Screening of potential diagnostic markers and therapeutic targets against colorectal cancer

XiaoQing Tian
DanFeng Sun
ShuLiang Zhao
Hua Xiong
JingYuan Fang

Department of Gastroenterology and Hepatology, Renji Hospital, School of Medicine, Shanghai Institute of Digestive Disease, Shanghai Jiao Tong University, Shanghai, People’s Republic of China

Objective: To identify genes with aberrant promoter methylation for developing novel diagnostic markers and therapeutic targets against primary colorectal cancer (CRC).

Methods: Two paired CRC and adjacent normal tissues were collected from two CRC patients. A Resi: MBD2b protein-sepharose-4B column was used to enrich the methylated DNA fragments. Difference in the average methylation level of each DNA methylation region between the tumor and control samples was determined by log2 fold change (FC) in each patient to screen the differentially methylated DNA regions. Genes with log2 FC value ≥4 or ≤−4 were identified to be hypermethylated and hypomethylated, respectively. Then, the underlying functions of methylated genes were speculated by Gene Ontology database and pathway enrichment analyses. Furthermore, a protein–protein interaction network was built using Search Tool for the Retrieval of Interacting Genes/Proteins database, and the transcription factor binding sites were screened via the Encyclopedia of DNA Elements (ENCODE) database.

Results: Totally, 2,284 and 1,142 genes were predicted to have aberrant promoter hypermethylation or hypomethylation, respectively. MAP3K5, MAP3K8, MAPK14, and MAPK9 with promoter hypermethylation functioned via MAPK signaling pathway, focal adhesion, or Wnt signaling pathway, whereas MAP2K1, MAPK3, MAPK11, and MAPK7 with promoter hypomethylation functioned via TGF-beta signaling pathway, neurotrophin signaling pathway, and chemokine signaling pathway. CREBBP, PIK3R1, MAPK14, APP, ESR1, MAPK3, and HRAS were the seven hubs in the constructed protein–protein interaction network. RPL22, RPL36, RPLP2, RPS7, and RPS9 were commonly regulated by transcription factors, and YY1 and IRF4 were hypermethylated.

Conclusion: MAPK14, MAPK3, HRAS, YY1, and IRF4 may be considered as potential biomarkers for early diagnosis and therapy of CRC.

Keywords: primary colorectal cancer, aberrant DNA methylation, microarray analysis, pathway enrichment analysis, transcription factor

Introduction
Colorectal cancer (CRC) is one of the most prevalent human malignancies. The 5-year survival rate can reach up to 90% for early stage CRC patients, but will decrease to <10% in patients with distant metastases. Thus, there is an urgent need to discover novel early diagnostic markers against this refractory disease.

Epigenetic modifications refer to any heritable changes in gene expression without changes in the DNA sequence, which are now considered primarily attributable to two molecular events: CpG DNA methylation and histone modification. Currently, aberrant DNA methylation is mainly seen in two forms: hypermethylation of CpG islands in the promoter regions of genes and global DNA hypomethylation (generally in pericentric heterochromatin). The former is defined as an increased level of DNA
methylation at a CpG island in a promoter area, which can lead to gene silencing and is thought to cooperate with other genetic mechanisms to alter key signaling pathways critical to colorectal tumorigenesis. Hypomethylation can result in genetic instability in somatic cells, thus increasing the mutation rates in tumor cells and chromosomal abnormalities. Similar to hypomethylation occurring in heterochromatin, hypomethylation of the promoters of some oncogenes may cause their overexpression, driving tumorigenesis. Extensive research work has also been carried out to investigate the role of aberrant DNA methylation in CRC, and some inspiring advances have been made. For instance, using Illumina Infinium HM27 DNA methylation assay, Hinoue et al have identified four DNA methylation-based CRC subgroups, each showing characteristic genetic and clinical features, which is consistent with the fact that CRC is a heterogeneous disease with distinct genetic and epigenetic alterations in different subtypes. At present, SEPT9 has been proven to be a reliable blood-based biomarker for stages II–IV CRC. However, practically available methylation markers are still very limited, especially those available for detection of early stage CRC. Thus, novel and effective methylation-based markers are imperatively needed to be developed.

In the present study, we performed a genome-wide search for genes with aberrant methylation in CRC by a methylated DNA immunoprecipitation-chip analysis, with the aim to acquire novel and effective diagnostic markers and therapeutic targets against primary CRC.

Materials and methods

Tissue samples

Two CRC tissue samples and two adjacent normal control samples were collected from two CRC patients (represented as LDM and YZK groups, respectively) during operation at Renji Hospital (Shanghai, People’s Republic of China) from January 2006 to June 2007. The two patients, a 57-year-old male and a 59-year-old female, were Han Chinese; both were diagnosed with primary CRC without lymphatic or distant metastasis, and the tumor cells were histopathologically diagnosed as poorly differentiated adenocarcinoma; the patients neither received preoperative treatments such as radiotherapy or chemotherapy nor had any chronic disease history before the operation. This study was approved by the ethics committee of Renji Hospital of Shanghai Jiao Tong University and was carried out under the provisions of the 1975 Declaration of Helsinki. For each patient, adjacent tumor-free parenchyma, 5 cm away from the tumor, was obtained to serve as the paired control. Immediately after the surgical removal, the tissue samples were frozen in liquid nitrogen and stored at –80°C.

Enrichment of methylated DNA fragments and promoter microarray hybridization

Genomic DNA was prepared from two pairs of CRC tissues. A Resi: MBD2b protein-sepharose-4B column (provided by Dr Han, Shanghai Hujing Biotech Co, Ltd.) was used to enrich the methylated DNA fragments as described by Tóth et al. The enriched methylated DNA fragments and the input control were sent to NimbleGen for labeling and hybridization using NimbleGen promoter tiling array (Homo sapiens HG18 RefSeq Promoter), which can detect 28,226 CpG sites. Microarray hybridization, washing, and scanning were performed using the NimbleGen Systems of Iceland as described previously.

Preprocessing of DNA methylation microarray data

Data were analyzed using NimbleScan™ (NimbleGen, Inc., Madison, WI, USA) and custom-written software. NimbleScan detects peaks by searching for four or more probes with a signal above the specified cutoff value using a 500 bp sliding window. The cutoff value is the percentage of a hypothetical maximum, which is the mean +6 (standard deviation). The log$_2$-transformed signal intensity ratios of Cy5 to Cy3 were then randomized 20 times to evaluate the probability of false positives. The cutoff value used here ranged from 90% to 15%. Each peak is assigned a false discovery rate (FDR) score based on randomization. In general, the following guidelines are used when reviewing FDR scores: 1) the lower the FDR value, the more likely the peak corresponds to a protein-binding site; 2) for most datasets, peaks with an FDR score ≤0.05 very often represent the high-confidence protein-binding site(s); 3) peaks with an FDR score between 0.05 and 0.2 are also indicative of a binding site; and 4) peaks with an FDR score >0.2 are generally not considered high-confidence binding sites.

Screening of differentially methylated DNA regions

The number of DNA methylation regions was determined in both the LDM and YZK groups. If DNA methylation was not detected, the value 0 was assigned to the missing data. Then, difference in the average methylation level of each DNA methylation region between the tumor and control samples, which is measured by log$_2$ fold change (FC) value, was determined in the LDM and YZK groups, respectively, in order to
screen the differentially methylated DNA regions (DMRs). Genes with log₂FC value ≥4 and ≤-4 were identified to be hypermethylated and hypomethylated, respectively.

**Functional annotation and pathway enrichment analysis**

Functional annotation and pathway enrichment analysis of genes with aberrant promoter methylation were performed based on Gene Ontology (GO) database (P<0.05) and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (P<0.05), respectively. Only GO terms and KEGG pathways annotating two or more genes were retained.

**Screening of transcription factors, proto-oncogene, and tumor suppressor genes**

With reference to the TRANSFAC (TRANScription FACtor) database, we further checked whether genes with DMR are transcription factors (TFs). Meanwhile, we also screened the known proto-oncogene and tumor suppressor genes with reference to Tumor Suppressor Genes (TSGenes) and Tumor-Associated Gene (TAG) databases.

**Construction of protein–protein interaction network**

To better understand the differentially expressed gene (DEGs) from an interactive perspective, we used STRING (Search Tool for the Retrieval of Interacting Genes/Proteins; http://www.string-db.org) database to build an interaction network of encoding products of genes with DMR. The resulting network was visualized using Cytoscape software. The connection degree of each gene in the network was calculated.

**Screening of TF binding sites (TSSs) and TFs**

First, data regarding the TF binding sites (transcription start sites [TSSs]) in the ENCODE database were downloaded. To improve the reliability, we tested the reproducibility of TSSs by screening those appearing in at least two independent samples. Next, for TFs within 1 kb upstream or 0.5 kb downstream, a TSS was screened for each gene. TFs enriched in at least ten genes with aberrant promoter methylation were obtained by Fisher’s exact test at FDR <0.01. FDR control is used in multiple hypothesis testing to correct for multiple comparisons.

**Results**

**Genes with aberrant methylation in the promoter region**

According to the differential methylation analysis of LDM and YZK groups, 2,284 and 1,142 genes were predicted to have aberrant promoter hypermethylation and hypomethylation, respectively (Table 1), with a ratio of approximately 2:1, indicating that gene hypermethylation has a dominant role in CRC compared to hypomethylation. These genes with aberrant methylation included some mitogen-activated protein kinase kinase kinase (MAPKKK)-encoding genes as well as mitogen-activated protein kinase (MAPK)-encoding genes, such as MAP3K5 (mitogen-activated protein kinase kinase kinase 5), MAP3K8 (mitogen-activated protein kinase kinase kinase 8), MAPK14 (mitogen-activated protein kinase 14), and MAPK9 (mitogen-activated protein kinase 9), which were observed to display aberrant hypermethylation, and MAP2K1 (mitogen-activated protein kinase kinase 1), MAPK11 (mitogen-activated protein kinase 11), MAPK7 (mitogen-activated protein kinase 7), and MAPK3 (mitogen-activated protein kinase 3), which were observed to have aberrant promoter hypomethylation.

**Pathway enrichment analysis of genes with aberrant methylation**

According to the pathway enrichment analysis, genes with hypermethylation were observed to be mainly enriched in the following pathways: calcium signaling pathway, neuroactive ligand–receptor interaction, and regulation of actin cytoskeleton (Table 2).

### Table 1 Results of screened genes with aberrant promoter hypermethylation or hypomethylation

<table>
<thead>
<tr>
<th></th>
<th>Peak count</th>
<th>Gene count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypermethylated in cancer</td>
<td>6,396</td>
<td>2,283</td>
</tr>
<tr>
<td>Hypomethylated in cancer</td>
<td>4,288</td>
<td>1,142</td>
</tr>
<tr>
<td>Total</td>
<td>10,684</td>
<td>3,425</td>
</tr>
</tbody>
</table>

**Table 2 Pathway analysis of the genes with aberrant promoter hypermethylation**

<table>
<thead>
<tr>
<th>KEGG pathway name</th>
<th>Gene count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium signaling pathway</td>
<td>44</td>
<td>6.11E-09</td>
</tr>
<tr>
<td>Neuroactive ligand–receptor interaction</td>
<td>47</td>
<td>0.000113936</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
<td>38</td>
<td>0.00026657</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>42</td>
<td>0.002032766</td>
</tr>
<tr>
<td>CAMs</td>
<td>22</td>
<td>0.012362503</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>30</td>
<td>0.015426148</td>
</tr>
<tr>
<td>Vitamin B6 metabolism</td>
<td>3</td>
<td>0.015943305</td>
</tr>
<tr>
<td>Wnt signaling pathway</td>
<td>23</td>
<td>0.024937712</td>
</tr>
<tr>
<td>Axon guidance</td>
<td>20</td>
<td>0.031486352</td>
</tr>
<tr>
<td>Glycosphingolipid biosynthesis – globo series</td>
<td>4</td>
<td>0.044406999</td>
</tr>
</tbody>
</table>

**Abbreviations:** KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; CAMs, cell adhesion molecules.
Table 3 Pathway analysis of the genes with aberrant promoter hypomethylation

<table>
<thead>
<tr>
<th>KEGG pathway name</th>
<th>Gene count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal adhesion</td>
<td>26</td>
<td>8.44E-06</td>
</tr>
<tr>
<td>ECM–receptor interaction</td>
<td>13</td>
<td>0.000331572</td>
</tr>
<tr>
<td>Protein digestion and absorption</td>
<td>10</td>
<td>0.00780902</td>
</tr>
<tr>
<td>TGF-beta signaling pathway</td>
<td>10</td>
<td>0.010034586</td>
</tr>
<tr>
<td>Neurotrophin signaling pathway</td>
<td>13</td>
<td>0.012476151</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>18</td>
<td>0.013928422</td>
</tr>
<tr>
<td>Glycosaminoglycan biosynthesis – chondroitin sulfate</td>
<td>4</td>
<td>0.023578401</td>
</tr>
<tr>
<td>Chemokine signaling pathway</td>
<td>16</td>
<td>0.03156309</td>
</tr>
<tr>
<td>Ubiquinone and other terpenoid-quione biosynthesis</td>
<td>2</td>
<td>0.046096841</td>
</tr>
</tbody>
</table>

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; ECM, extracellular matrix; TGF, transforming growth factor.

digestion and absorption, TGF-beta signaling pathway, and neurotrophin signaling pathway (Table 3).

Screening of TFs and TAG from genes with aberrant methylation

With reference to the TRANSFAC database, TFs with hypermethylation and hypomethylation were 126 and 39 accounting for 5.5% and 3.4% of the total genes with aberrant methylation, respectively (Table 4). Meanwhile, with reference to the TSGenes and TAG databases, 32 and 10 oncogenes (CDC25B, EWSR1, FGFR3, HRAS, RALA, REL, RET, RYK, SET, and TRIO) were observed to have promoter hypermethylation and hypomethylation, respectively. In addition, 86 and 49 tumor suppressor genes were observed to have promoter hypermethylation and hypomethylation, respectively (Table 4).

Construction of protein–protein interaction network using genes with aberrant methylation

The resulting protein–protein interaction (PPI) network is shown in Figure 1, and the top ten hub proteins were CREBBP (degree =35), PIK3R1 (degree =34), MAPK14 (degree =32), APP (degree =31), ESR1 (degree =30), MAPK3 (degree =28), HRAS (degree =27), ITGB1 (degree =26), RNPS1 (degree =25), and STAT3 (degree =25), successively.

Screening of TSSs and TFs

Twenty-one TFs were enriched in genes with promoter hypermethylation or hypomethylation (Figure 2), and most TFs were more significantly enriched in genes with aberrant promoter hypermethylation. Notably, TAF1 and HNF4A were exclusively enriched in genes with aberrant promoter hypermethylation.

Among the 21 predicted TFs, CTCF, TAF1, YY1, interferon regulatory factor 4 (IRF4), and ELF1 were the top five TFs that were predicted to be impaired in regulating most genes with aberrant promoter methylation (Figure 3). Among them, YY1 and IRF4 were observed to display aberrant promoter hypomethylation and hypermethylation, respectively, both of which commonly regulate genes encoding ribosomal proteins (RPs), such as RPL22, RPL36, RPLP2, RPS7, and RPS9, that were observed to have aberrant promoter hypermethylation.

Discussion

As genes with promoter hypermethylation were onefold more than those with promoter hypomethylation, it can be inferred that hypermethylation may have a more significant role in colorectal tumorigenesis. Among the genes with aberrant methylation, MAPKKK and MAPK genes seem to have particularly important roles in tumorigenesis, despite some opposite methylation aberrations were observed between them, indicating their complicated roles in colorectal tumorigenesis. MAP3K5, MAP3K8, MAPK14, and MAPK9 were observed to have aberrant promoter hypermethylation functioning via MAPK signaling pathway, focal adhesion, or Wnt signaling pathway, whereas MAP2K1, MAPK3, MAPK11, and MAPK7 were observed to have aberrant promoter hypomethylation functioning via TGF-beta signaling pathway, focal adhesion, neurotrophin signaling pathway, and chemokine signaling pathway. A previous study has implicated the alteration of MAPK signaling pathway and Wnt signaling pathway in CRC. MAPK pathways are evolutionarily conserved module that are involved in various cellular functions, such as growth, proliferation, differentiation, migration, and apoptosis, in which MAPK is activated upon phosphorylation by a mitogen-activated protein kinase kinase (MAPKK), which in turn is activated when phosphorylated by a MAPKKK. Schwartsmann et al have further pointed out that MAPK pathways are located at the downstream of growth-factor receptors in CRC. Slattery et al have proposed that genetic variation in the MAPK signaling pathway genes

Table 4 Screening of TFs, oncogenes, and tumor suppressor genes

<table>
<thead>
<tr>
<th>TF count</th>
<th>Oncogene count</th>
<th>Tumor suppressor count</th>
<th>Other TAG count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypermethylated in cancer</td>
<td>126</td>
<td>32</td>
<td>86</td>
</tr>
<tr>
<td>Hypomethylated in cancer</td>
<td>39</td>
<td>10</td>
<td>49</td>
</tr>
</tbody>
</table>

Abbreviations: TF, transcription factor; TAG, tumor-associated gene.
Figure 1
The constructed protein–protein interaction network.
Notes: A red node represents a gene with aberrant hypermethylation in the promoter region, and a green one represents a gene with aberrant hypomethylation in the promoter region. The size of a node is proportional to the connection degree.
(including MAP2K1, MAPK14, and MAPK3) influences CRC risk and survival after diagnosis.\textsuperscript{22} Mutations in Wnt signaling pathway gene members have been found in many colorectal carcinomas.\textsuperscript{23,24} In our study, aberrant methylation is also observed in the promoter region of genes involved in this pathway, such as MAPK9.

Additionally, MAPK14- and MAPK3-encoding products were also identified as hubs with high connection degrees in the constructed PPI network, implying their significant roles in colorectal tumorigenesis, despite their opposite methylation aberrations. The adverse effects of mutation of these two genes in CRC have been presented by Slattery et al\textsuperscript{22} but promoter methylation of these two genes in CRC has never been reported. Our finding suggests that the aberrant methylation may also have a role in CRC.

Among the genes observed with aberrant promoter hypomethylation, ten were known oncogenes, of which HRAS, FGFR3, and RET were observed to function via the endocytosis pathway, and also HRAS was observed to function via three additional pathways: focal adhesion, neutrophin signaling, and chemokine signaling.

HRAS encodes GTPase HRas, a small G protein in the Ras subfamily of the Ras superfamily of small GTPases. The hypomethylation of HRAS promoter in colon cancer cells has been discovered by Luo et al using methylation-specific polymerase chain reaction (PCR) assay, when they studied the effect of S-adenosylmethionine treatment on CRC cells.\textsuperscript{25} Feng et al have previously reported the significantly higher H-Ras protein level in the malignant tumor compared with the normal adjacent tissues,\textsuperscript{26} which may be partly attributed to hypomethylation. Additionally, they further reported close correlations between H-Ras expression and two MAPK pathway genes, MEK and ERK, which agree with the previous finding that mutations in MAPK pathway genes are frequently mostly affecting Ras and B-Raf in the extracellular signal-regulated kinase pathway.\textsuperscript{27} Meanwhile, HRAS-encoding product was also observed to have a higher connection degree according to the constructed PPI network, indicating that it may have a critical role in colorectal tumorigenesis.

The aberrant methylation occurring in gene promoter region will change the DNA conformation, impairing the combination of TFs with gene promoter, thus affecting their normal regulation of gene transcription. As 21 TFs were enriched in the genes with promoter hypermethylation or hypomethylation, it can be inferred that the normal

---

**Figure 2** The enriched TFs.

**Notes:** A red bar represents a TF enriched in genes with aberrant hypermethylation in the promoter region, and a green bar represents a TF enriched in genes with aberrant hypomethylation in the promoter region.

**Abbreviations:** TFs, transcription factors; FDR, false discovery rate.
transcriptional regulation by these 21 TFs of the resulting genes is presumably perturbed. In addition, our study also suggests that the transcriptional regulation by TFs of genes with aberrant promoter hypermethylation is more significantly affected compared to that of genes with aberrant promoter hypomethylation; additionally, there are two TFs exclusively enriched in genes with aberrant hypermethylation. Thus, all these findings consistently support our previous point that hypermethylation may have a more significant role in colorectal tumorigenesis.

Among the hypermethylated RP-encoding genes – RPL22, RPL36, RPLP2, RPS7, and RPS9 – the former three encode three protein components of the 60S subunit of a ribosome and the latter two encode two protein components of the 40S subunit. The downregulation of RPLP2^{28} and the overexpression of PRL36^{29} and RPS7^{30} have been reported previously. Lai and Xu have summarized the expression changes of different RPs in CRC reported by different researchers and found that the upregulation and downregulation of the same RP are not always consistent, and they further demonstrated...
that the extraribosomal functions of RPs may be critical for colorectal tumorigenesis.\textsuperscript{31} However, little is known about the aberrant methylation in these RP genes till now. In the present study, all these five RP-encoding genes were observed to be commonly regulated by \textit{YY1} and \textit{IRF4}. \textit{YY1} encodes a ubiquitously distributed TF Yin Yang 1 that belongs to the GLI-Kruppel class of zinc finger proteins, which is implicated in histone modification, and \textit{IRF4} encodes an IRF family of TFs, which is important for the regulation of interferon-inducible genes. \textit{YY1} overexpression has been demonstrated in human colon cancer by Chinnappan et al\textsuperscript{32} and the downregulated \textit{IRF4} expression has also been reported previously,\textsuperscript{33} which may have similar effects on the level of corresponding gene-encoding protein as the promoter hypermethylation and hypomethylation, respectively, despite the aberrant promoter methylation has never been reported in these two genes. Thus, the aberrant methylation occurring in both TFs, \textit{YY1} and \textit{IRF4}, and their downstream RP-encoding genes probably contributes to CRC occurrence.

Taken together, hypermethylation seems to have a more significant role in colorectal tumorigenesis. MAPKKK and MAPK genes (especially \textit{MAPK14} and \textit{MAPK3}) as well as the genes encoding PPI hub proteins (HRAS in particular) might be essential for CRC occurrence, which are supposed to function via one or more pathways in tumorigenesis. Meanwhile, some TFs, such as \textit{YY1} and \textit{IRF4}, together with their downstream target RP-encoding genes (\textit{RPL22}, \textit{RPL36}, \textit{RPLP2}, \textit{RPS7}, and \textit{RPS9}) are also supposed to have critical roles in colorectal tumorigenesis. Therefore, \textit{MAPK14}, \textit{MAPK3}, \textit{HRAS}, \textit{YY1}, and \textit{IRF4} may be considered as potential diagnostic markers and therapeutic targets against primary CRC. However, only two patients were involved in the present study. Samples collected from more patients are imperatively needed to validate the methylation status of genes identified earlier. Thus, the findings in the present study should be taken carefully.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


