S219T (E) was diffuse-scattered type GC. Obvious
histological heterogeneity was not seen in all spec-
imens (original magnification, ×100).

Fig. 2. Cluster analysis of 128 tags form eight serial analysis of gene expression
(SAGE) libraries and dendrogram showing similarities in expression patterns among
libraries. Tags were selected as described in the “Materials and Methods.” Brackets
indicate the cluster of tags commonly up-regulated in gastric carcinoma, which is
expanded in size on the right for visualization. On the dendrogram, two intestinal type
gastric carcinoma samples cluster together, indicating their high degree of similarity. Each
color represents a tag, whereas each column corresponds to a SAGE library sample. The
absolute abundance of the SAGE tag in the library (SAGE tag number) correlates with the
intensity of the red color (black, not present; intense red, highly abundant).
mucosa were APOE (50.0%) and S100A11 (41.3%; Fig. 3D). The mRNA expression levels of APOE were associated with T grade (P = 0.0139), N grade (P = 0.0006), and tumor stage (P = 0.0125; Table 4).

**REGIV Overexpression in GC.** Among the 20 up-regulated tags in each GC sample (Supplementary Table 2), we focused on REGIV because REGIV expression was narrowly restricted by Virtual Northern analysis by SAGEmap (Fig. 4A). Besides GCs, REGIV was detected at low levels in only eight libraries, including one colon cancer and two normal colon libraries. Quantitative RT-PCR analysis showed overexpression of the REGIV gene in 22 samples of the 46 GC samples (47.8%; Fig. 4B). When we focused on REGIV gene expression in GC, high levels of REGIV expression (>100 arbitrary units) were found in 14 of 46 samples (30.4%; Fig. 4C). Among various normal tissues obtained from an autopsy, obvious REGIV expression was found in noncancerous stomach, duodenum, ileum, colon, and pancreas, as reported elsewhere (22). However, the levels of REGIV expression were low (<60 arbitrary units).

Analysis of the amino acid sequence of the RegIV protein suggested that it may be secreted. To investigate whether RegIV is a secreted protein, we performed Western blot analysis of cell extracts and culture media of MKN-28 cells transiently transfected with pcDNA 3.1 or pcDNA-RegIV-V5. With an anti-V5 antibody, we detected an approximate M r 20,000 band corresponding to V5-tagged RegIV protein in cell extracts and culture media from RegIV-V5-expressing MKN-28 cells but not in control cells (Fig. 4D).

**DISCUSSION**

To identify potential molecular markers for GC and to better understand the development of GC at the molecular level, we performed SAGE analysis on 5 GC samples from 4 patients that showed distinct histological types and tumor stages. We analyzed with respect to (a) commonly up-regulated genes in GC compared with normal stomach, (b) up-regulated genes in advanced compared with early GC (genes potentially involved in tumor progression), (c) up-regulated genes in GC lymph node metastasis compared with primary GC (genes potentially involved in tumor metastasis), and (d) genes specifically ex-
pressed in GC. Quantitative RT-PCR analysis of 27 selected genes showed that COL1A1, CDH17, APOC1, COL1A2, YF13H12, CEACAM6, APOE, REGIV, FUS, and S100A11 were overexpressed in 40–80% of the 46 GC samples analyzed. Among them, TFF3, REGIV, and S100 calcium-binding proteins have been reported to be commonly up-regulated in GC by other SAGE studies (10, 12).

Among the 27 selected genes, only COL1A1 and CDH17 were overexpressed in >70% of the 46 GC samples, COL1A1 was most frequently overexpressed, and COL1A2 was also frequently overexpressed as determined by quantitative RT-PCR. Although COL1A1 expression has been demonstrated in tumor cells and tumor-associated stromal cells in multiple cancers (23, 24), COL1A1 and COL1A2 have been reported to be elevated in tumor endothelium as compared with normal endothelium (25), suggesting that they play an important role in angiogenesis and the formation of desmoplasia in GC. In fact, we found a significant association between tumor stage and mRNA expression level for both genes. CDH17 is a structurally unique member of the cadherin superfamily and is expressed in intestinal epithelial cells (26) and in intestinal metaplasia of the stomach (27). Although overexpression of CDH17 has been reported in intestinal type GC (27), the association between CDH17 and tumor invasion has not been examined in the present study. We showed that the high level of CDH17 expression was associated with advanced T grade, indicating that CDH17 is a candidate marker gene for tumor progression. However, a recent study of pancreatic cancer reported that high CDH17 expression correlates with good survival (28). Thus, the significance of the association of high CDH17 expression and advanced tumor invasion remains unclear. Organ specificity of CDH17 expression may be involved in tumor invasion and progression.

Frequently overexpressed genes in this study included 2 apolipoproteins, APOC1 was commonly up-regulated in GC, and APOE was a candidate marker for tumor metastasis. Although the expression status of these genes has not been previously examined in GC, it has been reported in certain cancers. APOC1 gene expression localizes to tumor-associated macrophages in breast carcinoma (24). In colorectal carcinoma, intense apolipoprotein E expression has been identified in macrophages surrounding the tumor area (29), suggesting that overexpression of these 2 apolipoproteins occur in tumor-associated macrophages. Macrophages appear to play a pivotal role in tumor angiogenesis, and in our previous observation, macrophage infiltration is significantly associated with tumor vessel density in GC (30). In addition, we found that a high level of APOE expression was associated with advanced T grade, N grade, and stage. Apolipoprotein E produced by tumor-associated macrophages may play an important role in tumor progression. Because APOE mRNA expression in lymph node metastasis tended to be higher than that in primary GC, APOE expression may be up-regulated in GC cells. In prostate cancer, apolipoprotein E expression was identified in cancer cells and correlated directly with Gleason grade (31). Whether GC cells or tumor-associated macrophages express apolipoprotein E remains unclear. Immunohistochemical analysis will be required to answer this question.

S100 calcium-binding proteins (S100A4, S100A9, S100A10, and S100A11) were among the 20 up-regulated genes. S100A4 is commonly up-regulated in GC. In fact, S100A4 expression was detected in 51 of 92 primary GC samples (55%; Ref. 32). Previous SAGE analysis of moderately differentiated GC indicated that 5 calcium-binding proteins (S100A2, S100A7, S100A8, S100A9, and S100A10) are overexpressed (10). S100A11 is potentially involved in tumor metastasis. However, no obvious up-regulation of S100A11 was identified in lymph node metastasis of GC. S100A11 may be important for stomach carcinogenesis, and overexpression of S100 calcium-binding proteins may be a common alteration in GC.

CEACAM6 is a member of the immunoglobulin superfamily (33) and functions as an intercellular adhesion protein (34). CEACAM6 overexpression independently predicts poor overall survival and disease-free survival in colorectal carcinoma (35). In GC, although frequent overexpression of CEACAM6 was identified in the present study, we found no association between the expression levels of CEACAM6 and clinicopathological features.

Overexpression of 2 genes related to wound-healing was identified in the present study. TFF3 functions in the maintenance and repair of the intestinal mucosa (36). TFF3 was commonly up-regulated in GC, and overexpression of TFF3 in GC has been reported previously (37). REGIV was a candidate gene specifically expressed in GC. REGIV is a member of the Reg gene family, which includes 3 other genes (22). REGIV expression is restricted to the gastrointestinal tract and pancreas and is up-regulated in response to mucosal injury in active Crohn's disease and ulcerative colitis (22). It has been reported that REGIV expression is increased in most colorectal cancers compared with normal tissues (38). Although overexpression of REGIV has been reported by conventional RT-PCR in 6 GC samples (12), the specificity of REGIV expression has not been investigated in our study. Virtual Northern and quantitative RT-PCR analysis showed REGIV expression to be narrowly restricted. We performed additional quantitative RT-PCR analysis of 10 colorectal cancers, 10 lung cancers, and 10 breast cancers (data not shown). Although REGIV expression was identified in all 10 colorectal cancers, the levels of REGIV expression were <100 arbitrary units. We also confirmed that the expression levels of REGIV in all 10 colorectal cancers were higher than those in normal colon. No REGIV expression was identified in lung or breast cancers. These results are consistent with the Virtual

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Case number</th>
<th>mRNA expression level</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH17</td>
<td>T gradea</td>
<td>1/2  20</td>
<td>0.065 ± 0.022</td>
</tr>
<tr>
<td></td>
<td>N gradea</td>
<td>3/4  26</td>
<td>0.275 ± 0.068</td>
</tr>
<tr>
<td></td>
<td>1/2/3</td>
<td>35</td>
<td>0.201 ± 0.053</td>
</tr>
<tr>
<td></td>
<td>Stageb</td>
<td>I/III 18</td>
<td>0.102 ± 0.036</td>
</tr>
<tr>
<td></td>
<td>III/IV</td>
<td>28</td>
<td>0.235 ± 0.064</td>
</tr>
<tr>
<td>FUS</td>
<td>T gradea</td>
<td>1/2  20</td>
<td>0.050 ± 0.017</td>
</tr>
<tr>
<td></td>
<td>N gradea</td>
<td>3/4  26</td>
<td>0.062 ± 0.014</td>
</tr>
<tr>
<td></td>
<td>1/2/3</td>
<td>35</td>
<td>0.066 ± 0.013</td>
</tr>
<tr>
<td></td>
<td>Stageb</td>
<td>I/III 18</td>
<td>0.033 ± 0.012</td>
</tr>
<tr>
<td></td>
<td>III/IV</td>
<td>28</td>
<td>0.072 ± 0.015</td>
</tr>
<tr>
<td>COL1A1</td>
<td>T gradea</td>
<td>1/2  20</td>
<td>6.84 ± 1.34</td>
</tr>
<tr>
<td></td>
<td>N gradea</td>
<td>3/4  26</td>
<td>20.66 ± 8.14</td>
</tr>
<tr>
<td></td>
<td>1/2/3</td>
<td>35</td>
<td>17.50 ± 6.11</td>
</tr>
<tr>
<td></td>
<td>Stageb</td>
<td>I/III 18</td>
<td>5.32 ± 1.14</td>
</tr>
<tr>
<td></td>
<td>III/IV</td>
<td>28</td>
<td>20.64 ± 7.54</td>
</tr>
<tr>
<td>COL1A2</td>
<td>T gradea</td>
<td>1/2  20</td>
<td>10.91 ± 1.86</td>
</tr>
<tr>
<td></td>
<td>N gradea</td>
<td>3/4  26</td>
<td>27.90 ± 9.14</td>
</tr>
<tr>
<td></td>
<td>1/2/3</td>
<td>35</td>
<td>23.87 ± 6.91</td>
</tr>
<tr>
<td></td>
<td>Stageb</td>
<td>I/III 18</td>
<td>9.45 ± 1.54</td>
</tr>
<tr>
<td></td>
<td>III/IV</td>
<td>28</td>
<td>27.62 ± 8.49</td>
</tr>
<tr>
<td>APOE</td>
<td>T gradea</td>
<td>1/2  20</td>
<td>6.17 ± 2.90</td>
</tr>
<tr>
<td></td>
<td>N gradea</td>
<td>3/4  26</td>
<td>11.29 ± 5.29</td>
</tr>
<tr>
<td></td>
<td>1/2/3</td>
<td>35</td>
<td>11.49 ± 2.66</td>
</tr>
<tr>
<td></td>
<td>Stageb</td>
<td>I/III 18</td>
<td>6.33 ± 3.22</td>
</tr>
<tr>
<td></td>
<td>III/IV</td>
<td>28</td>
<td>10.83 ± 2.78</td>
</tr>
</tbody>
</table>

* Mann-Whitney U test.
* Tumor staging of gastric carcinoma was done according to the Tumor-Node-Metastasis Stage Grouping (15).
Fig. 3. Validation of serial analysis of gene expression (SAGE) data by quantitative reverse transcription-PCR (RT-PCR). Fold change indicates the ratio of target gene mRNA level in gastric carcinoma (GC) to that in corresponding nonneoplastic mucosa. A, quantitative RT-PCR analysis of genes commonly up-regulated according to SAGE analysis. Of the 46 common up-regulated genes, 19 genes (41%) were commonly up-regulated in both GCs and nonneoplastic mucosa. The remaining 27 genes were up-regulated only in GCs.

A

B

C

D

Fig. 3 continued. Validation of serial analysis of gene expression (SAGE) data by quantitative reverse transcription-PCR (RT-PCR). Fold change indicates the ratio of target gene mRNA level in gastric carcinoma (GC) to that in corresponding nonneoplastic mucosa. A, quantitative RT-PCR analysis of genes commonly up-regulated according to SAGE analysis. Of the 46 common up-regulated genes, 19 genes (41%) were commonly up-regulated in both GCs and nonneoplastic mucosa. The remaining 27 genes were up-regulated only in GCs.
Northern analysis. Furthermore, we showed that the RegIV protein is secreted, suggesting that RegIV may serve as a serum tumor marker. The number of samples we studied was small, and serum RegIV protein levels have not been examined. Additional investigation will clarify whether the RegIV protein can serve as a serum tumor marker. The role of REGIV gene overexpression in stomach carcinogenesis

Fig. 4. A, Virtual Northern analysis shows REGIV expression to be narrowly restricted. Poorly differentiated gastric cancer (GC G189) showed strong expression of REGIV, whereas moderately differentiated GC (GC G234) did not. In our SAGE analysis, REGIV expression was detected as follows: 1933 in S219T; 714 in W246T; 1105 in P208L; 113 in W226T; and 259 in P208T. B, quantitative RT-PCR analysis of REGIV in primary GC and corresponding nonneoplastic mucosa. Fold change indicates the ratio of REGIV mRNA level in GC to that in corresponding nonneoplastic mucosa. Of the 46 GC samples, overexpression (T/n/2) of REGIV was identified in 22 (47.8%). C, quantitative RT-PCR analysis of REGIV in primary GC samples and various noncancerous tissues. In GC, a high level of REGIV expression (>100 arbitrary units) was identified in samples 31, 9, 33, 14, 30, 18, 27, 45, 21, 7, 15, 22, 16, and 46. In various noncancerous tissues, a high level of REGIV expression was not identified. REGIV expression was found in noncancerous stomach, duodenum, ileum, colon, and pancreas. The units are arbitrary, and we calculated REGIV mRNA expression by standardization to 1.0 μg of total RNA from HSC-39 as 1.0. D, anti-V5 Western blot assay of V5 epitope-tagged RegIV protein. Cells and media from MKN-28 cells transfected with pcDNARegIV-V5 (RegIV) or pcDNA 3.1 (empty) constructs were lysed, resolved by SDS-PAGE, and immunoblotted with monoclonal mouse anti-V5 antibody. We confirmed by anti-β-actin Western blot that contamination of cells in culture medium was minimal.

GC samples, overexpression (T/n/2) was detected at the following frequencies: 31 (67.4%) for APOCI and 24 (52.2%) for YF13H12. B, quantitative RT-PCR analysis of genes potentially involved in tumor progression according to SAGE analysis. Of the 46 GC samples, overexpression (T/n/2) was detected at the following frequencies: 34 (73.9%) for CDH17 and 19 (41.3%) for FUS. C and D, quantitative RT-PCR analysis of genes potentially involved in tumor metastasis according to SAGE analysis. C, mRNA expression levels of indicated genes in nonneoplastic mucosa, tumor, and lymph node metastasis. The units are arbitrary, and we calculated the target mRNA expression level by standardization to 1.0 μg of total RNA from HSC-39 as 1.0. T, tumor; n, = nonneoplastic mucosa; L, lymph node metastasis; NL, normal lymph node from autopsy. D, mRNA expression levels of indicated genes in 46 GC samples. Of the 46 GC samples, overexpression (T/n/2) was detected at the following frequencies: 23 (50%) for APOE and 19 (41.3%) for S100A11.
GENE EXPRESSION PROFILE OF GASTRIC CARCINOMA

Table 5 The 10 most up-regulated tags in lymph node metastasis of gastric carcinoma in comparison with primary gastric carcinoma

<table>
<thead>
<tr>
<th>Tag sequence</th>
<th>P208T</th>
<th>P208L</th>
<th>UniGene ID</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCGGGCCCCG</td>
<td>0(0)b</td>
<td>1105 (17)</td>
<td>Hs.274411</td>
<td>SCAND1</td>
<td>SCAN domain containing 1</td>
</tr>
<tr>
<td>TATAGGGTTA</td>
<td>0(0)</td>
<td>975 (15)</td>
<td>Hs.24950</td>
<td>RGS5</td>
<td>Regulator of G-protein signalling 5</td>
</tr>
<tr>
<td>CAGGGCCCAAC</td>
<td>0(0)</td>
<td>780 (12)</td>
<td>Hs.417004</td>
<td>S100A11</td>
<td>S100 calcium binding protein A11 (calgizzarin)</td>
</tr>
<tr>
<td>CAAGGCCACCCAG</td>
<td>0(0)</td>
<td>780 (12)</td>
<td>Hs.164906</td>
<td>RNPC2</td>
<td>RNA-binding region (RNP1, RRM) containing 2</td>
</tr>
<tr>
<td>GCCCAAGGCTCA</td>
<td>86 (1)</td>
<td>1560 (24)</td>
<td>Hs.10499</td>
<td>AP0E</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>TATTCCCCCCT</td>
<td>86 (1)</td>
<td>1430 (22)</td>
<td>Hs.78224</td>
<td>FLJ10815</td>
<td>Hypothetical protein FLJ10815</td>
</tr>
<tr>
<td>CAAGCAGGAC</td>
<td>0(0)</td>
<td>650 (10)</td>
<td>Hs.245451</td>
<td>RNASE1</td>
<td>Ribonuclease, RNAse A family, 1 (pancreatic)</td>
</tr>
<tr>
<td>TAGAGAGGCA</td>
<td>0(0)</td>
<td>650 (10)</td>
<td>Hs.457718</td>
<td>H3F3B</td>
<td>H3 histone, family 3B (H3.3B)</td>
</tr>
<tr>
<td>CTGGCGCTGCG</td>
<td>0(0)</td>
<td>585 (9)</td>
<td>Hs.22640</td>
<td>P24B</td>
<td>Integral type I protein</td>
</tr>
<tr>
<td>GCTGCTCCTCT</td>
<td>0(0)</td>
<td>585 (9)</td>
<td>Hs.343579</td>
<td>CLDN5</td>
<td>Claudin 3</td>
</tr>
<tr>
<td>CAAGCAGGAC</td>
<td>0(0)</td>
<td>650 (10)</td>
<td>Hs.245451</td>
<td>MLP1L4</td>
<td>Mitochondrial ribosomal protein L14</td>
</tr>
</tbody>
</table>

a The absolute tag counts are normalized to 1,000,000 total tags per sample.
b Number in parentheses indicates the absolute tag counts.

remains unclear. A possible involvement of REGIV in drug (5-fluorouracil or methotrexate) resistance was reported recently (38). Thus, REGIV may inhibit apoptosis and may participate in tumor cell growth.

We found that FUS and YF13H12 were overexpressed in GC. FUS was first identified as the 5’-part of a fusion gene with CHOP in myxoid liposarcomas with the translocation t(12;16)(q13;p11), and FUS protein was found to bind to RNA (39). No studies have analyzed FUS expression in human cancers, including GC. However, it has been shown that expression of the FUS domain restores liposarcoma development in CHOP-transgenic mice (40), suggesting that gain-of-function mutation of both FUS and CHOP is important. In the present study, FUS was a candidate marker for tumor progression, and we showed that a high level of FUS expression was associated with advanced N grade and stage. We also found YF13H12 gene overexpression in GC. However, YF13H12 function remains unclear, and there are no reports on YF13H12 gene expression. Additional studies will elucidate the biological role of FUS and YF13H12 protein in GC.

Although we found several genes to be overexpressed in GC by SAGE, there were some exceptions of genes overexpressed by SAGE but not by quantitative RT-PCR. It is possible that inconsistent results between SAGE and quantitative RT-PCR represent more than one gene. For example, TTTAATTTGT, represented in both GOLPH2 and G3BP, is commonly up-regulated in GC; however, the expression levels of both GOLPH2 and G3BP were not frequently up-regulated by quantitative RT-PCR. Whether discrepancies between SAGE and quantitative RT-PCR are attributable to differences in methodology remains to be determined. Some GC samples that we analyzed showed overexpression of both GOLPH2 and G3BP by quantitative RT-PCR. Recent evidence indicates that G3BP may serve as an important downstream effector of Ras signaling, and G3BP has been shown to be overexpressed in cancers of the colon, thyroid, breast, and head and neck (41). Thus, genes not frequently overexpressed may play an important role in restricted cases of GC.

Interestingly, among the 20 up-regulated tags in each GC sample, the 2 intestinal-type samples showed distinct tumor stages but showed many of the same tags. Cluster analysis showed that the two intestinal-type GC libraries were the most similar to each other. These results led us to speculate that morphologic phenotype reflects the gene expression profile. Our present results may be due to the selection of samples that represented similar histological features among many variations of intestinal type GC. Additional studies should investigate gene expression profile with respect to morphology. Comparison of expression patterns of W226T and W246T will provide a list of genes involved in tumor progression without the potential bias of histology. Our cluster analysis also showed that the gene expression pattern of SAGE gastric cancer-G234, which is a moderately differentiated tumor and is categorized as an intestinal type GC, was not similar to that of our 2 intestinal type GC samples but is similar to that of SAGE gastric cancer-G189, which is a poorly differentiated tumor and is categorized as a diffuse type GC. The gene expression patterns of GC in Japan may differ from those in the United States. Because we analyzed a limited number of GC samples, additional experiments are needed.

In conclusion, our present SAGE data provide a list of genes potentially involved in invasion, metastasis, and carcinogenesis of GC. We identified several genes by quantitative RT-PCR that have not previously been implicated in GC. Among these, a high level of REGIV expression was detected in GC, and expression of REGIV was narrowly restricted. Because the RegIV protein is secreted, it may serve as a biomarker for diagnosis of GC.

ACKNOWLEDGMENTS

We thank Dr. Kenneth W. Kinzler for SAGE software; Drs. Yukio Takeshima and Mayumi Kaneko for lung and breast cancer tissues; and Kenshi Tominaga and Yoshie Kaneko for excellent technical assistance and advice.

REFERENCES

GENE EXPRESSION PROFILE OF GASTRIC CARCINOMA

Gene Expression Profile of Gastric Carcinoma: Identification of Genes and Tags Potentially Involved in Invasion, Metastasis, and Carcinogenesis by Serial Analysis of Gene Expression

Naohide Oue, Yoichi Hamai, Yoshitsugu Mitani, et al.