Molecular Characterization of Colorectal Signet-Ring Cell Carcinoma Using Whole-Exome and RNA Sequencing

**Abstract**

**BACKGROUND:** Signet-ring cell carcinoma (SRCC) is a very rare subtype of colorectal adenocarcinoma (COAD) with a poor clinical prognosis. Although understanding key mechanisms of tumor progression in SRCCs is critical for precise treatment, a comprehensive view of genomic alterations is lacking. **MATERIALS AND METHODS:** We performed whole-exome sequencing of tumors and matched normal blood as well as RNA sequencing of tumors and matched normal colonic tissues from five patients with SRCC. **RESULTS:** We identified major somatic alterations and characterized transcriptional changes at the gene and pathway level. Based on high-throughput sequencing, the pattern of mutations and copy number variations was overall similar to that of COAD. Transcriptome analysis revealed that major transcription factors, such as SRF, HNF4A, ZEB1, and RUNX1, with potential regulatory roles in key pathways, including focal adhesion, the PI3K-Akt signaling pathway, and the MAPK signaling pathway, may play a role in the tumorigenesis of SRCC. Furthermore, significantly upregulated genes in SRCCs were enriched for epithelial-mesenchymal transition genes, and accumulation of mucin in intracytoplasm was associated with the overexpression of MUC2. **CONCLUSION:** The results indicate that the molecular basis of colorectal SRCC exhibits key differences from that of consensus COAD. Our findings clarify important genetic features of particular abnormalities in SRCCs.

**Introduction**

Colorectal cancer is the third most common malignancy and a leading cause of cancer-related deaths worldwide [1,2]. Colorectal cancer represents a group of histologically heterogeneous tumors, and its histological subtype is determined by the major component of cancer cells [3–5]. There are several histological subtypes in colorectal cancer; adenocarcinoma is the most common type, and other subtypes include mucinous carcinoma, signet-ring cell carcinoma (SRCC), squamous cell carcinoma, and undifferentiated carcinoma [2–4]. The identification of the histological subtype is important for cancer patient management because it plays an important role in determining tumor biology and aggressiveness. In colorectal cancer, histological subtype is considered an important prognostic factor.
SRCC is a rare subtype of colorectal cancer, accounting for approximately 1% of patients with colorectal cancer [5, 6]. Signet-ring cell is characterized by intracytoplasmic mucin which displaces the nucleus to one side. SRCC is defined as a carcinoma with a signet-ring cell component of greater than 50% [4, 6]. Previous studies have reported that colorectal SRCC is associated with young age, advanced tumor stage, high rates of metastasis, and poor prognosis [5, 7]. These aggressive behaviors are associated with specific molecular features such as microsatellite instability, a high frequency of CpG island methylator phenotypes, and frequent BRAF and KRAS mutations [3, 4]. Despite the poor prognosis, limited studies have characterized colorectal SRCC owing to the low incidence. Thus, clinicopathological features are not well understood, and few studies have compared SRCC with typical adenocarcinoma of colorectal cancer, especially at the molecular level. In addition, the genomic characteristics of SRCC have been only partially defined by low-throughput molecular studies so that comprehensive genomic profiling is needed for better understanding molecular mechanisms of SRCC.

In this study, we conducted a comprehensive analysis of five SRCCs by both whole-exome and RNA sequencing with tumors and matched normal. Whole-exome sequencing (WES) identified DNA aberrations in somatic mutations and copy number alterations for individual SRCCs. The transcriptome profiling also identified differentially expressed genes between tumor and normal samples, their involved pathways, and transcriptional regulators that may play a critical role in SRCCs. Our results revealed that the molecular basis of colorectal SRCC exhibits key differences from that of consensus colorectal adenocarcinoma (COAD).

Materials and Methods

Sample Collection

This study was approved by the Institutional Review Board (IRB) of Samsung Medical Center (IRB approval no. SMC 2013-11-008). Written informed consent was obtained from each patient. The study subjects were five Korean patients diagnosed with the SRCC subtype of colorectal cancer at Samsung Medical Center, Seoul, Korea. Tumor and matched normal tissues were obtained from surgical specimen. Peripheral blood mononuclear cells (PBMC) were obtained from each patient. PBMC and normal mucosa were used for genomic and transcriptomic analysis as a control, respectively. Samples were snap-frozen and stored in liquid nitrogen until use. Histology, clinical stage, and analyzed lesions are summarized in Table 1. Hematoxylin and eosin staining results for patient tumor samples are shown in Supplementary Figure 1.

Supplementary Table 1. Patient Characteristics of Five SRCCs

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>SRCC1</th>
<th>SRCC2</th>
<th>SRCC3</th>
<th>SRCC4</th>
<th>SRCC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>46</td>
<td>40</td>
<td>38</td>
<td>60</td>
<td>29</td>
</tr>
<tr>
<td>Gender</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>Tumor location</td>
<td>Descending colon</td>
<td>Rectum</td>
<td>Ascending colon</td>
<td>Rectum</td>
<td>Rectosigmoid junctions</td>
</tr>
<tr>
<td>MSI status</td>
<td>MSS</td>
<td>MSS</td>
<td>MSS</td>
<td>MSS</td>
<td>MSS</td>
</tr>
<tr>
<td>TNM stage</td>
<td>IVB</td>
<td>IVA</td>
<td>IVA</td>
<td>IVA</td>
<td>IVA</td>
</tr>
<tr>
<td>Metastasis</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lymphatic invasion</td>
<td>Positive</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Perineural invasion</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Vascular invasion</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
</tbody>
</table>

TCGA Colorectal Cancer Data Sets

We downloaded the 629 clinical, 489 raw mutation annotation, 623 mRNASeq, and 616 SNP6 copy number of COAD from Broad GDAC firehose (http://gdac.broadinstitute.org, version: 2016.01.28). Based on clinical information, all COADs are adenocarcinoma or mucinous carcinoma. Therefore, we used TCGA COAD data sets as a reference in this study. We choose 32 of 623 samples for differentially expressed gene analysis because those had matched normal-tumor pairs. Lists of TCGA samples and its clinicopathological features are summarized in Supplementary Table 3.

Isolation of Genomic DNA and RNA

Genomic DNA and RNA in tissues were purified using the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA). Genomic DNA from peripheral blood was extracted using the QIAamp DNA Blood Mini Kit (Qiagen). Genomic DNA concentration and purity were measured using a NanoDrop 8000 UV-Vis Spectrometer (Thermo Scientific Inc., Wilmington, DE) and a Qubit 2.0 Fluorometer (Life Technologies Inc., Grand Island, NY). To estimate DNA degradation, median DNA size and ΔCt values were measured using a 2200 TapeStation Instrument (Agilent Technologies, Santa Clara, CA) and real-time PCR (Agilent Technologies), respectively. For RNA, concentration and purity were measured using the NanoDrop and Bioanalyzer (Agilent Technologies).

Whole-Exome Sequencing

Genomic DNA (1 μg) from each sample was sheared using the Covaris S220 (Covaris, Woburn, MA) and used to construct a library using SureSelect XT Human All Exon V5 and a SureSelect XT Reagent Kit, HSQ (Agilent Technologies) according to the manufacturer’s protocol. This kit is designed to enrich 335,756 exons of 21,058 genes, covering ~71 Mb of the human genome. After enriched exome libraries were multiplexed, the libraries were sequenced using a HiSeq 2500 sequencing platform (Illumina, San Diego, CA). Briefly, a paired-end DNA sequencing library was prepared by gDNA shearing, end-repair, A-tailing, paired-end adaptor ligation, and amplification. After hybridization of the library with bait sequences for 16 hours, the captured library was purified and amplified with an index barcode tag, and library quality and quantity were measured. Sequencing of the exome library was carried out using the 100-bp paired-end mode of the TruSeq Rapid PE Cluster Kit and TruSeq Rapid SBS Kit (Illumina). Sequencing data and results are summarized in Supplementary Table 1.
Exome-seq Data Analysis
Sequencing reads were aligned to the UCSC hg19 reference genome (downloaded from http://genome.ucsc.edu), using Burrows-Wheeler Aligner [8] version 0.7.5a, with default settings. PCR duplicates were marked using Picard-tools-1.93 (http://picard.sourceforge.net/), data cleanup was performed with GATK, and variants were identified with GATK-3.5 [9]. Then, point mutations were identified using MuTect [10] and VarScan2 [11] with paired samples. An in-house Perl script and ANNOVAR [12] were used to estimate (Supplementary Figure 2). Fusion genes were purity based on whole-transcriptome sequencing was performed with and false discovery rate (FDR) resulting ratio was adjusted according to tumor purity. Number alterations were identified using CNVkit [14], and the resulting ratio was adjusted according to tumor purity.

RNA Sequencing
Library construction for RNA sequencing was performed using a TruSeq RNA Sample Preparation v2 Kit (Illumina). Isolated total RNA was used in a reverse transcription reaction with poly (dT) primers, using SuperScript II Reverse Transcriptase (Invitrogen/Life Technologies) according to the manufacturer’s protocol. Briefly, an RNA sequencing library was prepared by cDNA amplification, end-repair, 3′ end adenylation, adapter ligation, and amplification. Library quality and quantity were measured using the Bioanalyzer and Qubit. Sequencing of the RNA library was carried out using the 100-bp paired-end mode of the TruSeq Rapid PE Cluster Kit and the TruSeq Rapid SBS Kit (Illumina).

RNA-seq Data Analysis
Reads from the FASTQ files were mapped to the hg19 human reference genome using STAR version 2.5.0a in 2-pass mode [15], and gene quantification was performed using RSEM [16]. Expressed genes were defined as those with a transcripts per million (TPM) value of more than 10 across all samples to reduce the false-positive rate. Differentially expressed genes (DEGs) were identified using the edgeR package version 3.14.0 with a cutoff (|log2 fold-change| > 2 and false discovery rate (FDR) < 0.001) [17]. DEGs were mapped to the pathway using the REACTOME pathway database [18]. Tumor purity based on whole-transcriptome sequencing was performed with ESTIMATE [19] (Supplementary Figure 2). Fusion genes were detected using deFuse [20] (v0.6.1) and JAFFA (v1.06) [21] with default settings. The overlapped fusion transcripts between two prediction tools were considered as candidates.

Prediction of Transcription Regulators
iRegulon (Cytoscape plugin) detects master transcription regulators from a set of DEGs [22]. Regulators were predicted among the set of DEGs using iRegulon. Briefly, Cytoscape networks were created by importing the list of DEGs. The set of nodes (genes) was submitted to iRegulon and analyzed using the following options: 1) motif collection (10-kb region, 9,713 PWMs), 2) track collection (750 ChIP-seq tracks of ENCODE uniform signals), 3) putative regulatory region (10kb centered around TSS), 4) motif rankings database (10-kb region centered around TSS, 7 species), and 5) track of rankings database (10-kb region centered around TSS, ChIP-seq-derived).

Gene Set Enrichment Analysis
A gene set enrichment analysis (GSEA) [23] was conducted to analyze an SRCC-specific upregulated gene set. A GSEA prereanked analysis was performed by inputting a list of genes sorted according to fold changes of each TCGA COAD and SRCC.

qPCR Assay
Total RNA was extracted from patient tissue (RNAPrep Mini kit, Qiagen), and 500 ng RNA was subjected to reverse transcription using reverse transcriptase kit (Bioneer). Real-time quantitative PCR amplification was performed with a SYBR Green (ABI) in a real-time system (ABI, USA). Human-specific PCR primers (Bioneer) were used to analyze expression of the following genes: TWIST1, SNAI1, SNAI2, ZEB1, and GAPDH. mRNA levels of specific genes were calculated as ΔΔCt and normalized to GAPDH.

Western Blot
To prepare tissue extracts, tissues were lysed using a protein lysis buffer including a protease inhibitor. Then, 30 to 60 μg of protein extract was incubated with primary antibodies against TWIST1 (ab50887, Abcam), SNAI1 (#3895, Cell Signaling), SNAI2 (ab38551, Abcam), and ZEB1 (sc-25388, Santa Cruz). β-Actin (#3700, Cell Signaling) was used as normalized protein controls in Western blotting.

Results
Somatic Alterations in Colon SRCC
Somatic mutations frequently observed in TCGA nonhypermethylated COADs [24] were also identified in SRCCs (Figure 1A, Supplementary Figure 3). Most mutations were commonly found in SRCCs and TCGA COADs. However, APC nonsense mutation was observed at only one case (SRCC2) among five SRCCs (binomial test, $P = .006$), but there was a high frequency in COADs (81%). Mutations in three genes (USP6, DDX11, and CCDC166) having a very low frequency in COADs (≤1%) were relatively high in SRCCs (Supplementary Figure 4). Next, we performed a sequence context analysis (Figure 1B). Transition/transversion ratios were broadly consistent with those of COADs. A high frequency of C→T transitions is predominantly observed in COADs [25], and similarly, this transition type represented a large proportion of sequence changes in SRCCs. Furthermore, we identified chromosomal alterations in SRCCs. According to previous studies, COADs have recurrent chromosomal alterations, such as gains of 1q, 7p and q, 8p and q, 12q, 13q, 19q, and 20p and q and deletions of 1p, 4q, 5q, 8p, 14q, 15q, 17p and q, 18p and q, 20p, and 22q [24,26]. These alterations were consistently detected in SRCCs (Figure 2). Taken together, the results of somatic mutation and copy number alteration analyses at the whole-exome level revealed that SRCCs and COADs have overall similar properties at the genomic level, but APC gene mutation was uncommonly detected.

In addition, we found four novel fusion transcripts in the two SRCC cases (Supplementary Figure 5, A and B). One of the four fusions was interchromosomal fusion, and the remaining three were intrachromosomal fusions. Three of the four fusions showed the highest expression fold change (tumor vs. normal) of the partner gene in samples with fusion among the five SRCC samples (Supplementary Figure 5C).

Differentially Expressed Genes and Their Transcriptional Regulators
We investigated DEGs between SRCC tumors and matched normal colonic tissues. We detected 596 DEGs (upregulated: 383,
downregulated: 213) with $|\log_{2} \text{fold-change}| > 2$ and FDR $< 10^{-3}$ as criteria and mapped these genes to pathways (Figure 3, A and B). Upregulated genes were included in several pathways, including “extracellular matrix (ECM) organization,” “focal adhesion,” “integrin signaling pathway,” and “ECM-receptor interaction,” and these genes may be related to phenotypic traits of signet-ring cells and metastasis ability [27–30]. Other upregulated pathways were involved in cell survival and proliferation, such as the “PI3K/AKT and MAPK signaling pathways,” which are well-known signatures of cancer cells [31–35].

To identify the master transcription factors that regulate the DEGs, we conducted an iRegulon analysis (Figure 3C). We found four transcription factors ($\text{SRF}$, $\text{HNF4\text{A}}$, $\text{ZEB1}$, and $\text{RUNX1}$) that were significantly enriched in DEGs and compared their expression levels between tumor and normal tissues for each SRCC and TCGA COAD (Figure 3D, Supplementary Figure 6). All four transcription factors exhibited significantly altered gene expression levels (tumor vs. normal), and three of four showed different patterns in comparisons with TCGA COADs. For example, serum response factor (SRF) is a transcription activator that binds to the serum response element in the promoter region of target genes [36]. SRF gene expression was highly elevated in tumors compared with normal tissues in SRCCs but not in TCGA COADs. Hepatocyte nuclear factor 4 alpha (HNF4A) is a nuclear transcription factor that regulates the morphology and function of epithelial cells, and its gene expression was lower in SRCC tumors than normal tissues. These results indicate that altered expression levels of these transcription factors play a role in the tumorigenesis of SRCC.

**SRCC-Specific Upregulated Genes**

To identify signet-ring cell–specific dysregulated genes, we compared the expression levels between five SRCCs and 32 TCGA COADs (Figure 4A). After the removal of genes that did not exhibit expression differences ($P > .01$), we found 759 genes (208 up- and 366 downregulated genes in both SRCCs and COADs, 179 upregulated in SRCCs but downregulated in COADs, and 6 downregulated in SRCCs but upregulated in COADs). To identify the functional roles of SRCC-specific upregulated genes (179 genes; Figure 4A, bottom-right), we performed a GSEA. First, we selected the genes that had high fold-change values in SRCCs, which included the upregulated genes of both COADs and SRCCs and SRCC-specific upregulated genes. Then, we ordered genes by the fold change values (Figure 4A, from top-right to bottom-right).

As expected, enriched gene sets were “Colorectal Adenoma Up” and “Colorectal Adenoma Down” (Figure 4B upper). Second, we selected and sorted genes that had low fold-change values in COADs. Interestingly, SRCC-specific upregulated genes were enriched in gene sets of “Stem Cell Up Regulation” and “Epithelial-Mesenchymal Transition (EMT)” (Figure 4B lower and Figure 4C). The EMT is a process involved in the loss of polarity in epithelial cells and cell-cell adhesion. The progression to
Mesenchymal cells is related to the acquired migration ability, invasiveness, and increased production of ECM components [37]. In addition, this EMT process can generate cells with stem cell–like properties [38]. Taken together, our results suggest that SRCCs may acquire potential stem cell–like characteristics via the EMT process.

**Figure 2.** Copy number alterations in both SRCCs and TCGA COADs. Notes: The copy number alterations are plotted across chromosomes (red: amplification, blue: deletion, black: diploid).

**Figure 3.** Differentially expressed genes, pathways, and their transcription factors. Notes: (A) Heatmap of differentially expressed genes in SRCCs. Enriched pathways among up/downregulated genes are also presented. (B) Pathway-based functional interaction network. Genes belonging to same pathway are closely linked to each other. (C) Prediction of transcription factors that regulate differentially expressed genes. (D) Boxplot of the expression levels of SRF and HNF4A. Expression levels in TCGA COADs with tumor and matched normal samples were compared.
The main characteristics of signet-ring cells are a loss of cell-cell interactions and accumulation of intracellular mucin. Typically, epithelial cells express high levels of E-cadherin (CDH1), whereas mesenchymal cells express N-cadherin (CDH2) [39]. We investigated the expression of genes encoding E and N-cadherin and their transcription factors to estimate the EMT signature in SRCCs (Figure 5A). The expression of E-cadherin was lower in tumor tissues than in normal tissues, whereas N-cadherin expression was higher in tumor tissues than in normal tissues. Gene expression levels of transcription factors that directly or indirectly regulate E-cadherin, such as SNAI1, SNAI2, ZEB1, ZEB2, TWIST1, TCF4, ETS1, and RUNX2, increased during the tumorigenesis of signet-ring cells. In addition, ELF3 is a negative regulator of SNAI1, ZEB, and TWIST1, which exhibit decreased expression in SRCC tumors [40]. These results indicate that SRCCs undergo the EMT process, which is the main cause of the loss of cell-cell interactions.

To determine which mucin-producing genes are activated in SRCCs, we compared expression levels in tumor vs. normal samples of human mucin genes (Figure 5B). Nonexpressed mucin genes are not shown here. Among mucin genes, MUC2 was more highly expressed (3.7-fold) in tumor than in normal tissues, but other genes were either decreased or unchanged. In COADs, the expression levels of all mucin genes, including MUC2, decreased (Supplementary Figure 7). These results indicate that mucin accumulation in signet-ring cells is caused by the abnormal expression of MUC2 [41–43].

**Discussion**

In this study, we first characterized somatic alterations on the genome-wide scale for colorectal SRCC by whole-exome and RNA sequencing. Previous SRCC studies are focused only few cancer gene mutations and expressions or case reports. However, this work provides genomic and transcriptomic alterations based on massive parallel sequencing simultaneously. As compared with those for SRCCs and TCGA COADs, we demonstrated interesting findings that have not been previously reported about SRCCs. Mutations in the APC gene are the initial genetic alterations in colorectal tumorigenesis [44]. However, the low APC mutation rate in SRCCs indicates that they undergo a developmental process different from that of COAD. We also validated the low APC mutation rate (17%) in additional 6 SRCC cases (Supplementary Figure 9A). A previous study reported a distinct pattern of KRAS gene mutations in SRCCs; specifically, there is a lower mutation frequency at codon 12 (2 of 16) but a higher mutation frequency at codon 61 (4 of 16) as compared with those of non-SRCCs [45]. In our study, two samples had KRAS mutations at codon 12 (SRCC2: G12C, SRCC4: G12D), and mutation at codon 61 was not detected. A BRAF mutation was not detected in our SRCCs (Figure 1A, Supplementary Figure 9A). This might be explained by their microsatellite stable (MSS) status. Previous studies have reported a high frequency of BRAF (V600E) mutations in SRCCs; specifically, there is a lower mutation frequency at codon 12 (2 of 16) but a higher mutation frequency at codon 61 (4 of 16) as compared with those of non-SRCCs [45]. In our study, two samples had KRAS mutations at codon 12 (SRCC2: G12C, SRCC4: G12D), and mutation at codon 61 was not detected. A BRAF mutation was not detected in our SRCCs (Figure 1A, Supplementary Figure 9A). This might be explained by their microsatellite stable (MSS) status. Previous studies have reported a high frequency of BRAF (V600E) mutations in SRCCs, including high-level microsatellite instability (MSI-H) [46–48]. In TCGA COAD with hypermutated tumors, the mutation frequency in BRAF (V600E) is approximately 41% [24], but this frequency is much lower in nonhypermutated tumors. We
have also validated the gene mutations through the Sanger sequencing and both whole-exome and RNA sequencing (Supplementary Figure 3). Recurrent fusion genes in colorectal cancer [49] were not detected in SRCC samples, but we found several novel fusion genes (Supplementary Figure 5). However, the functional implications of these fusion genes are unknown. We found SRCC-specific expression changes for transcription factors such as $HNF4A$, $ZEB1$, and $RUNX1$ that may regulate target genes in SRCCs. These target genes are involved in butyrate and fatty acid metabolism, and PPAR signaling pathway. Short-chain fatty acids, which include butyrate, are produced in the human colon by bacterial fermentation and suppressor of colorectal cancer [50–52]. In addition, peroxisome proliferator-activated receptors (PPARs) are ligand-activated transcription factors that are bound to oxidized fatty acids and regulate the differentiation of cells [53,54]. This means that butyrate and fatty acid metabolism, and PPAR signaling pathway are related to each other. In SRCCs, butyrate and fatty acid metabolism, and PPAR signaling pathways were all downregulated. Thus, our results support the previous reports about tumor suppressive role of butyrate and fatty acid metabolism, and PPAR signaling pathway in colorectal cancer.

Finally, we examined the relationship between clinicopathological features and genetic abnormalities in SRCCs. Highly upregulated genes were enriched in the EMT process. We selected four of eight transcription factors that regulate highly upregulated genes in SRCCs to measure the mRNA and protein expression levels using qPCR and Western blotting, respectively (Supplementary Figure 8, A and B). Although the mRNA expressions of qPCR are similar to those of RNA-Seq, protein expressions are different from those of mRNA expressions (Supplementary Table 2). There may be several reasons for discrepancy between mRNA and protein expressions such as posttranscriptional mechanisms, difference of in vivo proteins half-lives, and experimental error in measuring the amount of both mRNA and protein [55]. In addition, $MUC2$ overexpression is considered as a major cause of the accumulation of mucin in signet-ring cells. Four transcription factors ($SRF$, $HNF4A$, $ZEB1$, and $RUNX1$) and $MUC2$ expressions were also validated using public expression data of SRCCs (GSE79793 [56]), and those gene expressions were corresponded to our study except for $ZEB1$ (Supplementary Figure 9B). Our results provide important insight to the understanding of molecular alterations in SRCCs.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2018.04.007.

References


