Analysis of gene expression profile in colon cancer using the Cancer Genome Anatomy Project and RNA interference

Zhi Gang HUANG,† * Zhi Hua RAN,† Wei LU‡ & Shu Dong XIAO†

†Shanghai Institute of Digestive Diseases, Renji Hospital, Shanghai Second Medical University, *Department of Gastroenterology, Tongji Hospital, Tongji University, and ‡Department of Gastroenterology, Huashan Hospital, Medical College of Fudan University, Shanghai, China

OBJECTIVE: To investigate the changes in the gene expression profile in colon cancer to further identify gene markers that may be useful in the management of this disease.

METHODS: Data from serial analysis of gene expression (SAGE) collected by the Cancer Genome Anatomy Project (CGAP) were used to detect the difference in gene expression between normal tissue and colon cancer, and were further confirmed in a sample of 20 patients using RT-PCR. To identify the functions of differential genes in regulating the cell growth of colon cancer, RNA interference (RNAi) was used to block one of these genes in the colon cancer cell line HCT-116.

RESULTS: Expression changes of greater than two-fold in two SAGE libraries of colon cancer compared to two of normal tissue were observed for 216 tags of a total of 195 160 transcript tags (54 up-regulated genes and 136 down-regulated genes). Subsequent analysis of 17 genes by RT-PCR confirmed the reliability of this analysis. RNAi-mediated blockage of one of these genes, transforming growth factor (TGF)β1, significantly reduced the growth of a colon cancer cell line.

CONCLUSIONS: The combination of CGAP analysis and RNAi provides an excellent system to rapidly define the specific genes that are up-regulated in cancer to impact the growth of cancer cells. Further study on these differential overexpressed genes may provide gene markers for the detection and treatment of colon cancer.

KEY WORDS: colon cancer, gene expression profile, RNA interference, SAGE.

INTRODUCTION

Colon cancer is one of the most common cancers in China with an annually increasing incidence; the rise of colon cancer mortality is the fastest of all gastrointestinal cancers. Colon cancer typically develops over 7–12 years from normal cells to malignant cells and involves multiple genetic events. It is generally believed that one of initiating steps in colon carcinogenesis is mutation in the adenomatous polyposis coli (APC) tumor suppression gene. APC mutations are detected in about 70–80% of sporadic colon cancer. Another pathway activating colon cancer is associated with microsatellite instability (MSI), which accounts for about 15% of sporadic colon cancer. Despite this information of how these genetic alterations lead to the development and progression of colon cancer, the mechanism of carcinogenesis remains to be resolved.

The multistep process and long development period provide opportunities for prevention of colon cancer. To date, the methods of detecting colon cancer, including occult blood testing and colonoscopy, are insufficient.
for early detection of colon cancer. Better screening methods for early diagnosis are badly needed to improve accuracy, efficacy and compliance.

To identify relevant marker genes that are important in the pathogenesis of colon cancer, the present study aimed to use the data from serial analysis of gene expression (SAGE) collected by the Cancer Genome Anatomy Project (CGAP) to analyze the changes in gene expression profile between colon cancer and normal tissue, and further confirm the findings in a sample of 20 patients using reverse transcriptase-polymerase chain reaction (RT-PCR). In addition, this study aimed to determine whether RNA interference (RNAi)-mediated reduction in the level of genes identified by SAGE in a colon cancer cell line might result in decreased proliferation in the cell line. These studies were intended to rapidly identify new target genes, which when disrupted, significantly affect tumor growth. Furthermore, our results might suggest that the combination of SAGE and RNAi is a powerful and highly efficient system for identification of target genes that could be helpful for therapy and early diagnosis of colon cancer.

METHODS

Tissue samples

Colon cancer tissues and their corresponding normal mucosa were obtained from 20 patients who underwent surgical resection in Shanghai Renji Hospital. The tissues were snap-frozen in liquid nitrogen within 20–30 min of harvesting and stored at −70°C. Total RNA was extracted from the bulk tissue samples using the Trizol reagent (Invitrogen, Grand Island, NY, USA) and stored at −70°C.

Cell lines

The colon cancer cell lines SW-1116 and Lovo were obtained from Shanghai Institute of Digestive Disease, and HCT-8, Hce-8693 and HCT-116 were purchased from Shanghai Institute of Cell Biology. All cells were propagated in RPMI 1640 (Gibco, Grand Island, NY, USA) except HCT-116 cells, which were propagated in McCoy’s 5 A (Gibco), and supplemented with 10% fetal bovine serum (FBS), 2 µmol/L L-glutamine, 50 U/mL penicillin, and 50 µg/mL streptomycin. Total RNA was extracted with the same methods as from tissue samples.

Analysis of SAGE

Two colon adenocarcinoma (Tu-98 and Tu-102) and two normal tissue (NC-1 and NC-2) SAGE libraries were chosen from CGAP. Comparison was performed between the cancer and normal SAGE libraries using analysis tools in CGAP. Significant expression was reported when the odds ratio (OR) was significantly higher than twofold or lower than 0.5-fold. A total of 195 160 transcripts were analyzed.

Semi-quantitative RT-PCR

For validation of SAGE results by RT-PCR, total RNA of tissue samples and cell lines were further synthesized into single-stranded cDNA using M-MLV reverse transcriptase (Promega Co., Madison, USA). Semi-quantitative PCR was performed using a Thermocycler (Biometra, Goettingen, Germany). The primers for PCR were all obtained from Sangon (Shanghai, China). All genes examined by semi-quantitative RT-PCR were normalized to a control gene (β-actin) and then a ratio comparing differential expression in samples versus control was calculated.

RNA oligonucleotides

Small interfering RNA (siRNA) oligonucleotides with 3’ overhanging dTdT dinucleotides were designed for TGFβ1, TGFβ1-1 (Sense: 5’-CAUUAGAGCUUGCUCAAUG CCUt-3’; Antisense: 3’-tGIAUUCGACGCAUACCCG-5’); and TGFβ1-2 (Sense: 5’-CCACAUUICU GAAUGUACGCUt-3’; Antisense: 3’-tGGUAGAAGAACUUCAGUGA-5’). The DNA sequence of scrambled TGFβ1 was 5’-AATTCGAA-CGTATGCCTCCGG-3’. All siRNA were synthesized using in vitro siRNA construction kit (Ambion Inc. Austin, USA). Positive control DNA templates supplied with the siRNA construction kit were used to generate siRNA specific to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Transfection of siRNA

Approximately 2 × 10⁴ HCT-116 cells were plated per well in six-well plates or 1 × 10⁵ cells per well in 24-well plates in McCoy’s 5A supplemented with 10% FBS. After 24 h, siRNA were transfected by mixing with siPORT Lipid (Ambion) to result in a final siRNA concentration of 100 nmol/L per well. The cells were harvested at 48 h after transfection, and then RT-PCR and immunofluorescent staining were performed.

Cell cycle analysis

At 48 h after transfection, HCT-116 cells were harvested and fixed with 95% ethanol. After rinsing, centrifugation and digestion by RNaseA, the cells were stained by Propidium iodide (10 µg/mL) and then analyzed on a flow cytometer (Epics, Coulter, Fullerton, USA).
Soft agar colony assay
At 24 h after RNA transfection, the cells were mixed with cell culture medium containing 5% agar to result in a final agar concentration of 0.3%. Then 0.8 mL of this cell suspension was immediately plated in 24-well plates coated with 0.8 mL/well of 0.5% agar in culture medium. The colonies were counted in triplicate wells at 10 days after plating and the colony formation rate was calculated.

Statistical analysis
SPSS 11.0 was used to analyze the data. The results of RT-PCR, cell cycle analysis and colony formation were assessed by \( t \)-test. Significance level was defined as \( \alpha = 0.05 \).

RESULTS
SAGE analysis
The SAGE libraries derived from two samples of colon cancer and two of normal colonic epithelium were analyzed. Comparison between the cancer and normal samples was performed. The results revealed that 216 transcript tags were differentially expressed at greater than twofold. There were 33 transcripts that were elevated in colon cancer by 10-fold and 16 transcripts by 20-fold compared with normal tissue. In contrast, there were 54 transcripts that were decreased in cancers by 10-fold and 25 transcripts by 20-fold. From both practical and biological perspectives, these changes were very interesting.

Validation of a subset of differentially expressed genes
To examine the reliability of the SAGE data and to identify molecules important in the pathogenesis of colon cancer, 17 genes (16 up-regulated and one down-regulated) were chosen from the above-mentioned 216 differentially expressed genes (Table 1). These genes were chosen based on the level of differential expression. Eight genes (DPEP1, CLIC1, DEFA6, BNF1, IGF2, H19, TGF\( \beta \)1, RUNX2) were elevated by \( \geq 20 \)-fold in the cancer, five genes (DEFA5, SPARC, HSPCA, NPM1, GNB2L1) were elevated by 10–20 fold and three genes (RAB13, PSMC4, PYGB) by 2–10 fold. RNA from the cell lines and the paired tumor and normal samples were reverse transcribed into cDNA and then RT-PCR was performed.

The results showed that nine gene expressions were detected in the SW-1116 cell line: TGF\( \beta \)1, HSPCA, H19, CLIC1, DPEP1, NPM1, GNB2L1, PYGB and PSMC4. Additional expressed genes including BNF1, RUNX2, RAB13 and IGF2 were also detected in HCT-8 cell line; all gene expressions in HCT-8 except IGF2 were also confirmed in the Hce-8693 cell line; and only six gene expressions were detected in the Lovo cell line (Table 2). In the tumor samples, the following gene expressions...
were positive: TGFβ1, HSPCA, H19, CLIC1, DPEP1, NPM1, GNB2L1, PYGB, PSMC4, BNF1, RUNX2, RAB13, IGF2, SPARC and GUCA2B.

Three genes (TGFβ1, PSMC4 and HSPCA) were further analyzed by semiquantitative RT-PCR. Values for each gene were normalized to values obtained for β-actin, then a ratio of tumor:normal was obtained. The results showed TGFβ1, PSMC4 and HSPCA were overexpressed in 12/20 (60%), 10/20 (50%), and 7/20 (35%) of tumor specimens, respectively, compared with normal tissues (P < 0.01) and this had high consistency with SAGE data.

RNAi targeted against TGFβ1 reduced its expression in HCT-116 colon cancer cells

Two pieces of siRNA against TGFβ1 were synthesized, siRNA directed against GAPDH was used as a control. siRNA specific to scrabbled TGFβ1 was also synthesized as a control. Each siRNA was then transfected into HCT-116 cells, their effects were examined by RT-PCR or immunofluorescent stain at 48 h after transfection. The results clearly demonstrated that siRNA against TGFβ1 reduced the level of its transcription, and two pieces of siRNA had a similar role in inhibition of gene expression (Fig. 1).

Role of TGFβ1 in regulating cell cycle progression in vitro

Previous studies have shown that TGFβ1 is important in cellular proliferation and cell growth and induces a significant decrease in the S-phase of cell cycle. In present study, siRNA against TGFβ1 resulted in a very clear increase by 59–82.5% in the number of cells entering S-phase and a decrease by 23.1–35.9% in the number of cells entering G2-phase (Table 3).

Changes of expression of TGFβ1 alter the growth of HCT-116 in soft agar

At 24 h after transfection, the cells were placed into medium with soft agar, and colony formation rate was assayed after 10 days. The results showed that TGFβ1 siRNA decreased the colonies by nearly half (44.6–52% vs 85.3%) when the cell number per well was 25; a significant decrease of 60–61% versus 77.3% was found when the cell number per well was 50 (Table 4).

<table>
<thead>
<tr>
<th>Differential genes</th>
<th>SW-1116</th>
<th>Lovo</th>
<th>HCT-8</th>
<th>Hce-8693</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNF1</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IGF2</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>H19</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TGFβ1</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>DEFA6</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>DPEP1</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>DEFA5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SPARC</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CLIC1</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RUNX2</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>HSPCA</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>NPM1</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>GNB2L1</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RAB13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>PSMC4</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>PYGB</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>GUCA2A</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 2. Detection of 17 differential expressed genes in four cell lines of colon cancer (SW-1116, Lovo, HCT-8 and Hce-8693)

Table 3. Fractional changes in cell cycle phase after transfection of siRNA against transforming growth factor (TGFβ1) compared with control group

<table>
<thead>
<tr>
<th>Group</th>
<th>G1</th>
<th>Rate of cell cycle phase (%)</th>
<th>G2</th>
<th>S</th>
<th>G2/G1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGFβ1-1</td>
<td>36.58 ± 5.41*</td>
<td>25.55 ± 6.28*</td>
<td>37.87 ± 7.76*</td>
<td>1.75 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>TGFβ1-2</td>
<td>53.45 ± 6.67</td>
<td>30.35 ± 7.09**</td>
<td>16.20 ± 3.55*</td>
<td>1.73 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>54.56 ± 4.30</td>
<td>38.79 ± 4.25</td>
<td>6.64 ± 2.14</td>
<td>1.73 ± 0.03</td>
<td></td>
</tr>
</tbody>
</table>

Comparison with control group: *P < 0.001; **P < 0.05.
Table 4. Colony formation after transfection of siRNA against transforming growth factor (TGF)β1 compared with control group

<table>
<thead>
<tr>
<th>Cells per well</th>
<th>Rate of colony formation (%)</th>
<th>TGFβ1-1</th>
<th>TGFβ1-2</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td></td>
<td>44.6 ± 15.88**</td>
<td>52 ± 13.86*</td>
<td>85.33 ± 14.24</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>60 ± 7.15*</td>
<td>61 ± 8.65*</td>
<td>77.33 ± 8.26</td>
</tr>
</tbody>
</table>

Comparison with control group: *P < 0.01; **P < 0.001.

DISCUSSION

In this study, SAGE and informatics tools from CGAP were used to analyze gene expression in colon cancer and normal tissue. The CGAP was launched in 1997 by the National Cancer Institute of America, which was a platform bringing together expressed sequence tag (EST) and SAGE database and bioinformatics analysis tools. Data from SAGE collected by CGAP has been used successfully in locating new oncogenes, establishing overexpressed genes in tumors and helping construct specific microarray for study of tumors. Hough et al.5 analyzed the gene expression profile of ovarian cancer using SAGE data from CGAP in 2000, and the results demonstrated that many of the genes were up-regulated in ovarian cancer represented surface or secreted proteins such as claudin-3 and -4, HE4, mucin-1, epithelial cellular adhesion molecule, and mesothelin, and were further confirmed using immunohistochemistry.

Serial analysis of gene expression is a profiling method that associates individual mRNA transcripts with 10–15 base tags derived from specific positions near their 3’ termini. The abundance of each tag provides a quantitative measurement of the transcript level presented within the mRNA population studied. SAGE is not dependent on a preexisting database of expressed genes and therefore provides a more accurate view of the gene expression profile differentiated from DNA microarray. For the present study, SAGE libraries derived from two samples of colon cancer and two normal colonic tissues were analyzed. These libraries contained a combined total of 195 160 transcript tags, and then a comparison was performed and revealed 216 transcript tags that were differentially expressed greater than two-fold between normal and tumor samples.

Sixteen up-regulated genes and one down-regulated gene were selected to be further confirmed by RT-PCR in colon cancer cell lines (SW-1116, Lovo, HCT-8 and Hce-8693) and 20 specimens of colon cancer. Positive rates ranging from 35.3% to 76.5% in the cell lines and 88.2% in the specimens were found. The results confirmed by RT-PCR are consistently with SAGE data. Of these genes, BNF1 was first reported in 2003 to be over-expressed in breast cancer and to encode an extracellular matrix protein, and was also positive in lung and colon cancer.6 NMP1 is responsible for protein shuttling between the cytoplasm and the nucleus and its overexpressed condition is associated with cell proliferation and apoptosis, and NMP/ALK fusion protein may play an important role in NMP/ALK-mediated lymphomagenesis.7 The expressed status and role of NMP1 in colon cancer need to be further identified.

The differential transcriptions of TGFβ1, PSMC4 and HSPCA were further confirmed as overexpressed genes in 12 (60%), 10 (50%) and 7 (35%) of 20 resected specimens, respectively (P < 0.01), by semiquantitative RT-PCR. In our study, overexpression of TGFβ1 was not significantly associated with Duke’s stages and pathological degrees, but this could not be excluded due to the unfavorable effects of low sample size and disproportion of distribution. It has been reported that mutation of TGFβ receptor II (RII) induced TGFβ resistance was detected in 15% of colon cancers;8 however, the level of TGFβ1 increased in 60–75% of colon cancer cases in our report, in accordance with other studies, and this is hard to be explained by mutation of RII. It is essential to further identify the role of TGFβ1 in the pathogenesis of colon cancer.

PSMC4 is a subunit of 26S proteasome with characteristics of ATPase activity and molecular chaperone. Proteasomes are strongly expressed in cutaneous malignant melanoma and high levels of circulating proteasomes have been reported in patients with melanoma. Recently ubiquitin ligase subunits Skp2 and Cks 1 of proteasome complex were reported to be associated with poor prognosis in colorectal cancer.9 It was also found overexpressed transcripts of PSMC4 in colon cancer in present study; it needs to be further identified whether overexpressed PSMC4 would become a marker.

HSPCA is a molecular chaperone associated with conformation stability and function of a wide range of oncogenic proteins, including c-Raf-1, Cdk4, ErbB2, mutant p53, c-Met, Polo-1 and telomerase hTERT.10 It was overexpressed in a wide variety of malignant tumors such as breast cancer, endometrial cancer, and pancreatic cancer. HSPCA was overexpressed in colorectal cancer in our study, which suggests that HSPCA may have a role in the cell proliferation of colon cancer.

To identify relevant target genes that are important in the pathogenesis of colon cancer, it was additionally investigated whether reducing the expression levels of
these genes in a colon cancer line might alter its proliferation in present study. siRNA against TGFβ1 were successfully transfected into the cells of colon cancer cell line HCT-116 and inhibition of TGFβ1 expression was detected. siRNA directed against TGFβ1 resulted in an obviously increased number of cells entering S phase and a reduced number of cells in G2 phase. Colony assay further demonstrated that siRNA against TGFβ1 led to a significant reduction in colony formation as compared with the control group. It is hypothesized that inhibition of cell proliferation by siRNA against TGFβ1 might be associated with induction of apoptosis in S phase.

With relative ease, the present results using siRNA were able to confirm the relationship between the overexpressed genes and disease progression, which needed more labor-intensive efforts in the past. In addition, siRNA against TGFβ1 may prove to be a novel therapeutical method; the development of stable plasmids expressing siRNA oligonucleotides may be the next step in this direction.

In conclusion, it has demonstrated that the SAGE database of CGAP combined with use of RNAi in the study of gene expression of colon cancer provides an excellent system to define the role of specific genes that are upregulated in colon cancer. Further study on overexpressed genes including TGFβ1, PSMC4 and HSPCA is suggested, as these may have a role in the elucidation of the pathogenesis of colon cancer, and may be potential gene markers for diagnosis and treatment of colon cancer.

REFERENCES