Three DNA Methylation Epigenotypes in Human Colorectal Cancer

Koichi Yagi1,3, Kiwamu Akagi4, Hiroshi Hayashi1, Genta Nagae4, Shingo Tsuji1, Takayuki Isagawa1, Yutaka Midorikawa1, Yoji Nishimura5, Hirohiko Sakamoto5, Yasuyuki Seto3, Hiroyuki Aburatani1, and Atsushi Kaneda1,2

Abstract

Purpose: Whereas the CpG island methylator phenotype (CIMP) in colorectal cancer associates with microsatellite instability (MSI)-high and BRAF-mutation(+), the existence of an intermediate-methylation subgroup associated with KRAS-mutation(+) is controversial, and suitable markers for the subgroup have yet to be developed. Our aim is to clarify DNA methylation epigenotypes of colorectal cancer more comprehensively.

Experimental Design: To select new methylation markers on a genome-wide scale, we did methylated DNA immunoprecipitation-on-chip analysis of colorectal cancer cell lines and re-expression array analysis by 5-aza-2′-deoxycytidine/Trichostatin A treatment. Methylation levels were analyzed quantitatively in 149 colorectal cancer samples using matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry. Colorectal cancer was epigenotyped by unsupervised two-way hierarchical clustering method.

Results: Among 1,311 candidate silencing genes, 44 new markers were selected and underwent quantitative methylation analysis in colorectal cancer samples together with 16 previously reported markers. Colorectal cancer was clustered into high-, intermediate-, and low-methylation epigenotypes. Methylation markers were clustered into two major groups: group 1 showing methylation in high-methylation epigenotype, and group 2 showing methylation in high- and intermediate-methylation epigenotypes. A two-step marker panel deciding epigenotypes was developed with 95% accuracy: the 1st panel consisting of three group-1 markers (CACNA1G, LOX, SLC30A10) to extract high-methylation epigenotype, and the 2nd panel consisting of four group-2 markers (ELMO1, FBN2, THBD, HAND1) and SLC30A10 again to divide the remains into intermediate- and low-methylation epigenotypes. The high-methylation epigenotype correlated significantly with MSI-high and BRAF-mutation(+) in concordance with reported CIMP. Intermediate-epigenotype significantly correlated with KRAS-mutation(+). KRAS-mutation(+)-colorectal cancer with intermediate-methylation epigenotype showed significantly worse prognosis.

Conclusions: Three methylation epigenotypes exist in colorectal cancer, and suitable classification markers have been developed. Intermediate-methylation epigenotype with KRAS-mutation(+) correlated with worse prognosis. Clin Cancer Res; 16(1); 21–33. ©2010 AACR.

Colorectal cancer is the fourth most common cancer and the fourth leading cause of cancer death in the world (1). Cancer arises as a consequence of the accumulation of epigenetic alterations and genetic alterations (2–5). Gene mutations, such as KRAS, p53, and APC, are well-known genetic alterations occurring in colorectal cancer, which were shown in the model of “adenoma-carcinoma sequence” by Bert Vogelstein (6). Microsatellite instability (MSI) and chromosomal instability are also well-characterized genetic alterations, and these two exclusive pathways suggest the heterogeneous origin of colorectal cancer (7).

Gene silencing is an epigenetic gene inactivation mechanism by DNA methylation of its promoter region, and is involved in the initiation and progression of cancer (4, 8). A genome-wide search of genes using aberrant methylation as markers is useful for identifying novel tumor-suppressor genes and methylation markers (9–12). Several approaches to detect aberrantly methylated regions in cancer have been developed, although most of these methods utilize methylation-sensitive restriction enzymes and thus only limited regions of genome can be analyzed (13).
Translational Relevance

Whereas CpG island methylator phenotype (CIMP) in colorectal cancer is known to associate with microsatellite instability and BRAF-mutation(+), the existence of an intermediate-methylation subgroup associated with KRAS-mutation(+), referred to as CIMP-low/CIMP-2, is controversial, and suitable markers for the subgroup have yet to be developed, if such exist. We identified new colorectal cancer methylation markers through genome-wide analyses of DNA methylation and expression, and the new methylation markers classified clinical colorectal cancer cases into three distinct clusters: high-, intermediate-, and low-methylation epigenotypes. The intermediate-methylation epigenotype strongly correlated with KRAS-mutation(+). Although microsatellite-stable colorectal cancer with worse prognosis was not stratified by using KRAS status only, KRAS-mutation(+) colorectal cancer with intermediate-methylation epigenotype showed significantly worse prognosis. Considering that KRAS-mutation(+) colorectal cancer is resistant to anti-epidermal growth factor receptor therapy, epigenotyping using new markers might be useful to stratify KRAS-mutation(+) colorectal cancer with worse prognosis, and alternative therapy targeting for this group is suggested to be developed.

As for accumulation of gene methylation, the concept termed "CpG island methylator phenotype" (CIMP) was first represented in 1999 by Toyota et al. CIMP was characterized by CpG island methylation in multiple regions, the so-called MINT (methylated-in-tumor) loci, existing in a subset of colorectal cancer, and correlated with MSI (14). In 2006, Weisenberger et al. reported new five CIMP markers to decide CIMP status using quantitative methylation analysis, Methylight (15). Ogino et al. proposed that the CIMP-low subgroup, which had less extensive methylation of CIMP-related markers, was associated with KRAS mutation compared with the CIMP-high and CIMP(-) subgroups (16). Shen et al. reported that in addition to CIMP correlating with MSI (namely CIMP-1), another CIMP (namely CIMP-2) existed in the remaining colorectal cancer cases through genetic and epigenetic analysis (17). However, these reports about CIMP-low/2 analyzed methylation of genes/loci that had been previously reported and frequently analyzed, such as p16, hMLH1, and MINT loci. As Ogino and Fuchs suggested, CIMP markers so far were considered to be specific for CIMP-high and not ideal for the identification of CIMP-low, therefore, sensitive and specific markers for CIMP-low needed to be identified if CIMP-low really existed (18). On the other hand, Perucho and his colleagues insisted on the absence of methylator phenotype, by analyzing methylation in as many as 36 genes (19). These suggest that the existence of the CIMP-low/2 subgroup or even CIMP itself is controversial, and epigenotyping of colorectal cancer by more comprehensive approaches, e.g., hierarchical clustering using quantitative methylation data by a single detection method using novel regions selected on a genome-wide scale, were required.

To enable high-resolution comprehensive mapping of DNA methylation, we recently reported methylated DNA immunoprecipitation (MeDIP)-chip analysis on short oligonucleotide arrays using unbiased amplification of MeDIP products by in vitro transcription (20). In this study, we did MeDIP-chip analysis of both the MSI-high (MSI-H) colorectal cancer cell line HCT116 and the microsatellite-stable (MSS) colorectal cancer cell line SW480 using promoter tiling array to identify promoter regions methylated in colorectal cancer on a genome-wide scale. HCT116 and SW480 were treated with 5-aza-2'-deoxycytidine (5-Aza) and/or Trichostatin A, and gene re-expression of the cell lines was analyzed using expression microarray. Candidate genes silenced by promoter methylation in colorectal cancer were selected using these methylation and re-expression data, and methylation of these marker genes and reported CIMP markers was analyzed using a single, bisulfite-PCR-based quantitative method, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MassARRAY), to epigenotype colorectal cancer.

Materials and Methods

Clinical samples and cell lines. Two hundred and thirteen colorectal cancer samples and nine normal colon mucosa samples were obtained from colorectal cancer patients who underwent surgical operation at Saitama Cancer Center from January 1996 to August 2006, with written informed consents, and kept frozen until use. Colorectal cancer samples were microscopically examined for determination of cancer cell contents by two independent pathologists, and were dissected to enrich cancer cells when necessary. For molecular analysis, 149 samples that contained ≥40% of cancer cells were used. Normal colon mucosa was extracted with enough margin from cancer, and it was confirmed microscopically that neither muscle layer nor cancer cells were included. Three MSI-H colorectal cancer cell lines, namely, HCT116, DLD1, and LoVo, and three MSS colorectal cancer cell lines, namely, SW480, T84 and Caco-2\(^6\) (21–23), were obtained from the American Type Culture Collection. DNA of clinical samples and cell lines was extracted using QIAamp DNA Micro Kit (QIAGEN). RNA of HCT116 and SW480 was extracted using TRIzol (Invitrogen). Human peripheral lymphocyte DNA was obtained from Coriell Cell Repositories. RNA samples of normal adult and fetal colon were purchased from Ambion and Stratagene, respectively. This

\(^6\) www.sanger.ac.uk/perl/genetics/CGP
study was certified by the Ethics Committee of Tokyo University and Saitama Cancer Center.

**5-Aza and Trichostatin A treatment.** HCT116 and SW480 cells were seeded at a density of 3 × 10^5 cells/10-cm dish on day 0, and exposed to 3 μmol/L 5-Aza on days 1, 2, and 3. Trichostatin A treatment was done on day 3 at a dose of 300 nmol/L. Medium was changed every 24 h, and cells were harvested on day 4.

**MeDIP-Chip analysis.** MeDIP-Chip of HCT116 and SW480 using Human Promoter 1.0R tiling array (Affymetrix), which covered 199,543,165 bp and 4,071,296 CpG sites around over 25,500 promoter regions, was carried out as we previously reported with MeDIP-chip using ENCODE tiling array (20). Briefly, genomic DNA of HCT116 and SW480 was fragmented by sonication and immunoprecipitated by anti-5-methylcytocine polyclonal antibody (Megabase Research Products). MeDIPed sample and input sample underwent unbiased amplification by in vitro transcription. Amplified cRNA was converted into cDNA. cDNA was hybridized to Promoter tiling array for both MeDIPed sample and input sample. Hybridization was done twice. Within a window of 550 bp, the duplicated data from MeDIP DNA were compared with duplicated data from input DNA using Wilcoxon rank sum test to calculate P values to detect candidate methylation sites. Sites with P < 10^-5 were considered to be candidate methylation sites in this study. MeDIP-chip data are available at GEO datasets (#GSE14526).

**Expression microarray analysis.** mRNA expression in HCT116 and SW480 with or without 5-Aza/Trichostatin A treatment and normal adult and fetal colon samples were analyzed on GeneChip Human Genome U133 plus 2.0 oligonucleotide arrays (Affymetrix). For global normalization, the average signal in an array was made equal to 100. Expression array data were used for selection of methylation markers. Expression array data are available at GEO datasets (#GSE14526).

**Bisulfite treatment.** Genomic DNA bisulfite conversion of clinical samples, cell lines, and control samples was done as described previously (24) with some modification (See Supplementary Material and Methods).

**Methylation analysis.** Methylation analysis of clinical samples, cell lines, and control samples was done using MassARRAY (Sequenom; ref. 25). Bisulfite-treated DNA was amplified by PCR, and the PCR product was transcribed by in vitro transcription and the RNA was cleaved by RNaseA. Unmethylated cytosine (C) was converted to uracil (U) by bisulfite treatment, i.e., thymine (T) in PCR product, and finally adenine (A) in in vitro transcription product. Methylated cytosine (mc) was not converted, i.e., cytosine (C) in PCR product, and finally guanine (G) in in vitro transcription product. RNaseA cleaves RNA at the 3′ site of both T and C; T-specific cleavage was able to be done by containing dC instead of C in the mixture, C-specific cleavage by containing dT instead of T in the mixture. Methylation status was determined by mass difference between A and G in a cleaved RNA product. Quantitative methylation was calculated for each cleaved product. This analytic unit containing several CpG sites in a cleaved product was called “CpG unit”. We did only T-specific cleavage analysis because C-specific cleavage is not informative at the high-CpG-density region (26).

Primers were designed to include no CpG site or only one CpG site in 5′ region of primers, and are shown in Supplementary Table S1. Control samples were prepared as follows. First, human peripheral lymphocyte DNA was used as diploid human DNA, and was amplified by GenomiPhi v2 DNA amplification kit (GE Healthcare Life-sciences). The amplified DNA was not methylated at all in any CpG sites, and was used as unmethylated (0%) control. The amplified DNA was also methylated by Sso methylase (New England Biolabs) and used as fully methylated (100%) control. By mixing the 0% and 100% control samples, partially methylated control samples (25%, 50%, and 75%) were generated.

**Analysis of genetic alterations.** Analysis of MSI status and KRAS mutation (codons 12, 13, and 19) was done as described previously (27, 28). Samples showing instability in two or more of the Bethesda five markers were defined as MSI-H, and the ones in none or one as MSS. BRAF mutation (V600E) at exon 15 was determined by direct sequencing using pyrosequencing. P53 immunohistochemistry was done using DO-7 antimouse monoclonal antibody (Santa Cruz Biotechnology), and samples with nucleus staining were determined as p53-IHC (+). Analysis of MSI, KRAS mutation, BRAF mutation, and p53-IHC were done for all the clinical colorectal cancer samples.

**Statistical analysis.** The correlation between epigenotypes and clinicopathologic factors except age was analyzed by Fisher’s exact test. About patient age, ANOVA was used for analysis of significance among three groups, and comparison between two groups was adjusted for multiple comparisons using Holm’s method. Unsupervised two-way hierarchical clustering was done based on standard correlation and average linkage clustering algorithm in sample direction, and Euclidian distance and complete linkage clustering algorithm in marker direction using GeneSpring 7.3.1 software (Agilent Technology). Model-based clustering was carried out using R software. Decision tree for three epigenotypes was generated using Weka software. Kaplan-Meier survival analysis was done by JMP 7 software and P value was calculated by log-rank test. Survival analysis by Cox proportional hazard model was done by R software. In survival analysis, the end of follow-up period was 60 mo from the primary surgery and the mean follow-up time of the cases used for survival analysis was 49 mo. Death as a result of colorectal cancer was the primary end point and deaths by other causes were censored.
Results

Generation of colorectal cancer methylation markers. In MeDIP-Chip analysis of HCT116 and SW480 using Promoter tiling array, 12,782 and 10,232 candidate methylation sites were detected on a genome-wide scale, respectively, and 3,178 and 3,177 RefSeq genes were found to possess candidate methylation sites within 1 kb from transcription start site. In expression analysis of HCT116 and SW480 treated by 5-Aza with or without Trichostatin A, 9,988 and 10,550 genes showed re-expression >1.5-fold compared with untreated ones, respectively. To select candidate silencing genes, we set up four criteria: (a) genes that possess candidate methylation site within 1 kb

![Gene selection from array-based analysis. A, Venn diagram showed the intersection of two categories acquired by genome-wide array-based analysis. Numbers, number of included genes; numbers in parentheses, number of genes selected for MassARRAY analysis (above arrows) and genes whose primer pairs were validated and used for further analysis (below arrows). B, representative genes selected from three categories. CHFR was selected from HCT116, HLTF from SW480, and SFRP1 from both cell lines. MeDIP score is shown by $-\log_{10}P$ value and the region over 20 (i.e., $P < 10^{-2}$) was considered as a candidate methylated site.](image-url)
from transcription start site; (b) GeneChip score of normal adult or fetal colon >50; (c) GeneChip score of the cell line <50; and (d) upregulation >1.5-fold after 5-Aza or 5-Aza and Trichostatin A treatment. As shown in Fig. 1, 577, 313, and 421 genes were identified as candidate silencing genes fulfilling the four criteria in HCT116 only, in SW480 only, and in both cell lines, respectively. Among these, we selected 21, 10, and 24 genes as candidate new markers, respectively (Fig. 1). In addition, we also selected 13 previously reported CIMP markers and 6 previously reported silenced genes.

Quantitative methylation analysis. We first validated quantitativity 74 primer pairs by analyzing control samples (0%, 25%, 50%, 75%, and 100% methylated samples) using MassARRAY (Supplementary Fig. S1). A linear standard curve was drawn, and correlation coefficient \( R^2 \) was calculated at each CpG unit. CpG units with \( R^2 \leq 0.9 \) were excluded for further analysis. Primer pairs whose amplicon contained <3 CpG units with \( R^2 >0.9 \) were also excluded or redesigned. Finally, primers for 60 regions, including 44 candidate new markers, 11 CIMP markers, and 5 silencing genes (Supplementary Table S1), were validated and used in further analysis, whereas the other 14 markers were excluded. The 44 new markers included 16, 8, and 20 candidate genes silenced in HCT116 only, in SW480 only, and in both cell lines (Fig. 1A).

The amplicons of 60 analyzed regions included 15,352 base pairs and 791 CpG units derived from 1,455 CpG sites. Methylation scores were measurable at 640 CpG units by MassARRAY. In addition to CpG units with correlation coefficient \( R^2 \leq 0.9 \), CpG units where methylation...
Table 1. Clinical and molecular characteristics according to three epigenotypes

<table>
<thead>
<tr>
<th>Clinical or molecular features</th>
<th>All cases</th>
<th>HME</th>
<th>IME</th>
<th>LME</th>
<th>Outlier</th>
<th>P (3 epigenotypes)</th>
<th>P (IME vs LME)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample number</td>
<td>149 (100%)</td>
<td>17</td>
<td>60</td>
<td>54</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>85 (57%)</td>
<td>5</td>
<td>35</td>
<td>35</td>
<td>10</td>
<td>0.038*</td>
<td>0.56</td>
</tr>
<tr>
<td>Female</td>
<td>64 (43%)</td>
<td>12</td>
<td>25</td>
<td>19</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>62.8 ± 9.4</td>
<td>69.0 ± 8.9</td>
<td>60.9 ± 9.2</td>
<td>62.1 ± 9.6</td>
<td>65.2 ± 6.8</td>
<td>8.5 × 10^-3x</td>
<td>0.51</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximal</td>
<td>54 (36%)</td>
<td>15</td>
<td>20</td>
<td>10</td>
<td>9</td>
<td>8.7 × 10^-7x</td>
<td>0.090</td>
</tr>
<tr>
<td>Distal</td>
<td>95 (64%)</td>
<td>2</td>
<td>40</td>
<td>44</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucinous component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>27 (18%)</td>
<td>9</td>
<td>16</td>
<td>0</td>
<td>2</td>
<td>9.7 × 10^-8x</td>
<td>1.4 × 10^-5†</td>
</tr>
<tr>
<td>(-)</td>
<td>122 (82%)</td>
<td>8</td>
<td>44</td>
<td>54</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly diff. component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>20 (13%)</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>7.5 × 10^-3x</td>
<td>0.13</td>
</tr>
<tr>
<td>(-)</td>
<td>129 (87%)</td>
<td>11</td>
<td>51</td>
<td>51</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AJCC Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>2 (1%)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.053</td>
<td>0.29</td>
</tr>
<tr>
<td>II</td>
<td>42 (28%)</td>
<td>6</td>
<td>19</td>
<td>11</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>50 (34%)</td>
<td>6</td>
<td>19</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>55 (35%)</td>
<td>3</td>
<td>22</td>
<td>19</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>90 (60%)</td>
<td>9</td>
<td>36</td>
<td>38</td>
<td>7</td>
<td>0.27</td>
<td>0.23</td>
</tr>
<tr>
<td>(-)</td>
<td>58 (39%)</td>
<td>8</td>
<td>24</td>
<td>15</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (1%)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph vessel invasion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>117 (79%)</td>
<td>15</td>
<td>48</td>
<td>40</td>
<td>14</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td>(-)</td>
<td>32 (21%)</td>
<td>2</td>
<td>12</td>
<td>14</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venous invasion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>122 (82%)</td>
<td>12</td>
<td>49</td>
<td>45</td>
<td>16</td>
<td>0.48</td>
<td>1</td>
</tr>
<tr>
<td>(-)</td>
<td>27 (18%)</td>
<td>5</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsatellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-H</td>
<td>17 (11%)</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4.0 × 10^-15x</td>
<td>0.50</td>
</tr>
<tr>
<td>MSS</td>
<td>132 (89%)</td>
<td>4</td>
<td>58</td>
<td>54</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>13 (9%)</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2.0 × 10^-15x</td>
<td>1</td>
</tr>
<tr>
<td>(-)</td>
<td>136 (91%)</td>
<td>5</td>
<td>60</td>
<td>54</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>60 (40%)</td>
<td>3</td>
<td>38</td>
<td>14</td>
<td>5</td>
<td>2.3 × 10^-5x</td>
<td>7.4 × 10^-5†</td>
</tr>
<tr>
<td>(-)</td>
<td>89 (60%)</td>
<td>14</td>
<td>22</td>
<td>40</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p53-IHC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>70 (47%)</td>
<td>0</td>
<td>29</td>
<td>30</td>
<td>11</td>
<td>2.4 × 10^-5x</td>
<td>0.50</td>
</tr>
<tr>
<td>(-)</td>
<td>75 (50%)</td>
<td>17</td>
<td>29</td>
<td>22</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>4 (3%)</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The number of samples with each clinical or molecular characteristics is shown. Proximal location, cecum to transverse colon. Distal location, descending colon to rectum. Poorly diff. component, poorly differentiated adenocarcinoma component. Abbreviation: AJCC, American Joint Committee on Cancer.

*P value among three epigenotypes < 0.05.
†P value between IME and LME < 0.05.

Methylation of 60 markers in HCT116, SW480, and clinical samples. Thirty-six markers selected from candidate silencing genes in HCT116 were mostly methylated in HCT116 (mean ± SD, 92 ± 11%), and 28 markers from...
SW480 were mostly methylated in SW480 (80 ± 21%). Of 44 new marker genes, 37 (84%) showed methylation(+) in >10% of the 149 colorectal cancer samples, and 30 of the 37 genes showed methylation(−) in all the 9 normal samples (Supplementary Fig. S2). Among these 30 cancer-specific aberrant-methylation genes, 16 (53%) showed methylation(+) in >50% of 149 colorectal cancer samples. On the other hand, all the 11 CIMP-related markers showed methylation(+) in <50% of colorectal cancer samples, and 9 of 11 showed methylation(+) in only 8% to 20% of colorectal cancer samples. Considering that the tumor cell content of the sample was ≥40%, a clinical sample with methylation rate >35% was defined methylation(+).

Epigenotyping of colorectal cancer. To epigenotype colorectal cancer by DNA methylation, 149 colorectal cancer samples along with 9 normal colon mucosa samples and 6 colorectal cancer cell lines were classified by unsupervised two-way hierarchical clustering using the methylation rates of the above 60 markers (Fig. 2). Three methylation clusters were identified: high-methylation epigenotype (HME), intermediate-methylation epigenotype (IME), and low-methylation epigenotype (LME). All the 3 MSI-H cell lines and 17 colorectal cancers were clustered in HME, all the 3 MSS cell lines and 60 colorectal cancers were clustered in IME, and 54 colorectal cancers were clustered in LME. Nine normal samples were clustered outside from three colorectal cancer epigenotypes, and 18 outlier samples were excluded for further analysis. The Bayesian information criterion curve constructed by model-based clustering revealed that the most optimal number of clusters for colorectal cancer was three (data not shown), supporting the above classification of colorectal cancer into three epigenotypes.

HME consisted of 17 samples, and included 13 (76%) MSI-H, 12 (71%) BRAF-mutation(+), 3 (18%) KRAS-mutation(+), and no p53-IHC(+) samples (Fig. 2 and Table 1). IME consisted of 60 samples, and included 2 (3%) MSI-H, no BRAF-mutation(+), 38 (63%) KRAS-mutation(+), and 29 (48%) p53-IHC(+) samples. LME consisted of 54 samples, and included no MSI-H, no BRAF-mutation(+), 14 (26%) KRAS-mutation(+), and 30 (56%) p53-IHC(+) samples. There was a significant difference among the three epigenotypes in MSI status (P = 4.0 × 10^{-13}, Fisher’s exact test), BRAF-mutation (P = 2.0 × 10^{-13}), KRAS-mutation (P = 2.3 × 10^{-5}), and p53-IHC (P = 2.4 × 10^{-5}), in concordance with previous reports (17). When only IME and LME were compared, there was a significant difference in KRAS-mutation (P = 7.4 × 10^{-5}) and no difference in p53-IHC status (P = 0.45).

As for clinical information (Table 1), gender, proximal tumor location, mucinous component of the tumor, and poorly differentiated adenocarcinoma component were significantly different among the three epigenotypes (P = 0.038, 8.7 × 10^{-7}, 9.7 × 10^{-8}, and 7.5 × 10^{-7}, respectively), and HME patients were significantly older.
Types of methylation markers. Methylation markers were clustered into two major groups, group 1 and group 2 (Fig. 2). Group-1 markers seemed to be highly methylated in HME. Group-2 markers seemed to be highly methylated in both HME and IME, but less in LME. To estimate these tendencies, we classified 60 markers into four types by average methylation rate for each epigenotype (Fig. 3). Type-1 markers \((n = 13)\) showed significantly different methylation rate \((P < 0.01, t\)-test\) between HME and IME, but no difference \((P > 0.01)\) between IME and LME. Type-2 markers \((n = 22)\) showed significantly different methylation rate between HME and IME, and also between IME and LME. Type-3 markers \((n = 13)\) showed no different methylation rate between HME and IME, but significant difference between IME and LME. Type-4 markers \((n = 12)\) showed no difference both between HME and IME, and between IME and LME. Nineteen group-1 markers consisted of 10 type-1 and 9 type-2 markers, whereas 28 group-2 markers consisted mostly of type-2 \((n = 13)\) and type-3 \((n = 11)\) markers. These indicated that informative methylation markers were classified into two groups: group 1 to distinguish HME from two other epigenotypes, and group 2 to distinguish IME and LME. Among 11 CIMP-related markers, 7 were included in type 1; the other four were included in type 2, but three of the four except \textit{NEUROG1} showed very small difference of methylation rate between IME and LME, and therefore were very close to type 1.

Two-step panel method to predict colorectal cancer epigenotypes. To decide colorectal cancer epigenotypes easily without hierarchical clustering, we first extracted the best decision tree model by examining a random combination of 60 markers by Weka software (Supplementary Fig. S3). This decision tree classified colorectal cancer with 91% accuracy by two steps: the first step was to extract HME by one type-1 marker, \textit{LOX}, and the second step was to
divide the remains into IME and LME using two type-2 markers and two type-3 markers. As indicated by clustering method and marker types, this decision tree confirmed that three epigenotypes could be obtained by extracting HME at the first step and dividing the remains into IME and LME at the second step. To avoid high dependency on one marker in the decision tree method, we next set up two marker panels consisting of three to five markers by examining the entire combination of 60 markers again (Fig. 4A). In panel 1, samples were regarded as HME if two or three markers among LOX, CACNA1G, and SLC30A10 were methylation(+) . If none or one marker was methylation(+) in panel 1, the samples were proceeded to panel 2, in which samples were regarded as IME when three or more markers among ELMO1, FBN2, THBD, HAND1, and SLC30A10 were methylation(+), and samples were regarded as LME when two or less markers were methylation(+). These marker panels were selected as the best combination using a training set of 90 colorectal cancer samples, and validation was done using a test set of 41 colorectal cancer samples, giving 95% accuracy in epigenotype prediction. Panel 1 markers consisted of one type-1 and two type-2 markers, whereas panel 2 markers consisted of three type-2 and two type-3 markers (Figs. 3 and 4A).

If all the 149 colorectal cancer samples were classified by this panel method, there would be 19 HME, 63 IME, and 67 LME cases. Significant differences among three epigenotypes were maintained in MSI status, BRAF mutation, KRAS mutation, and p53-IHC. Significant difference in KRAS mutation was also maintained between IME and LME (P = 3.9 × 10^{-5}; Supplementary Table S2).

Fig. 5. Comparison of three colorectal cancer epigenotypes with classical CIMP status. A, colorectal cancer with methylation(+) in ≥6 among 11 CIMP markers (CACNA1G, IGF2, NEUROG1, RUNX3, SOCS1, hMLH1, p16INK4A, MINT1, MINT2, MINT17, and MINT31) was considered as CIMP-high; 1 to 5 of 11 as CIMP-low; and 0 of 11 as CIMP-negative. B, 113 among 149 samples (76%) were classified into the corresponding groups (CIMP-high to HME, CIMP-low to IME, and CIMP-negative to LME). Genetic and clinical characteristics were compared. Top, classical CIMP; middle, epigenotype decided by two-step panel method; bottom, epigenotype by clustering. Difference between CIMP-low and CIMP(-) or between IME and LME was analyzed by Fisher’s exact test (**, P < 0.01; *, P < 0.05).
**Prediction of clinical outcome by epigenotype.** In MSS colorectal cancer samples, unadjusted overall survival of KRAS-mutation(+) colorectal cancer and KRAS-mutation(−) colorectal cancer was analyzed by Kaplan-Meier method, and showed no significant difference \( (P = 0.074, \text{log-rank test; Fig. 4B}) \). If epigenotype information was added, KRAS-mutation(+) colorectal cancer with IME showed significantly worse prognosis \( (P = 0.027, \text{log-rank test; Fig. 4C and D}) \). When clinical factors shown in Table 1 were compared between KRAS-mutation(+) colorectal cancer with IME and the others, the mucinous component factor and the tumor location factor showed significant difference \( (P = 2.4 \times 10^{-3} \) and \( 1.8 \times 10^{-3} \), respectively). Survival analysis by Cox proportional hazard model was also done; KRAS-mutation(+) colorectal cancer with IME showed significantly worse prognosis in unadjusted analysis \( (P = 0.027) \) and in analyses adjusted for the mucinous factor, the location factor, and the both factors \( (P = 0.018, 0.043, \) and \( 0.029 \), respectively). When analyzing colorectal cancer at stages III and IV only, the significant difference was reproduced (Supplementary Fig. S4). By epigenotype information only, IME showed no significant difference in overall survival compared with LME \( (P = 0.18, \text{log-rank test}) \). The rationale of the sample size is as follows. Considering the 5-year survival rate of colorectal cancer at each clinical stage in Japan \(^\text{10}\) and the number of samples at each stage in this study, the average 5-year survival rate was calculated to be 49.8%. For both analyses between KRAS-mutation(+) colorectal cancer and KRAS-mutation(−) colorectal cancer, and between KRAS-mutation(+) colorectal cancer with IME and the others, the sample size of 121 patients was calculated to be sufficient to provide a survival rate difference of 25% with significance level of 0.05 and power of 80% \(^{29}\). The sample size used for these two survival analyses were 131 and 126, respectively, both of which were more than the required numbers.

**Comparison with classical CIMP classification.** To compare the three colorectal cancer epigenotypes with classical CIMP status, we decided the classical CIMP status of 149 colorectal cancer samples using 11 CIMP-related markers (Fig. 5A; refs. 14, 15). In this classification, 113 of 149 (76%) colorectal cancer samples were grouped into the corresponding group (Fig. 5B). There was no significant difference in KRAS mutation \( (P = 0.086 \) among three CIMP status; \( P = 0.12 \) between CIMP-low and CIMP(−); Fig. 5C, Supplementary Table S3). As for clinical outcome, there was no significant difference in overall survival between CIMP-low and CIMP(−) in MSS colorectal cancer \( (P = 0.32, \text{log-rank test}) \), nor by adding KRAS-mutation status \( (P = 0.45, \text{log-rank test}) \). CIMP determined by previously reported sets of markers \( (\text{set 1 with CACNA1G, IGF2, NUROG1, RUNX3, and SOCS1}; \text{ref. 15; set-2 with hMLH1, p16INK4A, MINT1, MINT2, and MINT31}; \text{refs. 30, 31}) \) were also analyzed (Supplementary Fig. S5). There was no significant difference in KRAS-mutation between CIMP-low and CIMP(−), or in clinical outcome (data not shown).

**Discussion**

In this study, colorectal cancer was epigenotyped using 44 new methylation markers selected through genome-wide methylation and re-expression analyses as well as 16 reported genes/loci. Colorectal cancer samples were clustered into three distinct epigenotypes, HME, IME, and LME. Methylation markers were clustered into group-1 markers showing high methylation in HME and group-2 markers showing high methylation in IME and LME. Although MSS and MSI-H cell lines were used to select new methylation markers, there was neither MSS-specific nor IME-specific methylation marker identified. Two-step classification was therefore required and developed to classify three epigenotypes: the first step to extract HME using group-1 markers, and the second step to classify IME and LME using group-2 markers.

Previous CIMP-related markers were all categorized into group-1 markers except NEUROG1, indicating that CIMP-related markers were suitable to distinguish HME from others. It was reported that BRAF mutation showed nonrandom pattern of CpG island methylation whereas KRAS mutation showed random pattern, by analyzing 16 regions \(^{32}\). In these 16 regions, 13 regions were analyzed in our study, and all the regions except NEUROG1 were identified as group-1 markers. It was considered that although BRAF-mutation(+) colorectal cancer should show nonrandom pattern of methylation of group-1 markers, KRAS-mutation(+) colorectal cancer should show only rare and random methylation of group-1 markers and would show nonrandom pattern of methylation of group-2 markers if analyzed.

Shen et al. reported that colon cancer was classified into three groups using methylation data of 27 regions, which were constructed using four different methylation detection methods: 13 regions by pyrosequencing, 7 by combined bisulfite restriction analysis, 6 by methylated CpG island amplification, and 1 by methylation-specific PCR \(^{17}\). In our study, methylation data were obtained by a single, highly quantitative method, and two-way hierarchical clustering analysis showed clearly not only the existence of three distinct epigenotypes, but also two clusters of methylation markers. Correlation of our epigenotypes with MSI status, and BRAF and KRAS mutation was similar to Shen’s CIMPs, but our study showed no correlation of LME with p53 status. This might be due to the difference between sequencing and immunohistochemistry. Although some reported that positivity by immunohistochemistry correlated with the presence of p53 mutation \(^{18, 33}\), others reported that immunohistochemistry was not a reliable technique for detecting p53 mutation \(^{34}\). The second possibility is that our IME and their CIMP-2 might be different clusters. In fact, they used several MINT loci to detect CIMP-2, by

\(^{10}\) http://minds.jcqhc.or.jp
analyzing methylation using a competitive PCR method, methylated CpG island amplification. MINT loci showed no or very small difference in methylation rate between IME and LME and belonged to group-1 markers in our study, and so were considered to be difficult to classify IME and LME.

In the panel method, panel 1 consisted of three group-1 markers, CACNA1G, LOX, and SLC30A10. CACNA1G possesses MINT31, one of the original CIMP markers by Toyota et al. (14), at 2 kb upstream of its transcription start site, and also was used as a CIMP-related marker by Weisenberger et al. (15). LOX is a tumor suppressor gene inactivated by methylation in mammalian cancer (11, 35). Panel 2 consisted of four group-2 markers (ELMO1, FB2N, THBD, and HAND1) and SLC30A10 again. Hypermethylation of FB2N, THBD, and HAND1 was reported in cancers other than colorectal cancer, such as pancreatic, gastric, and non-small cell lung cancer (35–37), and the antiproliferative effect of THBD was reported in melanoma cells (38). ELMO1 was reported to have a role in promoting cancer cell invasion in glioma cells (39), but its promoter hypermethylation was not previously reported. SLC30A10 belonged to type 2, showing significant difference in methylation rate between HME and IME and between IME and LME, so it was acceptable to use this marker in both panel 1 and panel 2.

It is interesting that BRAF-mutation(+) and KRAS-mutation(+) colorectal cancers clearly showed different epigenotypes, although the reason is unknown. There are some possibilities. First, activation of these oncogenes itself may induce different epigenetic alterations on a genome-wide scale. In fact, silencing of genes occurred in Kras-transformed NIH3T3 but not in untransformed NIH3T3, and essential effectors for the epigenetic gene silencing were reported (40). In addition, oncogene activation in normal cells is known to induce oncogene-induced senescence forming senescence-associated heterochromatin foci (41), where regions with repression mark, such as histone H3 lysine K9 methylation, were found to gather (42). Dynamic epigenetic changes including histone modification and DNA methylation might be induced in this process. Second, oncogene activation itself may not cause DNA methylation, but gene inactivation by methylation might be required to escape oncogene-induced senescence to be cancer, as disruption of p16 and/or p53 leads to escape from oncogene-induced senescence (43). IGFBP7, whose methylation was reported to be essential to escape from oncogene-induced senescence in BRAF-mutated melanoma cells (44), showed significantly higher methylation rate in HME colorectal cancer (Fig. 3). IGFBP3, reported to be induced in replicatively senescent human umbilical vein endothelial cells (45), showed high methylation rate in IME colorectal cancer (Fig. 3). Although colon epithelium is different from melanocytes and endothelial cells, these data and reports suggested that Raf and Ras activation may perhaps require different gene inactivation to be cancer, which might result in different epigenotypes, i.e., HME and IME.

Whereas previous meta-analysis showed that MSS colorectal cancer had worse prognosis than MSI-H colorectal cancer (46), which was correlated with HME and BRAF-mutation(+), our interest was whether epigenotypes and KRAS-mutation status could predict prognosis of MSS colorectal cancer. Correlation between prognosis and KRAS-mutation itself is controversial; KRAS-mutation was indicative of worse prognosis in 13 reports, but not in 10 reports (47). In our study, there was no significant difference in survival rate between KRAS-mutation(+) and KRAS-mutation(−) in MSS colorectal cancer (Fig. 4B). Whereas KRAS-mutation highly correlated to IME, KRAS-mutation(+) colorectal cancer with IME significantly correlated with poorer survival (Fig. 4C and D). Although the prognostic significance of IME needs to be further investigated by additional studies, including a larger cohort of patients, this poor prognosis was observed under therapy without anti–epidermal growth factor receptor (EGFR) or anti–vascular endothelial growth factor antibody. The anti-EGFR antibody cetuximab (Erbitux) was recently reported to be effective on KRAS-mutation(−) colorectal cancer, but not on KRAS-mutation(+) colorectal cancer (48), suggesting that chemotherapy different from anti-EGFR antibody should be developed especially for KRAS-mutation(+) IME colorectal cancer. Identification of the existence of IME and its strong correlation to KRAS-mutation(+) is considered to be useful in understanding the molecular genesis of colorectal cancer in this unfavorable group, and it might be possible to develop a novel target therapy based on epigenetic alteration and/or oncogene mutation.

In conclusion, colorectal cancer was clustered into three major epigenotypes with different genetic characteristics, suggesting different molecular genesis of colorectal cancer. Methylation markers were clustered into two major groups, and a two-step panel method to predict colorectal cancer epigenotypes was constructed. IME with KRAS mutation showed significantly worse prognosis in MSS colorectal cancer.

Disclosure of Potential Conflicts of Interest

The authors have applied for a patent for a methylation marker used in colorectal cancer detection.

Acknowledgments

We thank Dr. Shumpei Ishikawa and Dr. Keisuke Matusaka for pathological diagnosis, and Kaoru Nakano, Hiroko Meguro and Kaori Shiina for technical assistance.

Grant Support

Grants-in-Aid for Scientific Research (B) 19390345 (A. Kaneda) and (S) 20101006 (H. Aburatani), by a grant for Translational Systems Biology and Medicine Initiative (TSBMI) from the Ministry of Education, Culture, Sports,

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 7/29/09; revised 9/18/09; accepted 9/27/09; published OnlineFirst 12/22/09.

References


Clinical Cancer Research

Three DNA Methylation Epigenotypes in Human Colorectal Cancer

Koichi Yagi, Kiwamu Akagi, Hiroshi Hayashi, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-2006

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2010/01/07/1078-0432.CCR-09-2006.DC1

Cited articles
This article cites 28 articles, 5 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/16/1/21.full.html#ref-list-1

Citing articles
This article has been cited by 30 HighWire-hosted articles. Access the articles at:
/content/16/1/21.full.html#related-urls

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

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

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