modification, play critical roles in the development of CRC; in addition, an increasing number of genes involved in cell cycle control, DNA repair, tumor invasiveness, and the response to growth factors have been identified as targets of hypermethylation in CRC.[3–6] These epigenetic alterations are thought to be the main driving force in a subset of CRCs exhibiting concurrent hypermethylation of multiple loci, which is termed the CpG island methylator phenotype (CIMP).[7] CIMP-positive CRCs show characteristic clinicopathological and molecular features, including proximal tumor location, female sex, older age, high tumor grade, wild-type *TP53*, frequent *BRAF* and *KRAS* mutations, and MSI. In addition, several studies support the hypothesis that CRCs can be categorized into three subclasses based on aberrant CpG island methylation: CIMP-high (CIMP-H; also known as CIMP1), CIMP-low (CIMP-L; also known as CIMP2), and CIMP-negative (CIMP-N). CIMP-H CRCs are significantly associated with a *BRAF* mutation, *MLH1* methylation, and subsequent MSI.[7–11] A link between CIMP-L CRCs and *KRAS* mutations was first reported by Ogino et al,[12] and it was subsequently confirmed by other groups,[10,13] but much remains unknown about their respective molecular and clinicopathological features.

Reducing the incidence of cancers, such as CRC, will require a better understanding of the mechanisms underlying carcinogenesis and the molecular alterations occurring in premalignant lesions. For example, although experimental evidence has confirmed the presence of CIMP in CRCs, the role of CIMP in the progression of precancerous lesions toward cancer is not yet fully understood. In recent years, sessile serrated adenomas (SSAs) have been the origin of MSI-positive/CIMP-H cancers, which account for approximately 10% to 15% of sporadic CRCs.[14,15] In addition, studies have also shown that CIMP is frequently observed among MSI-negative CRCs.[10,16] From these and other studies of many tumors, a model was suggested in which CIN and epigenetic instability (CIMP) represent the two major pathways of CRC development, with up to 50% of CRCs being characterized as CIMP.[16,17] This means that the origin of a large fraction of CIMP cancers remains unclear.

High-resolution magnifying colonoscopy is a powerful diagnostic tool for detecting premalignant lesions. According to Kudo's classification, the pit patterns of nonneoplastic lesions are classified as type I (normal colon) or type II [hyperplastic polyp (HP)], whereas the pit patterns of neoplastic lesions are classified as types III, IV, and V.[18,19] Recently, we performed an integrative analysis of the morphological, pathological, and molecular signatures in colorectal precancerous lesions and identified a novel pit pattern (type II, open pits) that was specific to SSAs.[20] Those results depict an important relationship between morphological characteristics and molecular alterations that will significantly improve our ability to detect premalignant lesions. In the present study, our aim was to uncover the molecular evolution of CIMP cancers through an integrative analysis of many precursor and malignant colorectal lesions. Based on their genetic and epigenetic signatures, we propose a model in which CRCs develop via four distinct pathways. We also provide evidence that CIMP and CIN are not completely mutually exclusive, so that chromosomal aberrations may play important roles in a subset of CIMP cancers. These findings will improve our understanding of the pathogenesis of CRCs and could potentially contribute to better clinical management of premalignant lesions.

Materials and Methods

Study Population and Tissue Specimens

Colorectal tumor tissues were collected from Japanese patients who underwent endoscopic or surgical resection of a colorectal tumor at Akita Red Cross Hospital (Akita, Japan). A total of 368 specimens from 192 precursor lesions, 38 noninvasive carcinomas [carcinoma *in situ* (CIS)], 100 CRCs, and 38 samples of adjacent normal tissue were analyzed in this study. Informed consent was obtained from all patients before collection of the specimens. Approval of this study was obtained from the Institutional Review Board of Akita Red Cross Hospital and Sapporo Medical University (Sapporo, Japan). By using the standard phenol/chloroform procedure, genomic DNA was extracted from biopsy specimens obtained before endoscopic or surgical resection. CRC cell lines were maintained and treated with 5-aza-2′-deoxycytidine, as previously described.[21]

Endoscopic and Histological Analysis

High-resolution magnifying endoscopes (CF260AZI; Olympus, Tokyo, Japan) were used for all colonoscopic analyses. Colorectal subsite locations were defined as right side colon proximal to the splenic flexure (cecum, ascending colon, hepatic flexure, and transverse colon), left side colon distal to the splenic flexure (splenic flexure, descending colon, and sigmoid colon), and rectum (rectosigmoid and rectum). All detected colorectal tumors were observed at high magnification after staining with indigo carmine dye and 0.05% crystal violet. Surface microstructures were classified according to Kudo's pit pattern classification system.[18,19] Most often, one biopsy specimen was collected from each lesion for the extraction of genomic DNA. However, when two or more pit patterns were found in a single lesion (eg, adenoma to carcinoma transition), biopsy specimens were obtained for each respective pit pattern (see Supplemental Figure S1 at *http://ajp.amjpathol.org*). Thereafter, the lesions underwent endoscopic mucosal resection, endoscopic submucosal dissection, or surgical resection, after which histological analyses were performed (see Supplemental Figure S1 at *http://ajp.amjpathol.org*). Conventional adenomas, such as tubular adenoma and tubulovillous adenoma, were diagnosed using standard criteria. Serrated lesions, including HP, traditional serrated adenoma (TSA), and SSA, were classified based on criteria previously described by Torlakovic et al.[22] Serrated lesions that did not satisfy the criteria for SSA or TSA were defined as HP. Tumors were classified into three categories: precursor lesions (HP, tubular adenoma, tubulovillous adenoma, TSA, and SSA), CIS, and CRCs.

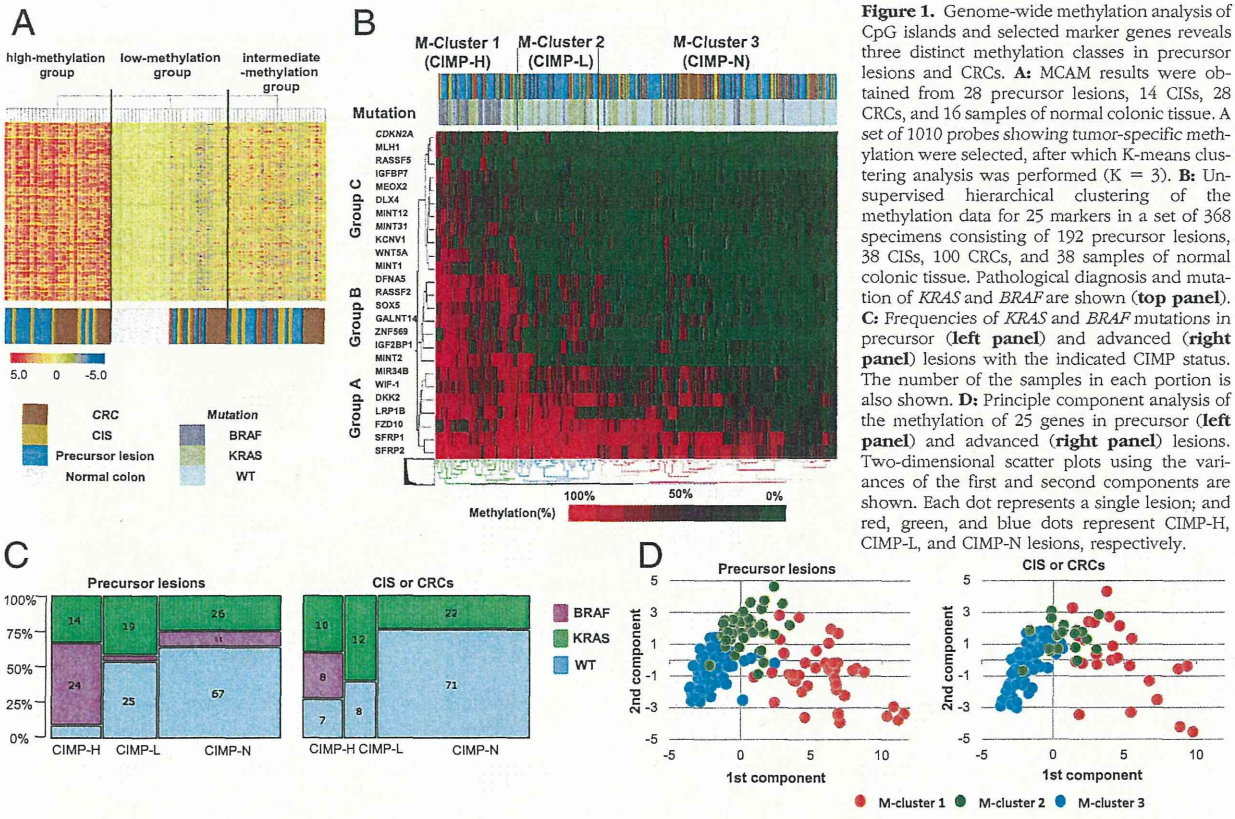

Figure 1. Genome-wide methylation analysis of CpG islands and selected marker genes reveals three distinct methylation classes in precursor lesions and CRCs. **A:** MCAM results were obtained from 28 precursor lesions, 14 CISs, 28 CRCs, and 16 samples of normal colonic tissue. A set of 1010 probes showing tumor-specific methylation were selected, after which K-means clustering analysis was performed (K = 3). **B:** Unsupervised hierarchical clustering of the methylation data for 25 markers in a set of 368 specimens consisting of 192 precursor lesions, 38 CISs, 100 CRCs, and 38 samples of normal colonic tissue. Pathological diagnosis and mutation of *KRAS* and *BRAF* are shown (**top panel**). **C:** Frequencies of *KRAS* and *BRAF* mutations in precursor (**left panel**) and advanced (**right panel**) lesions with the indicated CIMP status. The number of the samples in each portion is also shown. **D:** Principle component analysis of the methylation of 25 genes in precursor (**left panel**) and advanced (**right panel**) lesions. Two-dimensional scatter plots using the variances of the first and second components are shown. Each dot represents a single lesion; and red, green, and blue dots represent CIMP-H, CIMP-L, and CIMP-N lesions, respectively.

MCAM Data

Methylated CpG island amplification microarray (MCAM) analysis was performed as previously described.[23] A BioPrime Plus Array CGH Genomic Labeling System (Life Technologies, Carlsbad, CA) was used to label MCA amplicons from tumor samples with Alexa Fluor 647 (Life Technologies, Carlsbad, CA), and those from a pooled mixture of normal colonic tissue were labeled with Alexa Fluor 555 (Life Technologies). Labeled MCA amplicons were then hybridized to a custom human CpG island microarray (G4497A; Agilent Technologies, Santa Clara, CA), which included 15,134 probes covering 6157 unique genes. After washing, the array was scanned using an Agilent DNA Microarray Scanner (Agilent Technologies), and the data were processed using Feature Extraction software version 10.7 (Agilent Technologies), and analyzed using GeneSpring GX version 11 (Agilent Technologies). Unsupervised hierarchical clustering and statistical analyses were then performed.

Methylation Analysis by Bisulfite Pyrosequencing

Bisulfite pyrosequencing was performed as previously described.[24] Briefly, genomic DNA (1 μg) was modified with sodium bisulfite using an EpiTect Bisulfite Kit (Qiagen, Hilden, Germany). Pyrosequencing was then performed using a PSQ96 system with a PyroGold reagent Kit (Qiagen), after which the results were analyzed using Q-CpG software version 1.0.9 (Qiagen). Unsupervised hierarchical clustering, principal component analysis, and correspondence analysis of validation data were performed using JMP version 8 (SAS Institute Inc., Cary, NC). For the statistical analysis, quantitative methylation levels of each gene were converted to *z* scores, which were defined as follows: (Methylation Level in Each Sample - Mean Methylation Level in All Samples)/(SD of Methylation Levels in All Samples). Primer sequences are listed in Table 1.

Mutation and MSI Analysis

Mutations in codon 600 of *BRAF* and codons 12 and 13 of *KRAS* were examined by pyrosequencing using *BRAF* and *KRAS* pyrokits (Qiagen), respectively, according to the manufacturer's instructions. Mutation of *PIK3CA*, *AKT2*, and *PDK1* was analyzed by direct sequencing, as previously described.[25] Mutation of *TP53* was initially detected by PCR–single-stranded conformational polymorphism analysis, followed by direct sequencing, as previously described.[26] Mutation of *AKT1* was analyzed by pyrosequencing using the primer sequences listed in Table 1. MSI was assessed as previously described,[27] using primers proposed by the National Cancer Institute Workshop on Microsatellite Instability (*BAT25*, *BAT26*, *D5S346*, *D2S123*, and *D17S250*).[28] MSI was defined by the presence, in the tumor sample, of bands that were abnormal in size, compared with a corresponding normal

Table 2. Clinicopathological Features of the Colorectal Tumors Used in this Study

Feature	Total (N = 192)	CIMP-H (n = 42)	CIMP-L (n = 46)	CIMP-N (n = 104)	P value
Precursor Lesion					
Age (years)	69.61 ± 9.68	72.14 ± 8.01	72.43 ± 9.82	67.34 ± 9.73	<0.05
Sex					
F	73 (38.02)	21 (50)	26 (56.52)	26 (25)	<0.001
M	119 (61.98)	21 (50)	20 (43.48)	78 (75)	
Location					
Right	92 (47.92)	32 (76.19)	27 (58.7)	33 (31.73)	<0.001
Left	40 (20.83)	5 (11.9)	7 (15.22)	28 (26.92)	
Rectum	60 (31.25)	5 (11.9)	12 (26.09)	43 (41.35)	
Size (mm)	12.98 ± 9.61	13.73 ± 8.44	16.67 ± 12.55	8.38 ± 6.48	<0.001
Morphological characteristics					
Protruding type	82 (42.71)	17 (40.48)	21 (45.65)	44 (42.31)	
Flat type	110 (57.29)	25 (59.52)	25 (54.35)	60 (57.69)	
Depressed type	0	0	0	0	
Histological features					
HP	28 (14.58)	2 (4.76)	5 (10.87)	21 (20.19)	<0.001
SSA	29 (15.1)	26 (61.90)	1 (2.17)	2 (1.92)	
TSA	25 (13.02)	6 (14.29)	5 (10.87)	14 (13.46)	
Tubular adenoma	53 (17.6)	2 (4.76)	7 (15.22)	44 (42.31)	
Tubulovillous adenoma	57 (29.69)	6 (14.29)	28 (60.87)	23 (22.12)	
CIS					
	(N = 38)	(n = 8)	(n = 5)	(n = 25)	
Age (years)	66.71 ± 13.77	75.75 ± 7.78	73.6 ± 5.5	62.44 ± 14.61	<0.05
Sex					
F	13 (34.21)	3 (37.5)	2 (40)	8 (32)	
M	25 (65.79)	5 (62.5)	3 (60)	17 (68)	
Location					
Right	18 (47.37)	6 (75)	4 (80)	8 (32)	<0.05
Left	9 (23.68)	0	1 (20)	8 (32)	
Rectum	11 (28.95)	2 (25)	0	9 (36)	
Size (mm)	17.66 ± 9.75	20.63 ± 12.96	19.6 ± 10.21	16.32 ± 8.64	
Morphological features					
Protruding type	20 (52.63)	6 (75)	1 (20)	13 (52)	
Flat type	16 (42.11)	2 (25)	4 (80)	10 (40)	
Depressed type	2 (5.26)	0	0	2 (8)	
CRC					
	(N = 100)	(n = 17)	(n = 15)	(n = 68)	
Age (years)	67.81 ± 12.48	71 ± 9.54	72.13 ± 12.78	66.06 ± 12.82	
Sex					
F	43 (43)	14 (82.35)	8 (53.33)	21 (30.88)	<0.01
M	57 (57)	3 (17.65)	7 (46.67)	47 (69.12)	
Location					
Right	48 (48)	16 (94.12)	6 (40)	26 (38.24)	<0.001
Left	22 (22)	1 (5.88)	3 (20)	18 (26.47)	
Rectum	30 (30)	0 (0)	6 (40)	24 (35.29)	
Stage (UICC)					
I	38 (38)	8 (47.06)	3 (20)	27 (39.71)	
II	31 (31)	4 (23.53)	5 (33.33)	22 (32.35)	
III	23 (23)	5 (29.41)	7 (46.67)	11 (16.18)	
IV	8 (8)	0	0	8 (11.76)	

F, female; M, male; UICC, Union for International Cancer Control.

sample. A tumor sample was defined as MSI positive when two or more markers showed instability.

Array-Based Comparative Genomic Hybridization

Array-based comparative genome hybridization (array CGH) analysis was performed as previously described.[35] Briefly, 500 ng of genomic DNA from colorectal tumor specimens and sex-matched reference DNA from noncancerous colonic mucosa were labeled with Cy5 and Cy3, respectively, using a Genomic DNA Enzymatic Labeling Kit (Agilent Technologies), and were then hybridized to Human Genome CGH Microarray Kit 180A (G4449A; Agilent Technologies). DNA copy number aberrations (CNAs) were identified using the ADM-2 algorithm included in the Agilent Genomic Workbench software version 5 (Agilent Technologies). A copy number loss was defined as a log 2 ratio of <-0.5, and a copy number gain was defined as a log 2 ratio of >0.5. Unsupervised hierarchical clustering of CNAs was performed using JMP, version 8. The microarray data in this study

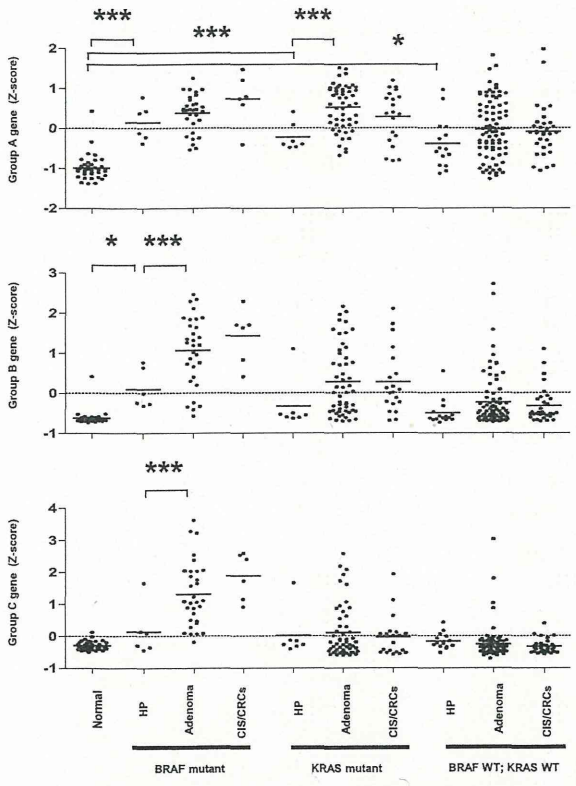

Figure 2. Association between methylation and *BRAF*/*KRAS* mutations. The levels of methylation of groups A to C genes in normal colon, precursor, and malignant lesions are shown as z scores. Colorectal lesions are divided into three groups, according to their *BRAF* or *KRAS* mutation status. Among the *BRAF* mutants, levels of methylation of group B and C genes are significantly higher in adenomas than HPs but are not further up-regulated in advanced lesions. $^{*}P < 0.05$, $^{***}P < 0.001$.

have been submitted to the Gene Expression Omnibus, and the accession number is GSE35534.

Statistical Analysis

To compare differences in continuous variables between groups, *t*-tests or an analysis of variance with a post hoc Tukey's test was performed. A Fisher's exact test or a χ^2 test was used for analysis of categorical data. $P < 0.05$ (two sided) was considered statistically significant. Statistical analyses were performed using JMP, version 8, and SPSS statistics 18 (IBM Corporation, Somers, NY).

Results

Three Methylation Subclasses in Precancerous and Malignant Colorectal Tumors

To clarify the epigenetic changes occurring early during colorectal tumorigenesis, we first performed global methylation analysis using MCAM in a series of normal colonic tissues (n = 16), precursor lesions (HP, n = 3; SSA, n = 5; TSA, n = 3; tubular adenoma, n = 6; tubulovillous adenoma, n = 10), CISs (n = 14), and CRCs (n = 28). Hierarchical clustering analysis using the MCAM results identified 1010 probe sets that detected frequent hypermethylation in tumor tissues (see Supplemental Figure S2 at *http://ajp.amjpathol.org*). Subsequent K-means clustering analysis using these probe sets revealed that the samples could be clearly categorized into three subclasses based on the level of their methylation (Figure 1A); subclasses with high and intermediate methylation were presumed to reflect CIMP-H and CIMP-L tumors, respectively. Among the 1010 probe sets, 538 unique genes were in the high-methylation group, whereas 259 genes were in the intermediate-methylation group (see Supplemental Figure S2C and Supplemental Tables S1 and S2 at *http://ajp.amjpathol.org*). A subset of the precursor lesions could also be categorized into these subclasses, indicating that CIMP-like methylation patterns were already established early during carcinogenesis.

To further characterize the genes that acquired methylation progressively during CIMP pathway tumorigenesis, we next performed MCAM analysis with a series of precursor lesions in which CIMP-N flat components were present, along with CIMP-positive protruding components within the same lesions (see Supplemental Figure S2D at *http://ajp.amjpathol.org*). Because both components were presumed to arise from the same origin, these lesions could represent an ideal model for analyzing the molecular progression of CIMP cancers. CIMP status was defined using classic CIMP markers (*MINT1*, *MINT2*, *MINT12*, *MINT31*, and *CDKN2A*), and tumors with methylation of three or more markers were defined as CIMP. When we analyzed three pairs of precursor lesions using MCAM, we identified 36 unique genes that were differentially methylated between CIMP-positive and CIMP-N precursor lesions (see Supplemental Table S3 at *http://ajp.amjpathol.org*). These genes were potentially the earliest targets of aberrant methylation during CIMP pathway tumorigenesis, and most of them overlapped with the genes identified in the initial MCAM analysis (see Supplemental Figure S2E at *http://ajp.amjpathol.org*).

Methylation Profiling Identified CIMP in Precancerous Lesions

Based on the results previously summarized, we selected a series of marker genes to characterize the methylation profile of precursor and malignant lesions. We initially selected 14 genes (*LRP1B*, *CDKN2A*, *WNT5A*, *MEOX2*, *ZNF569*, *GALNT14*, *SOX5*, *DFNA5*, *DLX4*, *SFRP2*, *WIF1*, *FZD10*, *KCNV1*, and *IGF2BP1*) identified in the MCAM analysis (see Supplemental Figure S2 at *http://ajp.amjpathol.org*). Among them, *IGF2BP1*, *KCNV1*, *DLX4*, *GALNT14*, and *ZNF569* were not previously reported to be methylated in CRCs, but RT-PCR analysis using multiple CRC cell lines confirmed that they were frequent targets of epigenetic silencing in CRC (see Supplemental Figure S3 at *http://ajp.amjpathol.org*). In addition, we selected 11 well-characterized markers (*MLH1*, *SFRP1*, *IGFBP7*, *DKK2*, *MIR34B*, *MINT1*, *MINT2*, *MINT12*, *MINT31*, *RASSF2*, and *RASSF5*) used for methylation analysis.[4,7,21,24,30–32]

Table 3. Clinicopathological Features of the Colorectal Tumors Analyzed Using Array CGH

Feature	C-cluster 1 (n = 49)	C-cluster 2 (n = 13)	C-cluster 3 (n = 22)	P value
Age (years)	69.51 ± 11.01	69.54 ± 8.97	67 ± 10.55	
Sex				
F	14 (28.57)	9 (69.23)	4 (18.18)	
M	35 (71.43)	4 (30.77)	18 (81.82)	
Location				
Right	27 (55.1)	9 (69.23)	6 (27.27)	
Left	8 (16.33)	4 (30.77)	7 (31.82)	
Rectum	14 (28.57)	0	9 (40.91)	
Precursor lesions				
HP	1 (2.04)	0	0	
SSA	2 (4.08)	0	0	
TSA	7 (14.29)	1 (7.69)	0	
Tubular adenoma	3 (6.12)	1 (7.69)	0	
Tubulovillous adenoma	1 (2.04)	0	1 (4.55)	
CIS	0	0	2 (9.09)	
CRCs	9 (18.37)	2 (15.38)	8 (36.36)	
SSA + CIS				
SSA portion	2 (4.08)	0	0	
CIS portion	1 (2.04)	1 (7.69)	0	
Tubulovillous adenoma + CIS				
Tubulovillous adenoma portion	12 (24.49)	1 (7.69)	0	
CIS portion	4 (8.16)	6 (46.15)	3 (13.64)	
Tubular adenoma + CIS				
Tubular adenoma portion	5 (10.2)	0	3 (13.64)	
CIS portion	2 (4.08)	1 (7.69)	5 (22.73)	
KRAS				
Mut	22 (45)	11 (85)	10 (45)	
Wt	27 (55)	2 (15)	12 (55)	
BRAF				
Mut	10 (20)	1 (8)	0	
Wt	39 (80)	12 (92)	22 (100)	
CIMP				
High	17 (35)	4 (31)	0	<0.01
Low	11 (22)	4 (31)	1 (5)	
Negative	21 (43)	5 (38)	21 (95)	

F, female; M, male; Mut, mutated; Wt, wild type.

We performed quantitative bisulfite pyrosequencing of the 25 markers in a total of 330 specimens consisting of 192 precursor lesions, 38 CISs, and 100 CRCs (Table 2). Consistent with the MCAM results, unsupervised hierarchical clustering using the pyrosequencing data revealed that, in addition to malignant lesions (CISs and CRCs), precursor lesions could also be divided into three relative subclasses (M-clusters 1 to 3; Figure 1B). The *BRAF* mutation was significantly enriched in M-cluster 1 tumors, in which most of the genes were highly methylated, suggesting that this subclass corresponded to CIMP-H (Figure 1B). M-cluster 2 included tumors with intermediate levels of methylation and a prevalent *KRAS* mutation, which corresponded to CIMP-L, whereas M-cluster 3 tumors exhibited the lowest methylation levels, which corresponded to CIMP-N. Among the precursor lesions, 37 (19.3%) exhibited a *BRAF* mutation, and most of those [24 (64.9%) of 37] were categorized as CIMP-H (Figure 1C). In CISs and CRCs, the *KRAS* mutation was most enriched in the CIMP-L group, but this tendency was less apparent among the precursor lesions (Figure 1C).

We next used principle component analysis to further evaluate our bisulfite pyrosequencing results and found that the first and second components accounted for 53.1% of the total variance (Figure 1D; see also Supplemental Table S4 at *http://ajp.amjpathol.org*). Two-dimensional plotting then showed that the characteristic pattern for each M-cluster was shared by the precursor and malignant lesions (Figure 1D; see also Supplemental Figure S4A at *http://ajp.amjpathol.org*).

Our hierarchical clustering analysis also showed that the marker genes could be categorized into three subgroups (Figure 1B). Group A genes, which were methylated in most of the samples, corresponded to the type A (age-related) genes originally proposed by Toyota et al.[7] Group B and C genes appeared to correspond to type C (cancer-specific) genes, whereas group C genes in this study were more strongly associated with CIMP-H. Interestingly, among *BRAF*-mutant precursor lesions, adenomas showed much higher levels of methylation of groups B and C genes than HPs (Figure 2). Similar results were obtained with *KRAS*-mutant precursors, although the difference was not statistically significant (Figure 2). By contrast, methylation was minimally up-regulated in adenomas in which both *BRAF* and *KRAS* were wild type, suggesting that accumulation of aberrant methylation, in concert with a *BRAF* or a *KRAS* mutation, may promote the progression from benign tumors to precancerous lesions.

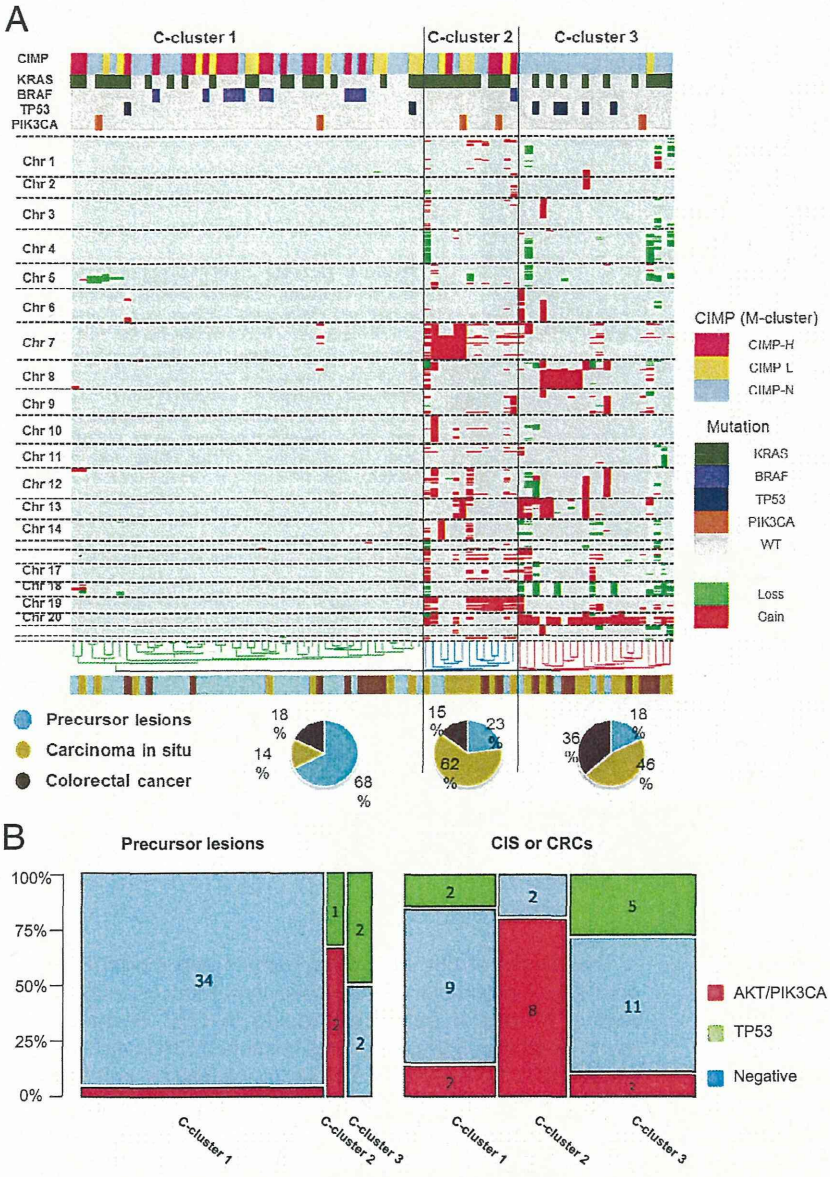

Figure 3. Distinct subclasses of precursor and malignant colorectal lesions are defined based on their CNAs. **A:** Unsupervised hierarchical clustering analysis using array CGH data from 40 precursor lesions, 25 CISs, and 19 CRCs. Lesions could be categorized into three subclasses (C-clusters 1 to 3). CIMP status and gene mutations are indicated (**top panel**), as are chromosome (Chr) numbers (**left panel**). Ratios of precursor lesions, CISs, and CRCs in each C-cluster are shown (**bottom panel**). **B:** Ratios of genetic defects in *AKT/PIK3CA* pathway genes and *TP53* mutations in precursor (**left panel**) and advanced (**right panel**) lesions with the indicated CNA status.

Clinicopathological Features of CIMP-Positive Precursor Lesions

The associations between the clinicopathological characteristics and the CIMP status of the precursor lesions, CISs, and CRCs are summarized in Table 2. Age, sex, and tumor location were matched among the three groups. As with CRCs, CIMP-positive precursor lesions were more prevalent than CIMP-N precursor lesions among female and older patients. Interestingly, CIMP-L was associated with larger diameters among precursor lesions, although this tendency was not apparent among malignant lesions. Most of the CIMP-H precursor lesions were SSAs, whereas most of the CIMP-L precursor lesions were tubulovillous adenomas (Table 2; see also Supplemental Figure S4B at *http://ajp.amjpathol.org*). Levels of groups B and C gene methylation were higher in SSAs than in the other precancerous lesions, whereas methylation of group A genes was higher in all precursor types than in normal colonic tissue (see Supplemental Figure S4C at *http://ajp.amjpathol.org*).

CNAs Are Late Events in Colorectal Tumorigenesis

Several studies have shown an inverse relationship between CIMP and CIN in CRC.[17,33,34] For that reason, we next used array CGH to analyze CNAs in 40 precursor lesions, 25 CISs, and 19 CRCs (Table 3). Unsupervised hierarchical clustering analysis using the array CGH data revealed that the tumors could be categorized into three subclasses, according to their CNA status (C-clusters 1 to 3): C-cluster 1 was enriched in tumors with the fewest

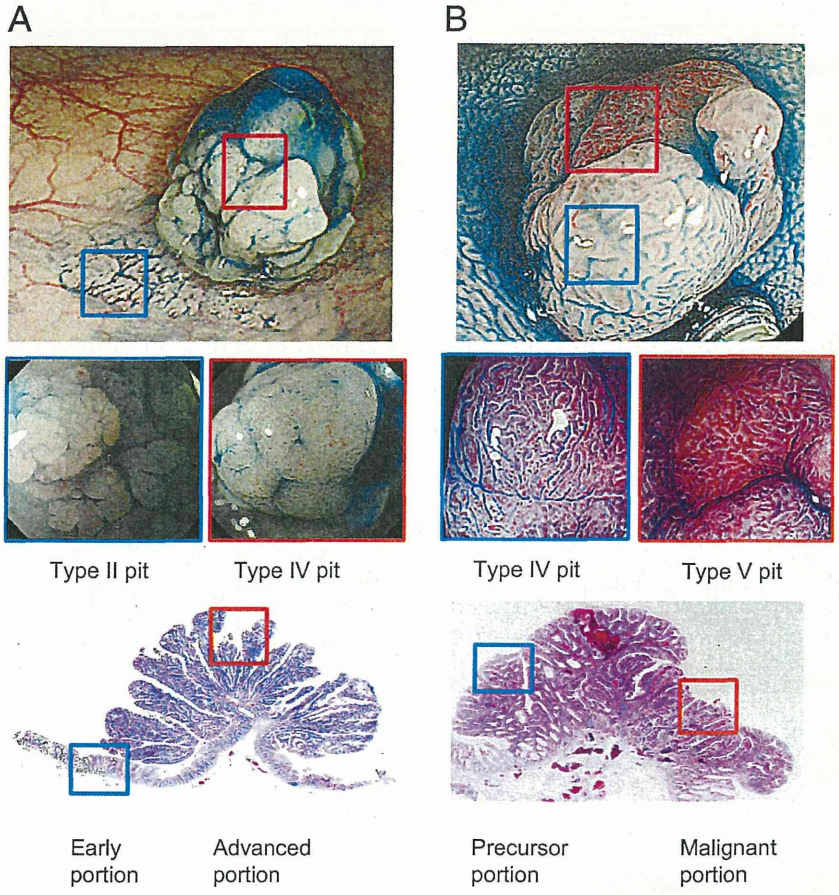

Figure 4. Endoscopic and histological findings in a set of mixed colorectal lesions. **A:** Endoscopic and histological findings from a representative precursor lesion in which a flat portion with an early pit pattern (type II, early portion) is present, along with a protruding portion with advanced pits (type IV, advanced portion). Both components are histologically premalignant (HP and TSA). Biopsy specimens were obtained from the respective portions (**blue** and **red boxes**), after which the molecular profiles are analyzed. **B:** Endoscopic and histological findings from a representative lesion in which a precursor portion (type IV, pit pattern) is present, along with a malignant portion (type V, pit pattern). Biopsy specimens were obtained from the respective portions (**blue** and **red boxes**), after which the molecular profiles are analyzed.

CNAs, whereas C-cluster 2 tumors were characterized by frequent copy number gains on chromosomes 7 and 19; both gains and losses were prevalent among tumors in C-cluster 3 (Figure 3A; see also Supplemental Figure S5A at *http://ajp.amjpathol.org*). Much of the precursor lesions [33 (82.5%) of 40] were enriched in C-cluster 1, whereas most of the malignant lesions (CISs and CRCs) were enriched in C-cluster 2 or 3, suggesting that CNAs occurred late during colorectal tumorigenesis (Figure 3A). Most of the CIMP-positive precursors and malignant lesions were categorized as C-cluster 1 or 2 (Figure 3A; see also Supplemental Figure S6A at *http://ajp.amjpathol.org*), and the methylation levels of the groups B and C genes were similar between C-cluster 1 and 2 tumors (see Supplemental Figure S6B at *http://ajp.amjpathol.org*). These observations suggested that CIMP and CNAs were inversely correlated in many colorectal tumors and that a subset of the CIMP-positive tumors (C-cluster 2) exhibited frequent copy number gains, particularly on chromosomes 7 and 19. As was previously seen, we found that most *BRAF*-mutant precursors and malignant lesions were enriched in C-cluster 1 (Figure 3A). By contrast, although most of the *KRAS*-mutant precursors were enriched in C-cluster 1, *KRAS*-mutant malignant lesions were equally distributed among all three C-clusters (Figure 3A; see also Supplemental Figure S6C at *http://ajp.amjpathol.org*).

The results of our integrative genetic and epigenetic analysis of precursor lesions were indicative of several distinct molecular pathways leading to CRC development. Notably, CIMP-positive/*BRAF*-mutant CRCs did not exhibit more CNAs than did pre-invasive lesions with the same *BRAF* mutations and CIMP-positive methylation profile, suggesting that such pre-invasive lesions may progress to CRC without additional CNAs. By contrast, CIMP-positive/*KRAS*-mutant precursors appeared to develop via CNA-independent and CNA-dependent pathways. The CNA-dependent pathway was characterized by frequent amplification of *BRAF* and *EZH2* on chromosome 7q and amplification of *AKT2*/*PAK4* and *DNMT1* on chromosome 19 (see Supplemental Figure S5B at *http://ajp.amjpathol.org*). More important, we found that most C-cluster 2 tumors exhibited genetic defects (mutations and/or CNAs) in genes whose products were implicated in the AKT/PIK3CA pathway, including *AKT1*, *AKT2*/*PAK4*, *PDK1*, and *PIK3CA* (Figure 3B; see also Supplemental Figure S5B at *http://ajp.amjpathol.org*).

Dynamics of the Molecular Signatures during the Progression of Colorectal Tumorigenesis

Our results suggested that acquisition of CNAs was essential for *BRAF* wild-type precursors to progress to more

Table 4. Histological and Molecular Signatures in a Set of Mixed Colorectal Lesions

Precursor Lesion Plus Precursor Lesion									
Type II or III pit patterns (early potion)					Type IV pit pattern (advanced potion)				
Pathological findings	Mutation	CIMP (M-cluster)	C-cluster	MSI	Pathological findings	Mutation	CIMP (M-cluster)	C-cluster	MSI
SSA	*BRAF*	CIMP-H	ND	Negative	Adenoma	*BRAF*	CIMP-H	ND	Negative
SSA	*BRAF*	CIMP-H	ND	Negative	Adenoma	*BRAF*	CIMP-H	ND	Negative
SSA	*KRAS*	CIMP-H	ND	Negative	SSA	*KRAS*	CIMP-H	ND	Negative
SSA	*KRAS*	CIMP-H	ND	Negative	Adenoma	*KRAS*	CIMP-H	ND	Negative
SSA	*BRAF*	CIMP-H	1	Negative	Adenoma	*BRAF*	CIMP-H	1	Negative
SSA	*KRAS*	CIMP-L	ND	Negative	SSA	*KRAS*	CIMP-H	ND	Negative
HP	*BRAF*	CIMP-H	ND	Negative	SSA	*BRAF*	CIMP-H	ND	Negative
HP	*BRAF*	CIMP-H	1	Negative	SSA	*BRAF*	CIMP-H	1	Negative
TSA	*KRAS*	CIMP-L	1	Negative	TSA	*KRAS*	CIMP-H	2	Negative
TSA	*KRAS*	CIMP-L	1	Negative	TSA	*KRAS*	CIMP-L	1	Negative
TSA	*KRAS*	CIMP-N	ND	Negative	TSA	*KRAS*	CIMP-L	ND	Negative
HP	*BRAF*	CIMP-N	ND	Negative	TSA	*BRAF*	CIMP-H	ND	Negative
HP	*BRAF*	CIMP-N	1	Negative	TSA	*BRAF*	CIMP-H	1	Negative
HP	*BRAF*	CIMP-N	ND	Negative	TSA	*BRAF*	CIMP-H	ND	Negative
HP	WT	CIMP-N	ND	Negative	TSA	WT	CIMP-L	ND	Negative
HP	*BRAF*	CIMP-N	ND	Negative	TSA	*BRAF*	CIMP-L	ND	Negative
Tubular adenoma	*KRAS*	CIMP-N	1	Negative	Tubulovillous adenoma	*KRAS*	CIMP-H	1	Negative
Tubulovillous adenoma	WT	CIMP-L	ND	Negative	Tubulovillous adenoma	WT	CIMP-L	ND	Negative
Tubulovillous adenoma	WT	CIMP-L	ND	Negative	Tubulovillous adenoma	WT	CIMP-L	ND	Negative
Tubular adenoma	*KRAS*	CIMP-N	2	Negative	Tubulovillous adenoma	*KRAS*	CIMP-N	3	Negative
Tubular adenoma	WT	CIMP-N	1	Negative	Tubular adenoma	*KRAS*	CIMP-N	1	Negative
Tubular adenoma	WT	CIMP-N	1	Negative	Tubular adenoma	*KRAS*	CIMP-N	1	Negative

Precursor Lesion Plus CIS or CRC									
Type II, III, or IV pit patterns (precursor potion)					Type V pit pattern (malignant potion)				
Pathological findings	Mutation	CIMP (M-cluster)	C-cluster	MSI	Pathological findings	Mutation	CIMP (M-cluster)	C-cluster	MSI
SSA	*BRAF*	CIMP-H	1	Negative	CIS	*BRAF*	CIMP-H	1	Positive
SSA	*BRAF*	CIMP-H	ND	Negative	CIS	*BRAF*	CIMP-H	ND	Positive
SSA	*BRAF*	CIMP-H	1	Negative	CIS	*BRAF*	CIMP-H	2	Positive
Tubular adenoma	*KRAS*	CIMP-H	1	Negative	CIS	*KRAS*	CIMP-H	1	Negative
Tubulovillous adenoma	*KRAS*	CIMP-H	1	Negative	CIS	*KRAS*	CIMP-H	1	Negative
Tubulovillous adenoma	*KRAS*	CIMP-L	ND	Negative	CIS	*BRAF*, *TP53*	CIMP-H	ND	Positive
Tubulovillous adenoma	*KRAS*	CIMP-L	1	Negative	CIS	*KRAS*	CIMP-H	2	Negative
Tubulovillous adenoma	*KRAS*	CIMP-L	1	Negative	CIS	*KRAS*	CIMP-L	2	Negative
Tubulovillous adenoma	WT	CIMP-N	1	Negative	CIS	*KRAS*	CIMP-L	2	Negative
Tubulovillous adenoma	*KRAS*	CIMP-N	2	Negative	CIS	*KRAS*	CIMP-L	2	Negative
Tubular adenoma	*KRAS*	CIMP-N	1	Negative	CIS	*KRAS*, *PIK3CA*	CIMP-L	2	Negative
Tubulovillous adenoma	*KRAS*	CIMP-N	1	Negative	CIS	*KRAS*	CIMP-N	2	Negative
Tubulovillous adenoma	*PIK3CA*	CIMP-N	1	Negative	CRC	*PIK3CA*	CIMP-N	2	Negative
Tubulovillous adenoma	WT	CIMP-N	1	Negative	CIS	*KRAS*	CIMP-N	1	Negative
Tubulovillous adenoma	WT	CIMP-L	ND	Negative	CIS	*KRAS*	CIMP-L	ND	Negative
Tubular adenoma	WT	CIMP-N	ND	Negative	CIS	WT	CIMP-L	ND	Negative
Tubulovillous adenoma	*KRAS*	CIMP-L	1	Negative	CIS	*KRAS*, *TP53*	CIMP-L	1	Negative
Tubulovillous adenoma	*KRAS*	CIMP-N	1	Negative	CRC	*KRAS*	CIMP-N	3	Negative
Tubulovillous adenoma	WT	CIMP-N	1	Negative	CIS	WT	CIMP-N	3	Negative
Tubulovillous adenoma	WT	CIMP-N	1	Negative	CIS	*KRAS*	CIMP-N	3	Negative
Tubular adenoma	WT	CIMP-N	1	Negative	CIS	WT	CIMP-N	3	Negative
Tubular adenoma	*KRAS*	CIMP-L	1	Negative	CRC	*KRAS*, *TP53*	CIMP-L	3	Negative
Tubular adenoma	WT	CIMP-N	3	Negative	CIS	WT	CIMP-N	3	Negative
Tubular adenoma	*TP53*	CIMP-N	3	Negative	CIS	*TP53*	CIMP-N	3	Negative
Tubular adenoma	WT	CIMP-N	3	Negative	CIS	WT	CIMP-N	3	Negative
Tubular adenoma	WT	CIMP-N	1	Negative	CIS	WT	CIMP-N	1	Negative
Tubulovillous adenoma	WT	CIMP-N	1	Negative	CIS	WT	CIMP-N	1	Negative

ND, no data; WT, wild type.

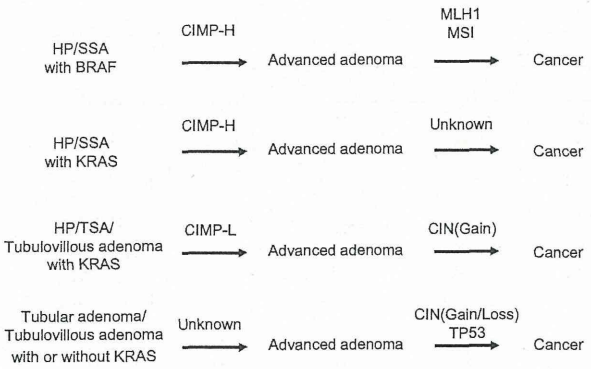

Figure 5. Model for development of CRCs via four distinct molecular pathways.

advanced tumors. To confirm this finding, we analyzed a series of colorectal lesions in which precursor components were present together with more advanced lesions within the same tumors. According to Kudo's classification, the aberrant pit patterns observed using magnifying colonoscopy are hallmarks of malignant tumors, and enabled us to distinguish between the precursor and advanced components (see Supplemental Figure S1 at *http://ajp.amjpathol.org*).[19] We first analyzed the precursor lesions (n = 22), in which portions with early pit patterns (type II or III) were present, along with more advanced pits (type IV), although both components were histologically premalignant (Figure 4A). Progression from precursor lesions with early pits to lesions with advanced pits was associated with the accumulation of DNA methylation, whereas genetic alterations (mutations and CNAs) were rarely acquired (Table 4; see also Supplemental Figures S7 and S8 at *http://ajp.amjpathol.org*). By contrast, progression from precursor (type II, III, or IV pit) to malignant (type V pit) lesions (n = 27) was accompanied by the occurrence of a wide variety of genetic changes, whereas methylation levels remained largely unchanged (Figure 4B and Table 4; see also Supplemental Figures S7 and S8 at *http://ajp.amjpathol.org*). For example, CIMP-H adenomas with a *BRAF* mutation acquired MSI as they developed into CISs, suggesting that inactivation of *MLH1* and subsequent genetic instability were late events in the CIMP-H pathway. In addition, CIMP-L and CIMP-N adenomas acquired mutations and CNAs as they developed into CISs and CRCs. Most malignant lesions that exhibited C-cluster 2–type CNAs were derived from tubulovillous adenomas and were characterized by a *KRAS* mutation and CIMP-L (Table 4). On the other hand, most advanced lesions with C-cluster 3–type CNAs were CIMP negative, and more than half of those lesions were derived from tubular adenomas (Table 4).

Discussion

In the present study, we performed integrated genetic and epigenetic analyses with many colorectal neoplasias, including premalignant and malignant lesions. Because of the tight association between CIMP and the clinicopathological features of CRCs, it was anticipated that epigenetic profiling of premalignant lesions would provide important information that would aid in selecting appropriate therapeutic options and predicting clinical outcomes.[35–37] Numerous studies have confirmed that CRCs arise through the accumulation of both genetic and epigenetic alterations; however, the interactions between these alterations early during carcinogenesis remained largely uninvestigated. In addition, only a small fraction of colorectal adenomas may develop into malignant tumors, and progression from adenoma to cancer generally takes >10 years.[38] Thus, the identification of genetic and/or epigenetic alterations that directly correlate with the malignant potential of precursor lesions could facilitate risk assessment and enable prevention of CRCs.

Our comprehensive methylation analysis revealed that the aberrant methylation patterns characterizing CIMP are established early during colorectal tumorigenesis. We found that CIMP-H precursor lesions are strongly associated with SSA and *BRAF*/*KRAS* mutations, whereas CIMP-L precursor lesions are associated with tubulovillous adenomas and frequent *KRAS* mutations. Recently, Yagi et al[39] reported that colorectal adenomas could be classified into high-, intermediate-, and low-methylation epigenotypes, and that the intermediate-methylation epigenotype correlated significantly with a *KRAS* mutation, which is consistent with our observations. These results are indicative of the important relationship between the histological type and the molecular features of premalignant colorectal lesions. By contrast, CIMP-N precursor lesions contained various histological types, including HP, tubular adenoma, and tubulovillous adenoma. Recent reports have shown that a subset of HPs with *BRAF* or *KRAS* mutations progress to SSAs or TSAs, but the time at which aberrant methylation occurs remains unclear.[15,40] Our present findings indicate that, among *KRAS*- or *BRAF*-mutant precursors, the methylation levels of several genes were significantly increased during the progression from HP to adenoma. Concurrent increases in the methylation of multiple genes were further confirmed in mixed lesions containing HP and adenoma components (Table 4).

CIN is an important driving force promoting colorectal tumorigenesis, and recent studies have shown an inverse relationship between CIMP and CIN.[17,33,34,41] Our genome-wide CNA analysis of precursor and malignant lesions revealed that most CNAs are acquired during the progression from adenomas to CISs/CRCs. Consistent with earlier reports, CIMP-H tumors showed few CNAs, whereas both copy number gains and losses were prevalent among CIMP-N tumors.[17,33,34] Furthermore, we discovered that a subset of *KRAS*-mutant/CIMP-L tumors exhibited a unique CNA pattern characterized by frequent copy number gains at chromosomes 7 and 19, with relatively few copy number losses.

In contrast to the tight association between a *BRAF* mutation and CIMP-H in a subset of CRCs, the relationship between a *KRAS* mutation and CIMP status is not fully understood, which likely reflects the molecular complexity of *KRAS*-mutant CRCs.[10,12,13,42] Our analysis sug-

gests that *KRAS*-mutant precursors progress to CRCs via three distinct pathways (Figure 5). First, such as *BRAF*-mutant/CIMP-H tumors, a subset of *KRAS*-mutant tumors, derived from SSAs, exhibit high levels of methylation and few CNAs. Although methylation of *MLH1* is essential for *BRAF*-mutant/CIMP-H adenomas to develop into cancers, *MLH1* methylation was infrequent among *KRAS*-mutant/CIMP-positive tumors. SSAs with a *KRAS* mutation are presumed to be the origin of MSI-negative/CIMP-H CRCs, although the molecular mechanisms underlying the progression from precursors to malignant lesions remain unknown.

Second, we identified a subclass of *KRAS*-mutant/CIMP-positive cancers originating from tubulovillous adenomas or TSAs, in which alteration in *AKT/PIK3CA* signaling was crucially involved. In these tumors, genes associated with an *AKT/PIK3CA* signaling pathway were commonly affected during the progression from adenomas to malignant lesions. Several studies have shown aberrant *AKT/PIK3CA* signaling to be critical for CRC development, and mutations in *PIK3CA*, *AKT1*, *AKT2*, and *PDK1* and amplification of *AKT2/PAK4* are all reportedly associated with a poor prognosis.[43–46] In the present study, concurrent amplification of *AKT2/PAK4* was specifically observed in *KRAS*-mutant/CIMP-positive cancers. Interestingly, we also observed frequent amplification of *BRAF* in this type of tumor. Although further study is needed to clarify its functional role in tumorigenesis, recent studies have shown that *BRAF* amplification promotes acquired resistance to MAPK/ERK kinase 1/2 inhibitors in CRC cells.[47,48]

The molecular profiles of mixed precursor lesions suggest that CIMP is acquired during the progression from flat to protruding-type adenomas (Figure 4A; see also Supplemental Figures S2D, S7A, and S8A at *http://ajp.amjpathol.org*), whereas the *KRAS* mutation status was unchanged between the two components. This means that a *KRAS* mutation precedes CIMP in the *KRAS*/CIMP pathway, and that the acquisition of epigenetic changes, in addition to the *KRAS* mutation, may promote adenoma cell proliferation.

By contrast, *KRAS*-mutant/CIMP-N tumors are derived from tubular adenomas or tubulovillous adenomas and are characterized by a frequent *TP53* mutation and high levels of CNA. An analysis of CIMP-N mixed lesions revealed that a *KRAS* mutation is found only in some advanced components, suggesting that a *KRAS* mutation occurs late during tumorigenesis in this pathway. Thus, *KRAS* mutations appear to play multiple roles during colorectal carcinogenesis. Further study will be required to fully characterize the functional diversity of *KRAS* mutations in CRCs; however, based on the genetic and epigenetic alterations found in CRC and the timing of the occurrence of *KRAS* mutations, we propose that CRCs develop via the four distinct pathways illustrated in Figure 5.

The results of several recent studies support the two-colon concept, which suggests that MSI-positive, CIMP-positive, and *BRAF*-mutant CRCs occur more frequently in the proximal colon.[9–11,13] By contrast, Yamauchi et al[49,50] recently reported that the frequencies of CIMP-H, MSI-high, and *BRAF* mutations in CRCs increased gradually along colorectal subsites from the rectum to the ascending colon. Because of the relatively few samples, we could not confirm this continuum concept in our present study. However, given its potentially significant impact on both basic and clinical research in CRC, testing this theory through further study of a larger population would seem warranted.

Our findings in this study have important implications for translating the molecular basis of carcinogenesis into a clinical benefit. The detection of high-risk precursor lesions is essential for preventing CRCs, and pit pattern observation using magnifying endoscopy enables us to detect neoplastic lesions with malignant potential.[19] We have dissected the morphological, histological, and molecular alterations in precancerous lesions of the colorectum and determined that they are directly linked to one another. Moreover, we provide strong evidence that aberrant pit patterns reflect histological changes and genetic and epigenetic defects in the precursor lesions, and that intratumoral variation in pit patterns could be predictive of the extent of the molecular abnormalities in a given tumor. Such a microstructure-based diagnostic system is readily available to the clinician, although specific skills are required for detailed pit pattern analysis. Further advances in the algorithm for pattern recognition may lead to the development of an innovative diagnostic system for detecting premalignant lesions and a reduction in CRC mortality.

Acknowledgments

We thank Dr. Yutaka Kondo for technical advice on MCAM analysis, Tomo Hatahira for technical assistance, and Dr. William F. Goldman for editing the manuscript.

References

1. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL: Genetic alterations during colorectal-tumor development. N Engl J Med 1988, 319:525–532
2. Grady W, Carethers J: Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008, 135:1079–1099
3. Jones PA, Baylin SB: The epigenomics of cancer. Cell 2007, 128: 683–692
4. Suzuki H, Tokino T, Shinomura Y, Imai K, Toyota M: DNA methylation and cancer pathways in gastrointestinal tumors. Pharmacogenomics 2008, 9:1917–1928
5. Lao VV, Grady WM: Epigenetics and colorectal cancer. Nat Rev Gastroenterol Hepatol 2011, 8:686–700
6. Curtin K, Slattery ML, Samowitz WS: CpG island methylation in colorectal cancer: past, present and future. Patholog Res Int 2011, 2011: 902674
7. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP: CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 1999, 96:8681–8686
8. Toyota M, Ohe-Toyota M, Ahuja N, Issa JP: Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. Proc Natl Acad Sci U S A 2000, 97:710–715
9. Weisenberger D, Siegmund K, Campan M, Young J, Long T, Faasse M, Kang G, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French A, Thibodeau