Loss of the Polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors

Running title: Activation of bivalent genes in colorectal cancer

Maria A. Hahn,1,5 Arthur X. Li,2,5 Xiwei Wu,3 Richard Yang,1 David A. Drew,4 Daniel W. Rosenberg,4 and Gerd P. Pfeifer1

1Department of Cancer Biology, 2Department of Information Sciences, 3Department of Molecular Medicine, Beckman Research Institute, City of Hope, Duarte, CA, and 4Center for Molecular Medicine, Univ. of Connecticut Health Center, Farmington, CT.

5These authors contributed equally to this work

Correspondence:
Dr. Gerd P. Pfeifer
Department of Cancer Biology
Beckman Research Institute
City of Hope
Duarte, CA 91010
626-301-8853
gpfeifer@coh.org

Disclosure of Potential Conflicts of Interest:
Under a licensing agreement between City of Hope and Active Motif (Carlsbad, CA), the methylated-CpG island recovery assay (MIRA) technique was licensed to Active Motif, and the author G.P.P. is entitled to a share of the royalties received by City of Hope from sales of the licensed technology. G.P.P. is a paid consultant of Zymo Research (Irvine, CA).
Abstract

In colon tumors, the transcription of many genes becomes deregulated by poorly defined epigenetic mechanisms that have been studied mainly in established cell lines. In this study, we used frozen human colon tissues to analyze patterns of histone modification and DNA cytosine methylation in cancer and matched normal mucosa specimens. DNA methylation is strongly targeted to bivalent H3K4me3- and H3K27me3-associated promoters, which lose both histone marks and acquire DNA methylation. However, we found that loss of the Polycomb mark H3K27me3 from bivalent promoters was accompanied often by activation of genes associated with cancer progression, including numerous stem cell regulators, oncogenes and proliferation-associated genes. Indeed, we found many of these same genes were also activated in ulcerative colitis patients where chronic inflammation predisposes them to colon cancer. Based on our findings, we propose that a loss of Polycomb repression at bivalent genes combined with an ensuing selection for tumor-driving events plays a major role in cancer progression.
Introduction

Tumorigenesis is a complex process that is driven by a number of genetic and epigenetic alterations, which often result in aberrant gene expression (1-3). Mutations are generally considered to be the primary drivers of tumorigenesis (4, 5). However, dysregulation of epigenetic regulatory mechanisms also contributes to malignant transformation. Methylation of cytosine at CpG sequences in DNA leading to the formation of 5-methylcytosine (5mC) is one of the most stable and most widely studied epigenetic modifications (6). In most human tumors, widespread hypermethylation of CpG-rich sequences (CpG islands) is observed along with genome-wide DNA hypomethylation (1, 7-14). DNA hypermethylation in cancer is not a random event; it commonly affects specific gene classes and is seen most frequently at targets of the Polycomb repression complex (15-18) including numerous homeobox genes (19, 20). However, many of the genes marked by Polycomb in normal tissues and cell types, for example embryonic stem cells, are expressed at very low levels in these cells (21, 22). The acquisition of the more permanent silencing mark, 5mC, at the promoters of Polycomb target genes does not fundamentally change their expression levels although plasticity of expression will be reduced. For these reasons, it has remained unclear whether methylation of Polycomb target genes plays an essential role in tumor promotion and is in fact a tumor-driving event (23).

Recently, genome sequencing has identified a high frequency of mutations in epigenetic regulatory factors or chromatin structural elements in human malignancies (24, 25). However, technical limitations have made it difficult to directly examine chromatin changes in normal and malignant tissues, with most studies being limited to in vitro cell culture models. In order to determine the role of histone modifications in targeting DNA methylation and in altering gene expression patterns in human primary tumors, we have conducted the first comprehensive analysis of frozen tissues from colorectal cancer (CRC) patients focusing on two histone methylation marks, the activating histone H3 lysine 4 trimethylation (H3K4me3) and the repressive histone H3 lysine 27 trimethylation (H3K27me3), which are thought to be critical for gene regulation.
We found that genes carrying both histone modifications (bivalent genes) in normal tissue are characterized by substantial variability and undergo reorganization of these modifications in colorectal cancer leading to activation of cancer-promoting genes as one important outcome that can confer tumor-driving properties onto an emerging malignant cell population.

**Material and Methods**

**Human tissue samples**

Human Duke's stage II colon cancers and matching normal mucosa were obtained from the Cooperative Human Tissue Network (CHTN).

**Gene expression analysis**

Total RNA from patient samples was purified by using the mirVana kit (Ambion, Life Technologies). For whole genome expression analysis, GeneChip® Human Gene 2.0 ST arrays (Affymetrix) were used. Validation of gene expression changes by real-time reverse transcription-PCR was performed as described previously (18). All expression data were normalized to GAPDH in the same sample.

**Analysis of 5mC patterns**

For analysis of genomic 5mC patterns, the methylated CpG island recovery assay (MIRA) was used as described previously (26). After genome amplification, the methylated DNA fraction was hybridized versus input DNA on human CpG island/promoter microarrays (NimbleGen). The observed MIRA patterns for single CpG islands were validated by combined bisulfite restriction analysis (COBRA) as described previously (18). Primer sequences are available upon request.
Chromatin immunoprecipitation

For chromatin immunoprecipitation, frozen tissues were crushed with a plastic pestle in ice-cold phosphate-buffered saline (PBS) and fixed for 10 min at room temperature in 1% formaldehyde. Chromatin immunoprecipitation and genome amplification protocols were performed as described previously (26). The following antibodies were used: anti-H3K4me3 (39159, Active Motif), and anti-H3K27me3 (07-449, Millipore). For obtaining the H3K4me3 profile, H3K4me3 antibodies were pre-blocked with an H3K9me3 peptide (Abcam) to remove minor cross-reactivity of this antibody. After genome amplification, immunoprecipitated DNA was hybridized versus input DNA on human CpG island/promoter microarrays (NimbleGen).

For sequential chromatin immunoprecipitation (ReChIP), the first immunoprecipitation was performed as described above except for the elution step which was performed with SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, 1x cOmplete Protease Inhibitor Cocktail, Roche Applied Science) for 10 min at 68°C on a shaker at 1000 rpm. After removal of beads, the samples were diluted with 1:10 ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl, 22 μg/ml BSA, 1x cOmplete Protease Inhibitor Cocktail) and incubated with protein A/G PLUS-agarose (sc-2003, Santa Cruz Biotechnology), pre-blocked with BSA and herring sperm DNA, for 30 min at 4°C. After removal of beads, chromatin was incubated with anti-H3K4me3 (39159, Active Motif), anti-H3K27me3 (17-622, Millipore) or IgG (17-622, Millipore) antibodies overnight at 4°C followed by further incubation with protein A/G PLUS-agarose. Beads were washed once with low salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl, 1x cOmplete Protease Inhibitor Cocktail) and twice with high salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl, 1x cOmplete Protease Inhibitor Cocktail). Chromatin was eluted with SDS lysis buffer for 10 min at 68°C on a shaker at 1000 rpm and crosslinks were reversed in presence of 300 mM NaCl at
65°C overnight. Primer sequences for qPCR of individual genes are available upon request. Analysis was done with tissues from patient #14. Results are for triplicate experiments (+/- S.D.).

**Bioinformatics analysis**

All analysis was performed using R statistical language, except for gene ontology analysis, which was performed using DAVID annotation tools and Ingenuity Pathway Analysis. The heat-maps were generated with Cluster v3.0 and Java Treeview v2.0. Refseq genes were downloaded from the UCSC hg18 annotation database. Promoter was defined as -1.4 kb to + 0.6 kb relative to the transcription start site (TSS). For heat-maps with epigenetic marks, only genes with gene body length greater than 2 kb were included. Genes that were differentially expressed during colorectal cancer formation were identified as those with a log2 fold change >1. Probes were considered positive if their normalized log2 ratios were above 1.5-fold. Peaks in each sample were defined as four or more consecutive positive probes with either one or no gaps. A promoter was defined as bivalent if it contained overlapping H3K4me3 and H3K27me3 peaks at expanded promoter areas (-2.4 kb < TSS < 0.6 kb). For heat-map analysis, genes were assumed activated in all tumors and bivalent in mucosa if those genes had a promoter bivalent status in at least 3 out of 4 normal colon samples and tumor-associated gene activation of log2 > 0.4 in four patient sample pairs.

Affymetrix microarray data were processed by Expression Console v 2.0 with default settings. Transcription profiles of unique gene were obtained by averaging the transcriptional levels of all different transcript isoforms of the same gene according to Affymetrix’s annotation. Genes with log2 intensity < 5 in more than 10% of biopsies were considered uninformative and were removed. Affymetrix expression array data were used to generate heat-maps as described previously (27). All array data were deposited in the GEO database (accession number GSE47076).
Results

Chromatin mapping studies with frozen tissue

Previous correlative studies, mostly performed with cultured cells (e.g., embryonic stem cells), indicated that the Polycomb-associated chromatin modification, H3K27me3, predisposes sequences towards DNA hypermethylation in cancer (15-17, 20). In order to study aberrations of chromatin and DNA methylation in primary tumors, we directly profiled H3K4me3, H3K27me3 and DNA methylation in four archived frozen colorectal tumors with diagnosed Duke’s stage II and four matching normal mucosa samples, using human promoter and CpG island microarrays containing ~28,000 CpG islands and ~20,000 promoters.

We used the methylated-CpG island recovery assay (MIRA) (19) to comprehensively identify DNA methylation changes in the analyzed tissues. By using an established bioinformatics approach (28), we identified between 484 and 2,098 tumor-associated hypermethylated DNA regions in the four tumors (tumor #6: 2,098 regions, tumor #8: 1,552 regions, tumor #12: 484 regions and tumor #14: 1,676 regions). We validated the aberrant methylation patterns by combined bisulfite restriction analysis (COBRA) for several differentially methylated regions located in the TMEFF2, LIFR, CDKN2A, WNT5A and MLH1 promoters (Fig. S1). In total, we analyzed eleven regions by bisulfite-based methods and observed that the two approaches yielded excellent correlation for all regions examined.

To judge the quality of chromatin immunoprecipitation (ChIP) conducted on the frozen tissues, we performed several bioinformatics comparisons of our ChIP data with DNA methylation and gene expression data obtained with the presumably more stable DNA or RNA material obtained from the same samples. As shown in Figure S2, ChIP patterns from frozen tissues are very reproducible between samples and provide remarkable consistency among independent ChIP analyses within the same tissue (R=0.89, p<0.0001) (Fig. S2A). Our data revealed the well-established negative correlation (R=-0.6, p<0.0001) between 5mC and H3K4me3 at CpG islands (29, 30) (Fig. 1A). Furthermore, changes of
H3K4me3 at promoters occurring between tumors and matching normal tissues were strongly linked to transcriptional changes analyzed at the RNA level (Fig. 1C). Genes gaining H3K4me3 at promoters were frequently activated in tumors and genes, which lost H3K4me3, were repressed. Genome-wide, DNA methylation changes at promoters also showed a strong negative correlation with H3K4me3 changes between tumors and matching normal tissues (Fig. 1B). Taken together, these data confirmed the reliability of our chromatin mapping approach using frozen tissue specimens.

Variability of H3K27me3

Analysis of H3K27me3-marked genomic regions showed many similarities between tumor and nonmalignant matching specimens (e.g., Fig. S2D and S3A). However, in contrast to the generally more invariant H3K4me3 patterns, we observed a substantial degree of instability of the H3K27me3 mark when comparing normal and tumor samples (Fig. 2A,B; and Fig. S3). The Pearson correlation coefficients of H3K27me3 when comparing between tumors and matching normal mucosa were significantly reduced (approx. 0.75) in comparison to the Pearson correlation coefficients for H3K4me3 when comparing between matching tumor and normal samples (approx. 0.91) (Fig. S3A). Inspection of H3K27me3 patterns revealed that the H3K27me3 mark was frequently lost along larger genomic areas such as the HOX and histone gene clusters in tumors (Fig. 2B and Fig. S3B), similar as has been reported recently by Bert et al. (10). However, these long-range areas of H3K27me3 loss represented only a relatively small fraction of all sequences that lost H3K27me3.

Quite surprisingly, a detailed examination of the chromatin profiles also revealed that loss of the H3K27me3 mark over larger sequence blocks can be alternatively associated with accumulation of aberrant DNA methylation at some promoters or with accumulation of the H3K4me3 mark at other (even nearby) promoters (Fig. 2B and S3B). Moreover, even a switch between transcriptional isoforms can occur in tumors when alternative routes between H3K4me3 and 5mC deposition are used upon H3K27me3 loss. For example, a loss of
H3K27me3 at the HOXC cluster in sample #6 is associated with aberrant DNA methylation at the short HOXC9 isoform promoter, at the HOXC10 promoter and at the HOXC4 short isoform (transcript variant 2) promoter. The same genomic locus shows accumulation of H3K4me3 at the HOXC5 promoter (transcript variant 1), HOXC6 promoter (transcript variant 1), HOTAIR long isoform (transcript variant 1) promoter, the HOXC4 long isoform (transcript variant 1) promoter and at the long HOXC9 isoform promoter (Fig. 2B). Thus, a loss of H3K27me3 at distinct promoters can be involved in H3K4me3 accumulation or aberrant DNA methylation and can have a major impact on transcript isoform usage in colorectal tumors. In order to evaluate if these observations apply to the cancer epigenome as a whole, and to determine whether alterations of the H3K27me3 patterns reflect transcriptional changes, we used the heat-map approach for promoter regions and sorted genes by H3K27me3 changes (Fig. 2C,D). We examined the 15% of genes that showed the greatest loss of H3K27me3 at promoters, and we additionally sorted these genes by H3K4me3 changes (Fig. 2C,D). We found that H3K4me3 alterations within this group of genes negatively correlated with accumulation of aberrant DNA methylation at promoters. At the same time, accumulation of H3K4me3 at promoters, which lost H3K27me3 in tumors, was frequently associated with gene activation (Fig. 2D). Thus, at a genome-wide level, we confirmed the two different correlates of H3K27me3 loss in colorectal cancer, gain of 5mC, or gene activation.

**Bivalent chromatin in ES cells and colonic mucosa**

Bivalent (H3K4me3 and H3K27me3 containing) chromatin status is a key regulatory mechanism during embryonic stem (ES) cell maintenance and differentiation (31, 32). We first asked to what extent bivalent promoter status is retained in human colonic mucosa versus ES cells. We found that the number of bivalent promoters varies between 3,414 and 3,967 within the normal colonic mucosal samples (Fig. S4A). Approximately 50-60% of these bivalent promoters in normal colon are also bivalent in ES cells, whereas 40-50% of bivalent promoters are specific for colon. Gene ontology analysis using DAVID
(http://david.abcc.ncifcrf.gov/) showed that the gene group retaining the bivalent state from ES cells to colon is more enriched with developmental genes in comparison to the gene groups that have lost the bivalent state during development, and those, which perhaps obtained it de novo in the colon (Fig. S4B). Previously, aberrant DNA methylation in cancer has been linked to the bivalent state of promoters (15). In order to further assess this correlation, we compared the bivalent promoter state in normal tissue to aberrant methylation in matching tumors (Fig. S5A) and observed that genes with bivalent promoters comprise the vast majority of promoters with cancer-associated DNA methylation. Strikingly, the presence of H3K27me3 at promoters in normal colon epithelial tissue was associated with 90% of all hypermethylated sites (Fig. S5B) and approximately 70% of aberrantly methylated promoters had a bivalent state in normal colon tissue (Fig. S5A).

Since the simultaneous presence of H3K4me3 and H3K27me3 at promoters might reflect different cell populations, we examined chromatin state at several candidate genes by sequential immunoprecipitation (Fig. 3A and Fig. S4C). These experiments confirmed presence of both histone marks on the same immunoprecipitated DNA fragments in normal mucosa.

**Chromatin changes at bivalent promoters**

We next evaluated all epigenetic changes that target bivalent promoters in colorectal cancer. We generated a heat map containing information for H3K27me3, H3K4me3 and 5mC changes occurring at promoters with bivalent status in all analyzed normal colonic mucosa samples (Fig. 3B; Table S1). We found that the majority of bivalent promoters undergo H3K27me3 changes at least in one of four analyzed tumors, pointing to a high instability or variability of bivalent promoters in colorectal cancer. Detailed analysis revealed that some genes such as *DMRTA2* and *HOXD11* lose H3K27me3 in tumors in comparison to normal tissue and may undergo either gene activation associated with accumulation of H3K4me3 or repression associated with aberrant DNA methylation in different individual patients (Fig. S3C). This fact may indicate a
random nature of the consequences of Polycomb loss at promoters, at least for certain genes. Clustering analysis indicated a large group of promoters characterized by a loss of H3K27me3, loss of H3K4me3 and gain of 5mC in all tumors (Fig. 3B; promoter groups marked by blue bars; Fig. S5C). Further, we observed a group of genes, which are characterized by a loss of H3K4me3 in all tumors and accumulation of 5mC in some tumors but accumulation of H3K27me3 in other tumors (Fig. 3B; promoter groups marked with black bars) providing alternative routes towards gene silencing (33). Both types of alterations, accumulation of H3K27me3 or of 5mC, are indeed characterized by gene repression (Fig. 3C). This observation suggests that the variability of chromatin at bivalent promoters can lead to two opposite outcomes, a gain or a loss of H3K27me3, and may result in gene repression according to both scenarios if 5mC is gained in the latter. A majority of promoters with aberrant DNA methylation in tumors was associated with a loss of H3K27me3 (Fig. S5D). This phenomenon, which has been referred to as 'Polycomb switching' (34), was true for promoters with and without bivalent state (Fig. S5C,D).

**Gene activation at bivalent promoters in cancer**

Our study identified a novel and distinct group of promoters where a loss of H3K27me3 is associated with accumulation or retention of H3K4me3 in tumors but generally little or no change in DNA methylation levels (Fig. 3B; promoter group marked by a green bar). These promoters were associated with strong gene activation (Fig. 3C). Analysis of the transcriptome of bivalent promoters revealed a very high transcriptional instability of bivalent genes in tumors. Approximately 26-46% of bivalent promoters were affected by transcriptional alterations (at least two-fold) in each tumor (Fig. 3D). We found that 15-20% of the bivalent genes undergo transcriptional activation in tumors (Fig. 3D).

In order to further elucidate the potential functional impact of bivalent promoter activation on colorectal cancer progression, we performed a detailed analysis of genes that are bivalent or non-bivalent in colonic mucosa and become activated at least two-fold in at least two out of four analyzed tumors
Non-bivalent genes activated in tumors were enriched for genes important in mitotic progression including mitotic kinases, mitotic cyclins and kinesins and checkpoint proteins (Fig. S6A; Table S2). The activated bivalent genes were strongly associated with the colonic transcriptome and transcriptional activity in colorectal cancer (Fig 4A). Importantly, this group of genes was strongly enriched with genes involved in transcriptional regulation and associated with early development (Fig. S6B; Table S2). The genes that become repressed in tumors and carry a bivalent status in mucosa were strongly linked to brain development (Fig. S6C) whereas repressed non-bivalent genes did not show any functional categories with major enrichment (data not shown). Significantly, the non-bivalent genes did not show any enrichment for transcription factors (Table S2), neither for activation nor for repression gene groups, suggesting that the enrichment of mitotic proteins in the non-bivalent group could be downstream of crucial regulators (e.g. cyclin D) encoded by the activated bivalent genes.

According to Ingenuity pathway analysis (www.ingenuity.com), the group of genes, which become activated in cancer and are bivalent in mucosa, was strongly enriched with genes associated with cancer as a disease (Fig. 4B) and with important aspects of the transformed phenotype (Table S3). For example, one of the main regulators of the epithelial to mesenchymal transition (EMT), SNAI2 was activated together with a loss of bivalent status in tumor tissues (Table S3). This group of genes contains many genes encoding transcription factors and genes associated with cellular growth and proliferation such as the oncogene cyclin D1 (CCND1), ribonucleotide reductase (RRM2) and MKi67, genes associated with cell adhesion and invasion including claudin1 (CLDN1) and EPCAM and other genes strongly implicated in cancer such as COX2/PTGS2 and the MET oncogene (Fig. 4D; Table 1; Table S3; Fig. S6D).

Three crypt stem cell markers, ASCL2, LGR5 and SOX9 (35-37) belong to the group of most frequently activated genes (in all tumors) with bivalent promoters in normal colonic epithelium (Fig. 4C,D; Table 1; Table S3). This fact was also confirmed for LGR5 in 11 sample pairs by Q-RT-PCR (Fig. 4C). Analysis of transcription factor binding sites by DAVID for bivalent tumor-
activated genes showed enrichment for SOX9 and FOXQ1 binding sites (Fig. S6E). Interestingly, the genes coding for both transcription factors were activated due to loss of H3K27me3 in all tumors. In addition, two other intestinal stem cell signature genes, OLFM4 and EPHB3 were activated together with a loss of the H3K27me3 mark in CRC (Fig. 4D). It is remarkable that the well-known proliferation marker, MKI67 (Ki67) and cancer stem cell markers such as Prominin 1 (CD133) and ALCAM (CD166) also became activated together with a loss of bivalent state in tumors (Table 1, Table S3). These data suggest that loss of the bivalent state accompanied by activation of critical cell cycle drivers and tumor-promoting genes is a general and pervasive mechanism in colorectal cancer.

In order to evaluate the possibility that the apparent loss of H3K27me3 in tumors reflects an intestinal stem cell origin and presence of H3K27me3 in mucosa reflects a differentiation process in which this mark is acquired at promoters of stem cell genes, we determined if cancer-associated transcriptional changes in tumors carry primarily a gene expression signature indicative of cells at the crypt base where such stem cells reside. We used previously published expression profiling of human crypt top versus bottom samples (38). From 3157 crypt bottom-specific genes, of which 306 genes carry a bivalent status in normal mucosa, we found only 26 crypt bottom genes including LGR5, ASCL2, MKi67 and CLDN1 to be activated in all tumors and bivalent in mucosa. At the same time, very few genes (ten) out of 2865 crypt top-specific genes, of which 251 genes carry bivalent character in mucosa, were repressed in all tumors and were bivalent in mucosa (Fig. S6F). These data clearly suggest that the tumors we analyzed do not express a predominant stem cell-like phenotype. Analysis of genes, which are not associated with crypt top or bottom according to Kosinski and colleagues (38), have bivalent status in mucosa and are activated in all tumors, revealed a strong association of these genes with cancer and epithelial neoplasia (Table S4). This fact suggests that this group of genes may play an important role in cancer development.
We then tested if activation of bivalent genes may occur at an early stage of colorectal cancer progression. Based on publically available gene expression profiles of 32 human adenomas and matching mucosa (39), we evaluated if bivalent genes, which become activated in all tumors, undergo transcriptional changes in colorectal adenomas. Bivalent genes were activated slightly preferentially over non-bivalent genes in adenomas (18% versus 15%). However, the majority (54%) of bivalent genes activated in stage II CRC were also activated in adenoma. Among the activated genes were stem cell markers including LGR5, SOX9 and ASCL2 and proliferation-associated genes such as MKI67 and RRM2 (Fig. 4E).

Transcriptional alterations of bivalent genes in inflamed mucosa
Chronic inflammation is a strong risk factor for cancer development (40). Similar to colon cancer, chronic inflammation of the digestive tract is associated with aberrant DNA methylation (41, 42), which frequently coincides with a loss of the Polycomb mark, H3K27me3 (18). We determined if activation of H3K27me3-marked genes occurs during inflammation and if these activated genes play a role in cancer. We used publically available expression profiles for 15 inflamed mucosal samples from ulcerative colitis (UC) patients, seven unaffected mucosal tissues from UC patients and 13 human colon biopsies from non-inflamed controls obtained from GEO dataset GSE38713 (43). Analysis of expression profiles for genes with bivalent status in mucosa revealed that almost 20% of bivalent genes already undergo transcriptional changes during inflammation (Fig. 5A and Table S1). Bivalent genes were not activated preferentially over non-bivalent genes (11% versus 10%, respectively). Remarkably, patterns of activation and repression of bivalent genes strongly divided biopsies with and without inflammation (Fig. 5B). According to Ingenuity pathway analysis, these activated bivalent genes are strongly linked to cancer (Fig. 5C, Fig. S6G) and are associated with functions associated with cancer development such as cellular movement, cell death and survival, and proliferation (Fig. S6H). Moreover, approximately 40% of genes (31/80) that are activated in all colorectal tumors
and have bivalent status in adjacent mucosa are activated in inflamed tissues from patients with ulcerative colitis (Fig. 5D). Among these genes, we found Foxq1, CLDN1, LEF1 and MKI67. However, in contrast to adenoma, expression of stem cell markers including SOX9, LGR5 and ASCL2 was not activated in inflamed tissues. This data provides additional evidence for bivalent gene activation in differentiated mucosa, in the absence of overt tumor formation, and demonstrates that the phenomenon of bivalent promoter activation occurs already during chronic inflammation, a condition strongly correlated with cancer predisposition.

Discussion

Our data suggest that loss of the bivalent (H3K4me3 and H3K27me3) chromatin state at promoters is accompanied by either gain of 5mC (when both marks are lost) or by gene activation (when H3K27me3 is lost) (Fig. 6). The latter event, as we demonstrate here for the first time, is likely to be a critical step in cancer pathogenesis responsible for the activation of crucial genes leading to progression to malignancy and metastasis. Among the activated bivalent genes were several key components of the WNT signaling pathway including LGR5, CD133, LEF1, TCF7, WNT2, and WNT3. This pathway is commonly found to have a high level of constitutive activity in colorectal cancer. Also, many cell cycle drivers and proliferation-associated genes were found in this category, including cyclin D, MKI67, RRM2, ETV4 and FGF19 along with many transcription factor genes such as SOX9, TP73, MYB, FOXA2, ETV4, and TEAD4, a transcription factor negatively regulated by the Hippo tumor suppressor pathway (Table S3). It is easily understandable that selection for activation of these sets of genes would provide a growth advantage to the tumor cell population. On the other hand, the DNA methylation pathway, if operative at bivalent promoters after loss of H3K27me3, may not immediately provide such a selective advantage because the affected genes are already expressed at a very low level while in the bivalent state in normal cells.
Two non-exclusive models of colorectal cancer initiation have been discussed. According to mouse models, it has been suggested that Lgr5+ cells from the crypt bottom comprise the tumor-initiating cell population (44). Other models favor a ‘top-down’ model, in which dysplastic cells can originate in the upper areas of crypts (45, 46). The latter model invokes a dedifferentiation process in which the cells re-acquire stem cell-like properties. Our data are consistent with a model in which colorectal cancer can arise from differentiated epithelial cells in which the bivalent chromatin state resolves into active or inactive forms. It is likely that the instability of bivalent genes is a mechanism that predisposes the colonic epithelium to cancer by being operative already during inflammatory processes. Elevated NF-κB signaling, as found in chronic inflammation, has been shown to enhance Wnt activation and to induce dedifferentiation of non-stem cells in the colon that acquire tumor-initiating capacity (45). Inflammation has been linked to methylation of Polycomb-marked genes in the intestinal epithelium of mice (18). We show here that activation of cancer-relevant genes occurs at bivalent promoters in inflamed differentiated mucosa from ulcerative colitis patients.

Besides inflammation, aging is a major risk factor for cancer. Similar to inflammatory conditions, the aging process has been associated with DNA hypermethylation of Polycomb target genes (47, 48). If the loss of the Polycomb mark underlies both DNA methylation silencing and the activation of bivalent genes in cancer cells, a mechanistic connection between inflammation, aging and cancer can be proposed, in which gene activation due to H3K27me3 loss may easily be a dominant driving force for the malignant change.

Questions remain as to the possible mechanisms underlying the high variability of the Polycomb mark. Earlier studies showed that the H3K27 methyltransferase EZH2 is downregulated in stressed and senescing populations of cells, which coincided with decreased levels of H3K27me3 at the INK4A-ARF locus (49). However, it has often been observed that EZH2 is overexpressed in human tumors (50, 51). In our samples, we observed that EZH2 was expressed at higher levels in tumors than in normal mucosa but we did not find substantial
differences in expression of the H3K27me3 demethylases *KDM6A* or *KDM6B*. It remains to be determined whether structural or functional aberrations of the Polycomb complex itself underlie the cancer-associated dysfunction of Polycomb marking at bivalent promoters or if bivalency of these loci itself predisposes them to inherent or transcription factor-driven variability and loss of the H3K27me3 mark resulting in activation of genes with functional importance for tumor progression.

**Disclosure of Potential Conflicts of Interest**

Under a licensing agreement between City of Hope and Active Motif (Carlsbad, CA), the methylated-CpG island recovery assay (MIRA) technique was licensed to Active Motif, and the author G.P.P. is entitled to a share of the royalties received by City of Hope from sales of the licensed technology. G.P.P. is a paid consultant of Zymo Research (Irvine, CA). The other authors declare no potential conflict of interest.

**Grant Support**

This work has been supported by NIH grant CA 084469 to G.P.P.
References


Figure legends

Figure 1. Verification of chromatin mapping with frozen tissues.
A. DNA methylation analysis and H3K4me3 histone mapping was conducted with normal (N) and tumor (T) samples from patients #6 (top) and #8 (bottom). An inverse correlation between 5mC and H3K4me3 signal is observed for each sample.

B. Changes in 5mC signal between tumor (T) and normal (N) samples are plotted versus changes in H3K4me3 signal for patients #6 and #8. An inverse correlation is observed. Analysis was performed for promoters overlapping with CpG islands. The Pearson correlations between 5mC and H3K4me3 data sets are indicated.

C. A heat-map analysis of changes (T versus N) in H3K4me3, 5mC, and gene expression levels is shown for four patient sample pairs. H3K4me3 changes positively correlate with changes in gene expression but negatively correlate with changes in 5mC. Promoters of genes longer than 2 kb were sorted by cancer-associated H3K4me3 changes. Each row represents a gene promoter. Red color indicates a gain and green represents a loss.

Figure 2. Frequent variability of the Polycomb mark H3K27me3 in colon tumors.
A. Representative snap-shots showing loss of H3K27me3 at the HOXB13 locus in three of four analyzed tumors (T) compared to matching normal mucosa (N) samples. The snap-shots indicate log2 ratios of signal of IP versus input. Transcriptional start sites and transcript directions are shown.

B. Loss of H3K27me3 in tumors is linked to aberrant DNA methylation or accumulation of H3K4me3 at promoters. Representative snap-shots of the HOXC cluster from patient #6 are shown. The colored circles mark changes of H3K4me3 or 5mC levels in tumor (T) versus normal (N) samples. Transcriptional start sites and transcript directions are shown.
C. Sorting of the genome by cancer-associated H3K27me3 changes at promoters. Each row represents a gene promoter. Values for each epigenetic mark were calculated as a mean of signal intensity log2 ratio for probes located at promoters. Red color indicates a gain, and green represents a loss. For analysis, we used genes larger than 2 kb and with positive values of H3K27me3 in at least one of the tissues (T or N).

D. Loss of H3K27me3 in tumors is frequently associated either with a gain of H3K4me3 and gene activation or with aberrant DNA methylation and gene silencing. The top 15% of genes with greatest loss of H3K27me3 from the sorting in panel C were sorted by H3K4me3 changes. Each row represents a gene promoter. Values for each epigenetic mark were calculated as a mean of signal intensity log2 ratio for probes located at promoters. Red color indicates a gain, and green represents a loss of the respective mark.

Figure 3. Patterns of instability of bivalent promoters in colorectal cancer.

A. Chromatin immunoprecipitation (ChIP) and sequential ChIP (ReChIP) of the FOXQ1 promoter for matched tumor and normal samples. Standard ChIP with anti-H3K4me3 and anti-H3K27me3 antibodies is shown in the left panel and sequential ChIP with both antibody combinations is shown in the right panel.

B. Clustering analysis of epigenetic changes at bivalent promoters in colon tumors. Analysis was done for promoters with bivalent status in all analyzed normal tissue samples. Each vertical row represents a gene promoter. Values for each epigenetic mark were calculated as a mean of signal intensity log2 ratio for probes located at promoters. Red color indicates a gain, and green represents a loss.

C. Transcriptional changes of bivalent promoters, which undergo similar epigenetic changes during colorectal cancer progression. The box-plots represent transcriptional levels for three groups of genes with bivalent promoter status in all analyzed normal tissues: genes with a gain H3K27me3 at promoters in tumors, genes which loose H3K27me3 at promoters and acquire cancer-
associated DNA methylation and genes which loose H3K27me3 and gain H3K4me3 at promoters. The data are merged for four patient samples.

**D.** Activation or repression of genes with bivalent promoters in normal mucosa. The graph indicates the fraction of bivalent genes, which become activated or repressed at least two-fold in tumor versus matching mucosa.

**Figure 4. Gene activation associated with bivalent promoters in colon tumors affects cancer-promoting and stem cell genes.**

**A.** Genes that become activated in colorectal cancer and carry bivalent promoter status in matching mucosa are strongly linked to colon and colorectal tumor transcriptomes. The top ten transcriptomes are indicated. Analysis was done for genes, which are at least two-fold activated in at least two out four tumors and are bivalent in matching mucosa. We performed data analysis by using the Unigene tissue specificity database (http://www.ncbi.nlm.nih.gov/uniGene) incorporated into DAVID (http://david.abcc.ncifcrf.gov/).

**B.** Top five diseases and disorders (www.ingenuity.com) associated with genes activated at least two-fold in two out four tumors and having bivalent promoter status in matching mucosa. The numbers of genes in each subgroup are indicated.

**C.** Activation of LGR5 in colorectal cancer. Transcript levels of LGR5 in colon tumors and matching mucosa were validated by Q-RT-PCR with further normalization to GAPDH.

**D.** Loss of H3K27me3 affects stem cell marker genes and genes strongly linked to colorectal cancer. Representative snap shots for patient #8 displaying epigenetic changes associated with gene activation in colon tumors are shown.

**E.** Bivalent genes are activated in adenomas. The heat-map displays transcriptional changes in colorectal adenomas for bivalent genes, which are activated in colorectal tumors and are also activated at least 2-fold in at least one third of the adenomas. Expression data for 32 human adenomas and matching mucosa were obtained from GEO data set GDS2947. For clustering analysis, we
used only genes with bivalent state in at least 3 out 4 mucosa tissues and activated in all four analyzed colorectal cancers. Gene symbols are indicated.

Figure 5. Activation of genes with bivalent promoters in ulcerative colitis.  
A. Transcriptional changes in ulcerative colitis for genes with bivalent promoter status in mucosa. Expression profiles for 13 human colon biopsies from non-inflamed controls, 15 inflamed mucosa samples from ulcerative colitis patients and 7 unaffected mucosa tissues from ulcerative colitis patients were obtained from GEO dataset GSE38713. Bivalent genes were divided into three groups: genes which become activated at least 2-fold in at least one third of inflamed tissues versus non-inflamed controls, genes which became repressed at least 2-fold, and genes which were not activated or repressed.  
B. Differentially expressed bivalent genes in inflamed tissues from patients with ulcerative colitis. Clustering analysis was performed for bivalent genes, which underwent at least 2-fold transcriptional activation or repression in at least one third of inflamed tissues from ulcerative colitis patients versus non-inflamed controls. For this analysis, we used genes with bivalent state in at least 3 out 4 mucosal tissues.  
C. Top five diseases and disorders (www.ingenuity.com) associated with genes activated at least two-fold in inflamed biopsies versus non-inflamed controls and having bivalent promoter status in all four analyzed mucosal tissues. The numbers of genes in each subgroup are indicated.  
D. Bivalent genes, which become activated in colorectal cancer and are activated also in inflamed tissues from ulcerative colitis patients. Clustering analysis is shown for bivalent genes, which underwent at least 2-fold transcriptional changes in at least one third of inflamed ulcerative colitis tissues and were activated in all tumors. Clustering analysis was done for genes that carry a bivalent promoter status in 3 out of 4 mucosal samples.
Figure 6. Fate of bivalent promoters in colorectal cancer.
The model depicts the state of promoters containing H3K4me3 (green) and H3K27me3 (red) in normal colonic mucosa and how this state has resolved in cancer cells. The disappearance of the H3K27me3 mark can be associated with DNA methylation and permanent repression of genes (bottom), or, alternatively, with activation of growth promoting genes (top).
Table 1: Bivalent genes activated in colorectal cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>Classification</th>
<th>Activation</th>
<th>Wnt – pathway target</th>
<th>Activation in ulcerative colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGR5*</td>
<td>Leucine-rich repeat containing G protein-coupled receptor 5</td>
<td>Crypt stem cell marker</td>
<td>Colon cancer stem cells, crypt stem cells</td>
<td>Yes</td>
</tr>
<tr>
<td>SOX9*</td>
<td>SRY (sex determining region Y)-box 9</td>
<td>Crypt stem cell marker, Transcription factor</td>
<td>Crypt stem cells, colon cancer</td>
<td>Yes</td>
</tr>
<tr>
<td>ASCL2*</td>
<td>Achaete-scute complex homolog 2</td>
<td>Crypt stem cell marker, Transcription factor</td>
<td>Crypt stem cells, colon cancer</td>
<td>Yes</td>
</tr>
<tr>
<td>CD133 (PROM1), Prominin1</td>
<td>Stem cell marker</td>
<td>Colon cancer stem cells, crypt stem cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXQ1*</td>
<td>Forkhead box Q1</td>
<td>Transcription factor</td>
<td>Colorectal cancer</td>
<td>Yes Yes</td>
</tr>
<tr>
<td>COX2 (PTGS2), Prostaglandin-endoperoxide synthase 2</td>
<td>Inflammation and cancer-associated cyclooxygenase-2</td>
<td>Colon cancer</td>
<td>Yes Yes</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>Met proto-oncogene (hepatocyte growth factor receptor)</td>
<td>Oncogene, hepatocyte growth factor receptor</td>
<td>Colon cancer, cancer stem cells, stem cells</td>
<td>Yes</td>
</tr>
<tr>
<td>CCND1, Cyclin D1</td>
<td>Cyclin, oncogene</td>
<td>Colon cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLDN1*</td>
<td>Claudin1</td>
<td>Tight junction molecule</td>
<td>Colon cancer</td>
<td>Yes Yes</td>
</tr>
<tr>
<td>EPCAM, Epithelial cell adhesion molecule</td>
<td>Adhesion molecule</td>
<td>Colon cancer, cancer stem cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MKI67, Antigen identified by monoclonal antibody Ki-67</td>
<td>General proliferation marker</td>
<td>Proliferating cells</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

* activated in all 4 tumors.
Figure 1
Figure 3
Figure 6
Loss of the Polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors

Maria A Hahn, Arthur X Li, Xiwei Wu, et al.

Cancer Res  Published OnlineFirst May 1, 2014.

Updated version
Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-13-3147

Supplementary Material
Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/05/01/0008-5472.CAN-13-3147.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.