Published in final edited form as:

Gastroenterology. 2014 August ; 147(2): 418–429.e8. doi:10.1053/j.gastro.2014.04.039.

Differences in DNA Methylation Signatures Reveal Multiple Pathways of Progression from Adenoma to Colorectal Cancer

Yanxin Luo^{1,2}, Chao-Jen Wong², Andrew M. Kaz^{2,3,4}, Slavomir Dzieciatkowski², Kelly T. Carter², Shelli M. Morris², Jianping Wang¹, Joseph E. Willis⁵, Karen W. Makar⁶, Cornelia M. Ulrich^{6,7}, James D. Lutterbaugh⁸, Martha J. Shrubsole⁹, Wei Zheng⁹, Sanford D. Markowitz⁸, and William M. Grady²

¹Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou, P.R. China

²Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

³Research and Development Service, VA Puget Sound Health Care System, Seattle, WA, USA

⁴Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA

⁵Department of Pathology, Case Medical Center, Case Comprehensive Cancer Center and Case Western Reserve University, Cleveland, OH, USA

⁶Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

⁷National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), University of Heidelberg, Heidelberg, Germany GDR

⁸Department of Medicine and Ireland Cancer Center, Case Western Reserve University School of Medicine and Case Medical Center, Cleveland, OH, USA

⁹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA

Abstract

Background & Aims—Genetic and epigenetic alterations contribute to the pathogenesis of colorectal cancer (CRC). There is considerable molecular heterogeneity among colorectal tumors,

© 2014 The American Gastroenterological Association. Published by Elsevier Inc. All rights reserved.

Corresponding author: Yanxin Luo, Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou 510655, P.R. China, phone: 8620-3825-0745, luoyx25@mail.sysu.edu.cn; William M. Grady, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. D4-100, Seattle, WA 98109, phone: 206-667-1107, fax: 206-667-2917 wgrady@fhcrc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest: No conflicts of interest exist for any of the authors.

Author Contributions: YL, SD, KTC, KWM, JEW, SMM, CMU, MJS, WZ and AMK acquired the data; CHW performed the statistical analysis; YL, JW and WMG drafted the manuscript; JW, CMU, SDM and WMG obtained funding and supervised the study.

Data Access: All methylation array data are available at the NCBI Gene Expression Omnibus under accession number GSE48684. The editors and reviewers can access the private data set by the following link: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=djghdmayauisvo&acc=GSE48684>

which appears to arise as polyps progress to cancer. This heterogeneity results in different pathways to tumorigenesis. Although epigenetic and genetic alterations have been detected in conventional tubular adenomas, little is known about how these affect progression to CRC. We compared methylomes of normal colon mucosa, tubular adenomas, and colorectal cancers to determine how epigenetic alterations might contribute to cancer formation.

Methods—We conducted genome-wide array-based studies and comprehensive data analyses of aberrantly methylated loci in 41 normal colon tissue, 42 colon adenomas, and 64 cancers using HumanMethylation450 arrays.

Results—We found genome-wide alterations in DNA methylation in the non-tumor colon mucosa adjacent to tubular adenomas and cancers. Three classes of cancers and 2 classes of adenomas were identified based on their DNA methylation patterns. The adenomas separated into classes of high-frequency methylation (adenoma-H), and low-frequency methylation (adenoma-L). Within the adenoma-H class a subset of adenomas had mutant *KRAS*. Additionally, the adenoma-H class had DNA methylation signatures similar to those of cancers with low or intermediate levels of methylation, whereas the adenoma-L class had methylation signatures similar to that of non-tumor colon tissue. The CpGs sites that were differentially methylated in these signatures are located in intragenic and intergenic regions.

Conclusions—Genome-wide alterations in DNA methylation occur during early stages of progression of tubular adenomas to cancer. These findings reveal heterogeneity in the pathogenesis of colorectal cancer, even at the adenoma step of the process.

Keywords

epigenetic modifications; colon cancer; progression; gene regulation

INTRODUCTION

Colorectal cancer (CRC) results from the progressive accumulation of gene mutations and epigenetic alterations, which induce the initiation and progression of these cancers. Although global DNA hypomethylation was one of the first DNA abnormalities identified in cancer, gene mutations were the first type of DNA alterations unequivocally demonstrated to drive cancer formation. Mutations have been found in adenomas, the precursor neoplasms to colon cancer, as well as in CRCs^[1]. The number of genetic alterations per adenoma genome is significantly smaller than that seen in CRCs^[2]. CRCs have hundreds of mutations and often display genomic instability^[3]. The genomic instability most commonly found in CRC is chromosome instability, which is recognized by the presence of aneuploidy and chromosomal gains and losses. A second form of genomic instability, microsatellite instability, is found in approximately 15% of CRCs and results from inactivation of the DNA mismatch repair system^[4]. Genomic instability is thought to predispose cells to mutations. A subset of the pro-tumorigenic mutations then drive CRC formation^[3, 4, 5, 6]. Based on the identification of multiple molecular subclasses of CRC, it appears that there are multiple molecular pathways that lead to CRC, although many aspects of these pathways are not well defined at this time^[6, 7].

More recently, aberrant DNA methylation has been found in CRC and appears to also play a role in driving CRC formation^[8]. The average CRC genome carries thousands of alterations in the DNA methylation status of CpG dinucleotides. These alterations often affect CpG dinucleotides found in promoter regions of genes and can induce the transcriptional repression of these genes^[9, 10]. The underlying mechanism responsible for the aberrant methylation seen in CRC is not known at this time. Importantly, there is a close association between the DNA methylation status of a locus and its chromatin structure. Regions bound by Polycomb group proteins (PcG) are commonly subjected to aberrant methylation in cancer^[11].

Analyses of established CRCs have also revealed that there is a molecular subclass of CRCs that has an excessive number of aberrantly methylated CpG dinucleotides^[12, 13]. These CRCs, which are designated as having a CpG Island Methylator Phenotype (CIMP), account for approximately 15–20% of all CRCs and often carry mutant *BRAF*^[14]. While virtually all CRCs have aberrantly methylated genes, the CIMP CRCs are recognized by an exceptionally high proportion of methylated loci^[14]. It appears that CIMP CRCs arise from sessile serrated polyps (SSPs), which is an observation supported not only by the occurrence of CIMP in approximately 30% of SSPs, but also by shared clinical features with CIMP CRCs, such as a predisposition to occur in the proximal colon, to occur in women, and to contain a high frequency of mutant *BRAF* and *MSI*^[7, 15, 16]. In contrast, CIMP is rarely found in tubular or tubulovillous adenomas^[17, 18].

The identification of candidate methylated genes in aberrant crypt foci and adenomas has suggested that epigenetic alterations may play a role in both the initiation and progression of conventional tubular adenomas and CRC as well as in CIMP CRCs^[19]. Moreover, aberrantly methylated genes can be found in the histologically normal colon of individuals with an increased predisposition to CRC and in older individuals, suggesting that aberrant DNA methylation may be one of the earliest molecular events that initiate CRC formation^[20, 21]. In order to further assess the role of epigenetic alterations in the initiation and progression of CRC, we carried out an epigenome-wide analysis of normal colon mucosa, tubular adenomas, and CRCs. We assessed the DNA methylation status of CpG dinucleotides in CpG islands (defined as regions with a fraction of C and G dinucleotides >50%), CpG shores (which are CpG rich regions located within 2 kb of islands) and CpG shelves (which are CpG-rich regions flanking shores)^[22, 23].

RESULTS

Identification and validation of methylated probes on the HumanMethylation450 arrays that are differentially methylated between normal colon, tubular adenomas, and CRC

We and others have reported previously that results from the HumanMethylation450 BeadChips (HM450) are technically robust but that there is a measurable false discovery rate^[24, 25]. Thus, we initially conducted technical validation studies and biological validation studies of a subset of differentially methylated CpGs (N=4) found on the HM450 arrays (described in detail in Supplemental Methods and Results). All of the CpG probes (N=4) that were either aberrantly methylated in cancers/adenomas were confirmed to be methylated by pyrosequencing. In addition, these 4 CpG probes were also assessed in two

independent collections of samples using pyrosequencing and were shown to have the same methylation pattern as seen in the samples run on the methylation arrays (Table S1, S2 and Figure S1, S2). These results demonstrate that the data generated from the HM450 arrays is reproducible and generalizable to colon adenomas and CRCs.

The methylation status in normal colon mucosa near concurrent CRC differs from that of normal colon mucosa from healthy individuals (no history of CRC and no concurrent CRC)

In order to assess the role of aberrant DNA methylation in the polyp→cancer sequence, we first determined the methylation status of normal colon mucosa using HM450 arrays, which assess the DNA methylation status of 485,577 CpG dinucleotides. Greater than 90% of the CpG islands in the genome are assessed using the HM450 array^[22]. We analyzed both the normal colon mucosa from people with no history of colon neoplasms, who are considered to be in an average risk group for CRC, and the normal colon mucosa from people with concurrent CRC, who are at increased risk of metachronous CRC^[26].

After filtering the data as described in the Methods section, we identified 343 differentially methylated probes using a q -value of $1E-3$ (Supplemental Data Table S3). As shown in Figure 1, these probes, which are located in 65 loci, can distinguish the DNA methylation levels in normal samples from CRC patients compared to normal mucosa from healthy individuals (q -value $< 1E-4$). The majority of these 343 probes (86%) have higher methylation levels in the CRC-associated mucosa compared to colon mucosa in cancer-free individuals. Additional studies will need to be done to determine if these probes are potential markers of a field cancerization process^[27, 28].

Epigenetic alterations are a common occurrence in colon adenomas

We next assessed the methylation status of CpG dinucleotides in 18 adenomas using the HM450 arrays. Adenomas are a well-recognized transition step between normal colon and colorectal cancer. The histologic heterogeneity of colon polyps and associated risk of CRC has been appreciated for many years. Recently, unique molecular features have been found in the different histologic types of adenomas, and it has been argued that these differences affect the likelihood of the polyps progressing to CRC^[7, 29]. This led us to compare the methylation state of adenomas and established CRCs to normal colon mucosa. We identified 86,460 differentially methylated probes (DMPs) with a q -value of $1E-5$. Nearly 40% were hypermethylated and 60% were hypomethylated in the adenomas compared to the normal colon mucosa.

The DNA methylation status of CpGs located in different classes of loci, including CpG islands, shores, and shelves as well as promoters, gene bodies, and intergenic regions was assessed. The HM450 array categorizes probes based on gene regions into three major gene feature groups: promoter (5'UTR, TSS200, TSS1500, and 1st exons), intragenic regions (body and 3'UTR), and intergenic regions^[30]. We found a higher proportion of hypomethylated probes in non-promoter regions (25% vs. 15% in promoter regions) and a higher proportion of hypermethylated probes in promoter regions compared to non-promoter regions (Supplemental Data Figure S3). Approximately 17%, 17% and 19% of probes in CpG islands, shores and shelves, respectively, are differentially methylated (colon adenoma

vs. normal colon), suggesting that the proportion of DMPs among these three classes of loci are not significantly different.

Next, we assessed the overall DNA methylation patterns of the adenomas. Cluster analysis of the 10,000 most variable CpG probes (2.5%) in the adenomas revealed 2 distinct epigenotypes, which we have termed adenoma-H (high methylator phenotype) and adenoma-L (low methylator phenotype) (Figure 2). Of note, using Leave-One-Out (LOO) cross-validation, we determined that the misclassification rate of the cluster results is 5.5%, which suggests that the adenoma-H and adenoma-L groups are truly unique entities and not a consequence of a chance association secondary to multiple comparisons. Comparison of methylated CpGs between the adenoma-H and adenoma-L groups revealed 1,196 differentially methylated probes (q -value $< 1E-4$; Supplemental Data Table S4). Among these DMPs, 89.7% are hypermethylated in the adenoma-H polyps compared to the adenoma-L group. Most of the probes that are hypomethylated ($N=58/124$, 47%) in the adenoma-H group vs. adenoma-L group are located in the intergenic or intragenic regions (Supplemental Data Figure S4). No association between DNA methylation and polyp size or histology was found.

Interestingly, the adenoma-L polyps have a methylation pattern similar to normal colon mucosa whereas the adenoma-H polyps are more similar to CIMP-negative CRC (Figure 3 and Supplemental Data Figures S5 and S6). We also found that *KRAS* mutations occur frequently in a subset of adenoma-H polyps ($N=7/11$, 63.3%), and that mutant *APC*, *BRAF*, and *PIK3CA* occur in small subsets ($N=5/18$, 27.8%; $N=1/18$, 5.6%; $N=3/18$, 16.7%, respectively) of all the adenomas. Mutations in *TP53*, *SRC*, *FBXW7*, or *CTNNB1* were not found in any of the adenomas (Table 1). Also, the adenoma-H polyps with mutant *KRAS* exhibit a unique methylation pattern compared to the adenoma-H polyps with wild-type *KRAS*.

In order to validate the discovery of unique epigenotypes of tubular adenomas, we assessed the methylation patterns of an independent collection of 24 adenomas using the HM450 arrays. We used the same set of probes identified in the discovery set of samples (Figure 2) and confirmed the previous cluster results. We also identified an adenoma-H and an adenoma-L subgroup, with obvious heterogeneity existing in the adenoma-H group. Mutation profiling results were similar in this second set of samples compared to the discovery set of samples. We found *KRAS* mutations occur frequently in a subset of adenoma-H polyps ($N=10/19$, 52.6%). Furthermore, we found mutations in *APC* ($N=10/24$, 41.7%), *BRAF* ($N=2/24$, 8.3%), *PIK3CA* ($N=4/24$, 16.7%) and *CTNNB1* ($N=1/24$, 4.2%), which is similar to the frequencies found in the discovery set of samples. *SRC*, or *FBXW7* mutations were not found in any of the adenomas in either set of samples. In addition, there are no significant differences in the age, gender and location of the adenomas between adenoma-H and adenoma-L clusters (Table 1).

Methylator phenotype in CRC

After assessing the methylation status of the normal colon mucosa and of colon tubular adenomas, we performed comprehensive DNA methylation profiling of 64 CRCs. We also

assessed the mutation status of *KRAS*, *BRAF*^{V600E}, *APC*, *SRC*, *FBXW7*, *TP53* and *PIK3CA* in the CRCs as we had done with the adenomas.

Using a recursively partitioned mixture model (RPM) and a hierarchical model (hCluster) clustering approach on the 10,000 most variable CpGs, we identified 3 distinct CRC subgroups, indicated as cluster 1 (N=26/64, 53%), cluster 2 (N=13/64, 22%) and cluster 3 (N=15/64, 25%), which we have termed Methyl-Low (low methylation pattern; M-L), Methyl-Intermediate (intermediate methylation pattern, M-I) and Methyl-High (high methylation pattern; M-H), respectively (Figure 3). The M-H subgroup is enriched for CIMP-High (CIMP-H) cancers (N=11/15, 68.8%)^[14]. The remaining M-H CRCs (N=4/15, 31.2%) were CIMP-Low (CIMP-L), which were characterized by having 1–2 methylated CIMP loci. Consistent with other studies, the M-H subgroup CRCs have frequent *BRAF*^{V600E} mutations (N=9/15, 60.0%) and are often MSI (N=7/15, 46.7%)^[14]. The M-H subgroup also has a relative paucity of *APC* mutations (N=2/15, 13.3%) and *TP53* mutations (N=2/15, 13.3%) when compared to the CRCs of the other cluster subgroups, in which there are no *BRAF*^{V600E} mutations, infrequent MSI (N=2/39, 4.1%) and a higher frequency of *APC* mutations (N=13/39, 34.7%) and *TP53* mutations (N=15/39, 42.9%) (Table 1).

CRCs in the M-I subgroup exhibit an intermediate methylation pattern when compared to the other 2 cluster subgroups (Figure 3). In the M-I subgroup, mutant *KRAS* (N=8/13, 61.5%) and mutant *APC* (N=4/13, 30.8%) were frequent, which is similar to the frequencies seen in other studies^[6, 31] (Table 1). The clinicopathological features of the M-I subgroup suggest that this subtype of cancers corresponds to the classifications CIMP2 or CIMP-Low that have been described^[6, 31]. The M-L group is enriched for non-CIMP cancers as determined by the CIMP MethylLight assay panel and has a relatively low frequency of methylated CpGs^[14]. Of note, the M-H, M-I, and M-L classes of CRCs identified in this study were validated through the analysis of two independent collections of CRCs, which identified the same 3 subsets of CRCs (described in detail in the Supplemental Methods and Results)^[6, 13].

Association of methylated loci with polycomb group binding sites in colon adenomas and CRCs

After identifying methylated CpGs that occur commonly in adenomas and CRC, we next assessed which of these methylated CpGs occur in loci commonly occupied by Polycomb group proteins (PcG). Previous studies have demonstrated that genes whose expression is affected by methylation often are Polycomb Group (PCG) Target loci^[11]. We found that 55% of the hypermethylated loci (defined by *q*-value < E-6 and a beta-value > 0.3 vs. normal colon; N=756) occurred in areas of bivalent chromatin in all 3 methylation classes of CRCs (Figure 4), which is consistent with previously published findings^[13]. In addition, we found that the PcG-marked hypermethylated loci found in the CRCs were also hypermethylated in the adenoma-H polyps, but not the adenoma-L polyps (Figure 4 and Supplemental Data Figure S7 and Table S5, S6). These findings suggest that the adenoma-H polyps may be more progressed towards CRCs than adenoma-L polyps and also suggest that the aberrant methylation of genes that are likely to be transcriptionally repressed by this epigenetic alteration occurs early in the polyp to cancer progression sequence.

Hypervariability of methylation occurs early in the adenoma-carcinoma progression sequence

In addition to assessing the methylation status of CpGs in the normal colon and colon neoplasms, we assessed the inter-sample variability in methylation of the CpGs in the polyps and CRCs. We observed higher inter-sample variability in the frequency of methylated loci across the CRCs and adenomas as compared to the normal samples, which had a small amount of inter-sample variability (Figure 5A and B). Of interest, the degree of epigenetic hypervariability found in the adenomas and CRCs is similar (Figure 5C), suggesting the variability in methylation status occurs early in CRC formation (Figure 5D).

Adenomas and CRCs with a high methylation pattern have a unique pattern of methylated intergenic and intragenic CpGs

Recent studies suggest that CIMP CRCs are derived from sessile serrated adenomas^[32], a CRC precursor lesion exhibiting unique morphologic features and epigenetic characteristics^[33, 34, 35]. Prior studies have shown that up to 30% of serrated adenomas have a CIMP methylation pattern and that tubular adenomas are rarely CIMP^[33]. These findings suggest that there are at least 2 unique polyp→cancer progression sequences that can lead to CRC^[7]. In light of the identification of CIMP in serrated adenomas, we carried out a detailed assessment of the epigenome in tubular adenomas and CRCs by assessing the methylation status of CpGs located in intergenic and intragenic regions. We compared the methylation status of these CpGs in the M-H CRCs and M-L CRCs and found 3,555 intergenic and 3,676 intragenic differentially methylated probes (q -value $< 1E-5$) (Supplemental Data Figure S8 and Table S7). When we compared the M-H to M-I/M-L CRCs, we identified 3,122 intragenic and 3,114 intergenic DMPs (q -value $< 1E-5$, Supplemental Data Table S8).

Adenoma-H polyps share an epigenetic signature with CRCs of low/intermediate methylation pattern (M-I/M-L)

Based on our observations of there being two discrete classes of adenomas based on methylation patterns (adenoma-L and adenoma-H) and on our identification of patterns of methylated intergenic and intragenic CpGs that distinguish M-H from M-L CRCs, we next assessed these CRC identified intergenic and intragenic CpGs in adenomas. Interestingly, this analysis revealed that the epigenetic signature of the adenoma-H polyp class is similar to the M-I/M-L CRCs, whereas the adenoma-L polyp class is similar to normal colon mucosa (Figure 6). Of note, similar results were found in a validation set of adenomas (Supplemental Figure S9). Multi-dimensional scaling analysis agreed with the cluster analysis, which showed adenoma-H polyps to be similar to M-I/M-L CRCs and adenoma-L polyps to be similar to normal colon mucosa (Supplemental Data Figure S10). These findings suggest that the adenoma-H polyps may be the origin of M-I/M-L CRCs, whereas adenoma-L polyps may be polyps that ultimately will not progress to CRC. Additionally, these studies reveal significant heterogeneity in the epigenome of adenomas and suggest that the epigenome may portend the fate of tubular adenomas.

DISCUSSION

In these studies we have found considerable genetic and epigenetic heterogeneity among not only CRCs but also among adenomas^[6, 36, 37]. The existence of different classes of CRC that differ based on DNA methylation patterns was first proposed by Issa and colleagues in 1999 when they identified a CpG Island Methylator Phenotype (CIMP) class of CRCs^[8]. Our studies provide additional insight into CIMP CRCs by showing that in the high methylation pattern CRCs (M-H) there is a low frequency of *APC* mutations, consistent with findings from the TCGA^[6], and suggests CIMP CRCs may arise by a WNT independent pathway^[38]. Furthermore, we identified a series of differentially methylated probes located in intragenic and intergenic regions that distinguish M-H from M-I/M-L CRCs. These probes lie in regions called tissue differential methylated regions (tDMRs), and their methylation status is tissue specific, perhaps reflecting the stem cell methylation pattern from which the tissue is derived^[23, 39]. These findings suggest that the M-H CRCs may be derived from a different stem cell precursor than M-I/M-L CRCs, which may explain the unique clinicopathological features of CIMP CRCs compared to non-CIMP CRCs^[14].

We also analyzed the methylation status of normal colon mucosa. Our analysis of the normal colon mucosa included samples from individuals with concurrent CRC, who may have a field defect in their colons that predisposes them to adenomas and CRC, and from individuals who are cancer free. Consistent with prior studies that have used limited gene panels, we observed differences in the methylation patterns between these two groups^[20, 40, 41]. Further studies using a prospective study design will be needed to confirm whether the methylation state in the normal colon predicts risk for developing CRC.

Prior studies of the methylation status of adenomas, using small sample sizes and small panels of candidate genes, have shown that methylation is present and heterogeneous in adenomas^[15, 42, 43]. Additionally, studies have demonstrated that CIMP can be detected in serrated polyps but is rare in tubular adenomas^[15, 44]. Using HumanMethylation450 arrays and a moderately sized set of 18 tubular adenomas, we observed that aberrant DNA methylation occurs commonly in tubular adenomas. Moreover, we identified subclasses of tubular adenomas based on their methylation status and *KRAS* mutation status. We studied the intergenic and intragenic CpGs, rather than promoter CpGs, because their methylation status appears to reflect the epigenome of the stem cell population from which the tumor cells are derived^[45]. We found that the adenoma-H polyps have a methylation pattern that is similar to that of M-I/M-L CRCs, whereas the adenoma-L polyps have a methylation pattern that is similar to that of normal colon epithelium. Importantly, these results were confirmed in an independent collection of 24 adenomas. These results suggest that adenoma-H polyps may be the precursors for M-I/M-L CRCs. Our findings also suggest adenoma-L polyps may have a low potential to progress to CRC and may represent the 90% of adenomas that do not evolve into CRC.

Our studies provide insight into the colon adenoma to cancer progression sequence. This sequence of histologic progression of normal colon epithelial cells to CRC is widely believed to be driven by the serial acquisition of gene mutations and epigenetic alterations. It appears likely that the different subclasses of CRC arise through different polyp to cancer

progression sequences with CIMP CRCs arising from serrated polyps and non-CIMP CRCs arising from tubular adenomas^[46]. Our studies demonstrate there are subclasses of adenomas recognized by their epigenotype and *KRAS* mutation status and raise the possibility that one of these subclasses, adenoma-H polyps, may be the precursors for CRCs with a low/intermediate methylation pattern (M-I/M-L CRCs).

In summary, our results confirm those of prior studies that have shown aberrant DNA methylation occurs early in CRC formation and may predispose histologically normal tissue to become neoplastic. In addition, we have found that the epigenetic state of the adenomas may influence the propensity of the adenoma to undergo malignant transformation and portend the epigenotype of the resulting CRC.

METHODS

Primary human tissue samples

DNA extracted from snap-frozen tissues was used for the studies using the HumanMethylation450 arrays. A detailed description of the samples used is available in Supplemental Methods.

DNA isolation and bisulfite conversion

Genomic DNA was extracted and bisulfite modified as previously described^[47].

Molecular characterization

The CpG Island Methylator Phenotype (CIMP) status and Microsatellite instability (MSI) status of the CRCs were assessed using the methods as described previously^[47, 48]. Gene mutation status of *KRAS*, *BRAF*^{V600E}, *APC*, *TP53* and *PIK3CA* was determined using the qBiomarker Somatic Mutation PCR System Arrays/Human Colon Cancer (Qiagen, USA) following the manufacturer's protocol, as described previously^[24].

HumanMethylation450 Array

Illumina Infinium HumanMethylation450 BeadChips (Illumina, San Diego, Ca, USA) were used for these studies. The processing of the DNAs on the methylation arrays was conducted in the Genomics Shared Resources at the Fred Hutchinson Cancer Research Center according to the manufacturer's specifications (Illumina). Data filtering and normalization procedures are described in Supplemental Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the outstanding service provided by the Genomics Shared Resources (FHCRC) and the Cooperative Human Tissue Network for the tissues they provided. We also thank the ColoCare team (Chris Velicer, Rebecca Holmes, Stephanie Zschäbitz, Kathy Vickers, Rachel Wilbur, Shannon Rush, and Sara Bates and others) for their assistance on these studies. In addition, we would like to thank Toshinori Hinoue and Peter W. Laird for kindly sharing their data. Finally, we would like to thank the study participants who kindly agreed to provide tissues for analysis.

Grant Support: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number RO1CA115513, P30CA15704, U01CA152756, U54CA143862, and P01CA077852 (WMG), P50CA95103 (ZW), R01CA121060, P30CA68485, K07CA122451 (MJS). The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Institutes of Health. Support for these studies was also provided by a Burroughs Wellcome Fund Translational Research Award for Clinician Scientist (WMG), Program of Introducing Talents of Discipline to Universities of China (B12003, JW) and International Science & Technology Cooperation Program of China (2011DFA32570, JW), National Natural Science Foundation of China (81201920, YL); and 5P50CA150964 (SDM).

REFERENCES

- Gal-Yam EN, Egger G, Iniguez L, et al. Frequent switching of polycomb repressive marks and DNA hypermethylation in the pc3 prostate cancer cell line. *Proc Natl Acad Sci U S A*. 2008; 105:12979–12984. [PubMed: 18753622]
- Houseman EA, Christensen BC, Yeh RF, et al. Model-based clustering of DNA methylation array data: A recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. *BMC Bioinformatics*. 2008; 9:365. [PubMed: 18782434]
- Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science*. 2013; 339:1546–1558. [PubMed: 23539594]
- Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology*. 2008; 135:1079–1099. [PubMed: 18773902]
- Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science*. 2007; 318:1108–1113. [PubMed: 17932254]
- Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–337. [PubMed: 22810696]
- Jass JR. Molecular heterogeneity of colorectal cancer: Implications for cancer control. *Surg Oncol*. 2007; 16(Suppl 1):S7–S9. [PubMed: 18023574]
- Toyota M, Ho C, Ahuja N, et al. Identification of differentially methylated sequences in colorectal cancer by methylated cpG island amplification. *Cancer Res*. 1999; 59:2307–2312. [PubMed: 10344734]
- Issa JP, Ottaviano YL, Celano P, et al. Methylation of the oestrogen receptor cpG island links ageing and neoplasia in human colon. *Nat Genet*. 1994; 7:536–540. [PubMed: 7951326]
- van Engeland M, Derks S, Smits KM, et al. Colorectal cancer epigenetics: Complex simplicity. *J Clin Oncol*. 2011; 29:1382–1391. [PubMed: 21220596]
- Ohm JE, McGarvey KM, Yu X, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet*. 2007; 39:237–242. [PubMed: 17211412]
- Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer*. 2004; 4:988–993. [PubMed: 15573120]
- Hinoue T, Weisenberger DJ, Lange CP, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*. 2012; 22:271–282. [PubMed: 21659424]
- Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with braf mutation in colorectal cancer. *Nat Genet*. 2006; 38:787–793. [PubMed: 16804544]
- Burnett-Hartman AN, Newcomb PA, Potter JD, et al. Genomic aberrations occurring in subsets of serrated colorectal lesions but not conventional adenomas. *Cancer Res*. 2013; 73:2863–2872. [PubMed: 23539450]
- Spring KJ, Zhao ZZ, Karamatic R, et al. High prevalence of sessile serrated adenomas with braf mutations: A prospective study of patients undergoing colonoscopy. *Gastroenterology*. 2006; 131:1400–1407. [PubMed: 17101316]
- Rex DK, Ahnen DJ, Baron JA, et al. Serrated lesions of the colorectum: Review and recommendations from an expert panel. *Am J Gastroenterol*. 2012; 107:1315–1329. quiz 1314, 1330. [PubMed: 22710576]

18. Kim YH, Kakar S, Cun L, et al. Distinct cpg island methylation profiles and braf mutation status in serrated and adenomatous colorectal polyps. *Int J Cancer*. 2008; 123:2587–2593. [PubMed: 18798261]
19. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol*. 2011; 8:686–700. [PubMed: 22009203]
20. Shen L, Kondo Y, Rosner GL, et al. Mgmt promoter methylation and field defect in sporadic colorectal cancer. *J Natl Cancer Inst*. 2005; 97:1330–1338. [PubMed: 16174854]
21. Ahuja N, Li Q, Mohan AL, et al. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res*. 1998; 58:5489–5494. [PubMed: 9850084]
22. Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 cpg sites in the human genome. *Epigenetics*. 2011; 6:692–702. [PubMed: 21593595]
23. Irizarry RA, Ladd-Acosta C, Wen B, et al. The human colon cancer methylome shows similar hypoand hypermethylation at conserved tissue-specific cpg island shores. *Nat Genet*. 2009; 41:178–186. [PubMed: 19151715]
24. Luo Y, Kaz AM, Kanngurn S, et al. Ntrk3 is a potential tumor suppressor gene commonly inactivated by epigenetic mechanisms in colorectal cancer. *PLoS Genet*. 2013; 9:e1003552. [PubMed: 23874207]
25. Busche S, Ge B, Vidal R, et al. Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia. *Cancer Res*. 2013; 73:4323–4336. [PubMed: 23722552]
26. Saini SD, Kim HM, Schoenfeld P. Incidence of advanced adenomas at surveillance colonoscopy in patients with a personal history of colon adenomas: A meta-analysis and systematic review. *Gastrointest Endosc*. 2006; 64:614–626. [PubMed: 16996358]
27. Shen L, Issa JP. Epigenetics in colorectal cancer. *Curr Opin Gastroenterol*. 2002; 18:68–73. [PubMed: 17031233]
28. Grady WM, Parkin RK, Mitchell PS, et al. Epigenetic silencing of the intronic microrna hsa-mir-342 and its host gene evl in colorectal cancer. *Oncogene*. 2008; 27:3880–3888. [PubMed: 18264139]
29. Yagi K, Akagi K, Hayashi H, et al. Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res*. 2010; 16:21–33. [PubMed: 20028768]
30. Dedeurwaerder S, Defrance M, Calonne E, et al. Evaluation of the Infinium methylation 450k technology. *Epigenomics*. 2011; 3:771–784. [PubMed: 22126295]
31. Shen L, Toyota M, Kondo Y, et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A*. 2007; 104:18654–18659. [PubMed: 18003927]
32. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*. 2010; 138:2088–2100. [PubMed: 20420948]
33. Yamamoto E, Suzuki H, Yamano HO, et al. Molecular dissection of premalignant colorectal lesions reveals early onset of the cpg island methylator phenotype. *Am J Pathol*. 2012; 181:1847–1861. [PubMed: 22995252]
34. Jass JR. Serrated adenoma of the colorectum and the DNA-methylator phenotype. *Nat Clin Pract Oncol*. 2005; 2:398–405. [PubMed: 16130936]
35. Gaiser T, Meinhardt S, Hirsch D, et al. Molecular patterns in the evolution of serrated lesion of the colorectum. *Int J Cancer*. 2012
36. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med*. 2009; 361:2449–2460. [PubMed: 20018966]
37. Lugli A, Jass JR. Types of colorectal adenoma. *Verh Dtsch Ges Pathol*. 2006; 90:18–24. [PubMed: 17867575]
38. Kawasaki T, Noshio K, Ohnishi M, et al. Correlation of beta-catenin localization with cyclooxygenase-2 expression and cpg island methylator phenotype (cimp) in colorectal cancer. *Neoplasia*. 2007; 9:569–577. [PubMed: 17710160]
39. Doi A, Park IH, Wen B, et al. Differential methylation of tissue- and cancer-specific cpg island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet*. 2009; 41:1350–1353. [PubMed: 19881528]

40. Worthley DL, Whitehall VL, Buttenshaw RL, et al. DNA methylation within the normal colorectal mucosa is associated with pathway-specific predisposition to cancer. *Oncogene*. 2010; 29:1653–1662. [PubMed: 19966864]
41. Ushijima T. Epigenetic field for cancerization. *J Biochem Mol Biol*. 2007; 40:142–150. [PubMed: 17394762]
42. Kim YH, Petko Z, Dzieciatkowski S, et al. CpG island methylation of genes accumulates during the adenoma progression step of the multistep pathogenesis of colorectal cancer. *Genes Chromosomes Cancer*. 2006; 45:781–789. [PubMed: 16708352]
43. Kim KM, Lee EJ, Ha S, et al. Molecular features of colorectal hyperplastic polyps and sessile serrated adenoma/polyps from Korea. *Am J Surg Pathol*. 2011; 35:1274–1286. [PubMed: 21836485]
44. O'Brien MJ, Yang S, Mack C, et al. Comparison of microsatellite instability, CpG island methylation phenotype, braf and kras status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. *Am J Surg Pathol*. 2006; 30:1491–1501. [PubMed: 17122504]
45. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010; 466:253–257. [PubMed: 20613842]
46. Bettington M, Walker N, Clouston A, et al. The serrated pathway to colorectal carcinoma: Current concepts and challenges. *Histopathology*. 2013; 62:367–386. [PubMed: 23339363]
47. Luo Y, Tsuchiya KD, Il Park D, et al. Ret is a potential tumor suppressor gene in colorectal cancer. *Oncogene*. 2013; 32:2037–2047. [PubMed: 22751117]
48. Grady WM, Rajput A, Myeroff L, et al. Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. *Cancer Res*. 1998; 58:3101–3104. [PubMed: 9679977]

Figure 1. Differentially methylated CpG probes in the normal colon mucosa between people with no history of CRC and people with concurrent CRC

The strip-plot shows the top 65 probes that distinguish these 2 groups of tissue samples with a q -value $< 1E-4$. The Y-axis represents the methylation level (β value) of each case ranging from 0 (completely unmethylated) to 1 (completely methylated). The corresponding CpG probe ID on the HumanMethylation450 array is indicated on the X-axis. Each dot represents the β value of a single case for each targeted CpG probe. (Normal-Cancer (Normal-C) = normal colon mucosa samples from patients with concurrent CRC; Normal-Healthy (Normal-H) = samples from patients without a history of CRC).

Figure 2. Identification and validation of clustering of colorectal adenomas and heatmap representation of DNA methylation array data

DNA methylation status was assessed using HumanMethylation450 arrays. Each column represents one sample and each row represents one of the top 5,000 most variable probes. The probes are arranged based on the order of unsupervised hierarchical cluster analysis using a correlation distance metric and average linkage method. The DNA methylation M-values are represented by using a color scale from green (low DNA methylation) to red (high DNA methylation). The presence of *TP53*, *PIK3CA*, *KRAS*, *BRAF*^{V600E}, *CTNNB1* or *APC* mutations is indicated by a colored block (no color = wild-type). Two subgroups (adenoma-L and adenoma-H) were identified using the clustering analysis on the discovery set of 18 adenoma samples (middle panel). These results were confirmed in an independently collected validation set of 24 adenomas using the same probes identified in the discovery set of adenomas (right panel). Normal colon samples (left panel, N=41) were used for reference.

Figure 3. Cluster analysis of CRCs and heatmap representation of DNA methylation array data
 DNA methylation status was assessed using HumanMethylation450 arrays. Each column represents one sample and each row represents one of the top 5,000 most variable probes. The DNA methylation M-values are represented by using a color scale from green (low DNA methylation) to red (high DNA methylation). Three subgroups were identified by clustering and are indicated above the heatmap (Methyl-Low, Methyl-Intermediate, and Methyl-High). The presence of *TP53*, *SRC*, *PIK3CA*, *KRAS*, *FBXW7*, *CTNNB1*, *BRAF*^{V600E} or *APC* mutations and CIMP status, which was determined using the Weisenberger panel of

CIMP CpGs with MethyLight assays, are indicated by a colored block (no color = wild-type). Normal colon samples (left panel, N=41) were used for reference.

Figure 4. Venn diagram of hypermethylated genes that are PcG-marked

A) PcG-marked hypermethylated genes in CRC samples (419 genes); **B)** PcG-marked hypermethylated genes in adenoma samples (545 genes); **C)** PcG-marked hypermethylated genes in adenoma and CRC samples (384 genes). The number in each area indicates the number of genes in that area.

Figure 5. Increased inter-sample methylation variability in colon adenomas and CRCs

A) The standard deviation of 1,000 randomly selected probes in normal colon samples and adenomas is shown; **B)** The standard deviation of 1,000 randomly selected probes in normal colon samples and CRCs is shown; **C)** The standard deviation of 1,000 randomly selected probes in adenomas and CRCs is shown; **D)** The standard deviation of 1,000 randomly selected probes in adenoma-L polyps and adenoma-H polyps is shown. The gray solid line in each image (from left low corner to the up right corner) is the identity line, which indicates the location of a1:1 ratio between the comparison sets, whereas the dashed red line indicates the best-fit linear regression of the CpG probes. The red line above the gray line means the group on the Y-axis has higher variability than the one on the X-axis.

Figure 6. Cluster analysis of M-H CRCs, M-I/M-L CRCs and adenoma-H adenomas using differentially methylated probes in intragenic/intergenic regions

Differentially methylated intragenic (A) and intergenic (B) CpG probes between M-H and M-I/M-L CRCs were used in unsupervised clustering analysis of the adenoma-H polyps and CRCs. The results show that the methylation pattern in adenoma-H polyps is similar to that in M-I/M-L CRCs but not to that in M-H CRCs. (C) Schematic of concept that there are multiple adenoma to cancer pathways, which can be identified by DNA methylation signatures in the normal colon, adenomas, and cancers (See Discussion).

Table 1
Clinical and genetic characteristics of DNA methylation-based subtypes of CRC and adenoma samples

	Cancer						Adenoma (Training Set)						Adenoma (Validation Set)					
	All		Cluster 1		Cluster 2		Cluster 3		All		High		Low		All		High	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total	64	100.0	36	56.3	13	20.3	15	23.4	18	100.0	11	61.0	7	39.0	24	100.0	19	79.2
Age																		
< 50	13	20.3	9	25.0	3	23.1	1	6.7	7	38.9	3	27.3	4	57.1	1	4.2	1	5.3
50 – 60	18	28.1	10	27.8	5	38.5	3	20.0	3	16.7	2	18.2	1	14.3	8	33.3	6	31.6
60–70	15	23.4	10	27.8	3	23.1	2	13.3	5	27.8	4	36.4	1	14.3	5	20.8	4	21.1
>70	18	28.1	7	19.4	2	15.4	9	60.0	3	16.7	2	18.2	1	14.3	10	41.7	8	42.1
Sex																		
Male	23	35.9	15	41.7	5	38.5	3	20.0	4	22.2	3	27.3	1	14.3	9	37.5	7	36.8
Female	41	64.1	21	58.3	8	61.5	12	80.0	14	77.8	8	72.7	6	85.7	15	62.5	12	63.2
Location																		
Proximal colon	28	43.8	9	25.0	8	61.5	11	73.3	12	66.7	8	72.7	4	57.1	14	58.3	11	57.9
Transverse colon	3	4.7	2	5.6	0	0.0	1	6.7	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Distal colon	25	39.1	17	47.2	5	38.5	3	20.0	5	27.8	2	18.2	3	42.9	9	37.5	7	36.8
Rectum	6	9.4	6	16.7	0	0.0	0	0.0	1	5.6	1	9.1	0	0.0	1	4.2	1	5.2
Unknown	2	3.1	2	5.6	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
MSI/MSS																		
MSI	9	14.1	2	5.6	0	0.0	7	46.7										
MSS	55	85.9	34	94.4	13	100.0	8	53.3										
Stage																		
I or II	21	32.8	11	30.6	4	30.8	6	40.0										
III or IV	43	67.2	25	69.4	9	69.2	9	60.0										
TP53																		
Mutant	23	35.9	15	41.7	6	46.2	2	13.3	4	22.2	1	9.1	3	42.9	5	20.8	4	21.1
Wild-type	41	64.1	21	58.3	7	53.8	13	86.7	14	77.8	10	90.9	4	57.1	19	79.2	15	78.9
SRG																		
Mutant	1	1.6	0	0.0	0	0.0	1	6.7	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Wild-type	63	98.4	36	100.0	13	100.0	14	93.3	18	100.0	11	100.0	7	100.0	24	100.0	19	100.0
PIK3CA																		
Mutant	7	10.9	3	8.3	3	23.1	1	6.7	3	16.7	2	18.2	1	14.3	4	16.7	4	21.1
Wild-type	57	89.1	33	91.7	10	76.9	14	93.3	15	83.3	9	81.8	6	85.7	20	83.3	15	78.9
KRAS																		
Mutant	29	45.3	16	44.4	8	61.5	5	33.3	8	44.4	7	63.6	1	14.3	11	45.8	10	52.6
Wild-type	35	54.7	20	55.6	5	38.5	11	73.3	10	55.6	4	36.4	6	85.7	13	54.2	9	47.4
FBXW7																		
Mutant	3	4.7	1	2.8	0	0.0	2	13.3	0	0.0	0	0.0	0	0.0	1	4.2	1	5.3

			Cancer						Adenoma (Training Set)						Adenoma (Validation Set)					
			All		Cluster 1		Cluster 2		Cluster 3		All		High		Low		All		High	
			N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<i>CTNNB1</i>	Wild-type		61	95.3	35	97.2	13	100.0	13	86.7	18	100.0	11	100.0	7	100.0	23	95.8	18	94.7
	Mutant		2	3.1	1	2.8	1	7.7	0	0.0	0	0.0	0	0.0	0	0.0	1	4.2	1	5.3
			62	96.9	35	97.2	12	92.3	15	100.0	18	100.0	11	100.0	7	100.0	23	95.8	18	94.7
<i>BRAF</i>	Wild-type																			
	Mutant		9	14.1	0	0.0	0	0.0	9	60.0	1	5.6	0	0.0	1	14.3	2	8.3	1	5.3
<i>APC</i>	Wild-type		55	85.9	36	100.0	13	100.0	6	40.0	17	94.4	11	100.0	6	85.7	22	91.7	18	94.7
	Mutant		19	29.7	13	36.1	4	30.8	2	13.3	5	27.8	2	18.2	3	42.9	10	41.7	10	52.6
	Wild-type		45	70.3	23	63.9	9	69.2	13	86.7	13	72.2	9	81.8	4	57.1	14	58.2	9	47.4