Defining a Chromatin Pattern that Characterizes DNA-Hypermethylated Genes in Colon Cancer Cells

Kelly M. McGarvey, Leander Van Neste, Leslie Cope, Joyce E. Ohm, James G. Herman, Wim Van Criekinge, Kornel E. Schuebel, and Stephen B. Baylin

Divisions of Cancer Biology and Biometry and Clinical Trials, The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland; Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium; and OncoMethylome Sciences, SA, Lierge, Belgium

Abstract

Epigenetic gene regulation is a key determinant of heritable gene expression patterns and is critical for normal cellular function. Dysregulation of epigenetic transcriptional control is a fundamental feature of cancer, particularly manifesting as increased promoter DNA methylation with associated aberrant gene silencing, which plays a significant role in tumor progression. We now globally map key chromatin parameters for genes with promoter CpG island DNA hypermethylation in colon cancer cells by combining microarray gene expression analyses with chromatin immunoprecipitation-on-chip technology. We first show that the silent state of such genes universally correlates with a broad distribution of a low but distinct level of the PcG-mediated histone modification, methylation of lysine 27 of histone 3 (H3K27me3), and a very low level of the active mark H3K4me2. This chromatin pattern, and particularly H3K4me2 levels, crisply separates DNA-hypermethylated genes from those where histone deacetylation is responsible for transcriptional silencing. Moreover, the chromatin pattern can markedly enhance identification of truly silent and DNA-hypermethylated genes. We additionally find that when DNA-hypermethylated genes are demethylated and reexpressed, they adopt a bivalent chromatin pattern, which is associated with the poised gene expression state of a large group of embryonic stem cell genes and is characterized by an increase in levels of both the H3K27me3 and H3K4me2 marks. Our data have great relevance for the increasing interest in reexpression of DNA-hypermethylated genes for the treatment of cancer. [Cancer Res 2008;68(14):5753–9]

Introduction

DNA sequence is the basic genetic building block of the genome. However, epigenetic gene regulation is critically important for normal cellular function, and dysregulation of epigenetic transcriptional control is often associated with disease. Aberrant promoter DNA hypermethylation is a critical event in the silencing of tumor suppressor genes in virtually all types of human cancer and is a frequent alternative to genetic loss of tumor suppressor gene function (1). It is of great concern to know the chromatin effects of DNA demethylation because there is a promise of successful use of DNA-demethylating agents as cancer treatment (2). Using colorectal cancer HCT116 cells that have both DNMT1 and DNMT3b genetically knocked out (double-knockout cells; ref. 3), we are able to examine histone modification changes in a system with profoundly reduced DNA methyltransferase activity.

We have previously shown in local promoter chromatin immunoprecipitation (ChIP) studies that several gene promoters retain many modifications that are typically associated with repression of transcription on DNA demethylation–induced gene expression in double-knockout cells (4). These genes, now actively transcribed, exist in what we have termed a semi-heterochromatic state (4). This state shares similarities with what others have described as bivalent promoter domains that contain both the active H3K4me2 mark and the repressive H3K27me3 mark (5). This chromatin status is thought to maintain a set of embryonic genes, which contain promoters with CpG islands, in a poised, low transcription state, which facilitates prevention of premature lineage commitment (5).

In the present study, we have sought to more globally establish whether what we have learned from the above chromatin analyses of subsets of colon cancer genes may define key features of the cancer epigenome. We have thus performed a promoter genomewide localization, by ChIP-chip analyses, of the above key active (H3K4me2) and repressive (H3K27me3) marks in wild-type HCT116 colorectal cancer cells compared with double-knockout cells lacking DNA methylation. In this study, we are able to orient our global results around control genes from the previous local ChIP studies (4) mentioned above. Moreover, we have mapped our ChIP-chip results to previous microarray expression results used in an approach to randomly scan colon cancer cells for CpG island DNA-hypermethylated genes (6). This has allowed us to correlate the chromatin status of a large number of verified DNA-hypermethylated genes in HCT116 and double-knockout cells for which both reverse transcription-PCR (RT-PCR) and microarray expression analyses had previously been matched (6) to globally determine the chromatin signature of CpG island–containing genes before and after removal of DNA hypermethylation.

Materials and Methods

Cell culture. HCT116 and double-knockout cells were maintained in McCoy's 5A modified medium supplemented with 10% fetal bovine serum (Gemini Bio-Products) and 1% penicillin/streptomycin (Invitrogen) and grown at 37°C in 5% CO₂ atmosphere.

ChIP-chip. ChIP was combined with DNA microarray analysis results (6) on HCT116 colorectal cancer cells and double-knockout cells in duplicate using antibodies specific for H3K27me3 (7), H3K4me2 (Upstate), or an IgG control (Upstate). Using 5 × 10⁶ cells per experiment, ChIP, DNA amplification, labeling, and array hybridization were done as previously described (8). Independent batches of HCT116 and double-knockout cells
were used to perform independent ChIP experiments, and ChIPs for each antibody were hybridized to independent Agilent array sets. Whole genome expanded promoter 244K arrays from Agilent Technologies that span roughly 5 kb upstream to 2 kb downstream from gene transcription start sites were used.

Analyses. The data preprocessing for the tiling arrays was done according to the manufacturer's recommendations using the Agilent ChIP Analytics 1.3 software (i.e., a median blanks subtraction followed by an interarray median normalization and a dye-bias median normalization). The data were imported into R for downstream analysis. The expression arrays were preprocessed as previously described (6).

The expression data were analyzed as described in ref. 6, but some modifications were made to determine the basal expression of genes. Expression values were calculated using an Agilent whole human genome microarray with the double-knockout sample labeled with Cy5 and the original HCT116 sample labeled with Cy3. Expression status was defined as the natural logarithm of the normalized single channel signal (i.e., the Cy5 channel for double-knockout cells and the Cy3 channel for the HCT116 cells). The upper boundary of the silent expression range was chosen so that 90% (38-42) of previously verified DNA-hypermethylated genes (6) were contained within this zone. This cutoff was also chosen because it clearly separates the general expression trends of the verified genes from a set of false-positive genes (6) as shown in Supplementary Fig. S1.

Composite graphs were created to reflect the general chromatin trends of a particular group of genes. To retain maximal data quality, only genes that had a minimal amount of probes within the promoter-associated region around the transcription start site were allowed to contribute data points to the composite profile. A sliding window of 100 bp was normally used to compute the average H3K27me3 and H3K4me2 profiles for both HCT116 and double-knockout cells. An extended window size was used in case the number of input genes was lower than 50 to smooth the profiles and give a better view of the general trends. The profiles made with an extended window size were compared with the standard 100-bp window profiles to ensure that only the visible noise was reduced and nothing about the trends or gene characteristics was changed. Finally, the mean H3K27me3 enrichment in a broad region from −2,250 to +2,250 relative to the TSS was determined for all genes on the tiling array and was used to visualize the general H3K27me3 trends.

Results and Discussion

We first markedly refined our Agilent 44K stratification of gene expression levels, as compared with our recent study (6), allowing us to define a very low range of basal expression for a group of 42 genes confirmed to have (a) fully absent expression by RT-PCR in association with CpG island DNA hypermethylation in HCT116 colorectal cancer cells (6) and (b) presence of expression by RT-PCR in association with loss of virtually all promoter DNA methylation in double-knockout cells (Fig. 1A; gene list—Supplementary Table S1). In addition, this change in expression levels between HCT116 and double-knockout cells was appreciated in the microarray expression analyses through which the hypermethylation status of the genes was discovered (6). The genes have a very low expression range in the microarrays in the HCT116 cells, which encompasses 90% (38 of 42) of verified hypermethylated genes (Fig. 1A). To place the chromatin state of these genes into a global perspective, we began by using the ChIP-chip approach to examine more than 4,500 genes with CpG island–containing promoters, which had a much higher range of signal for expression on the microarrays (‘actve genes’ in Fig. 1A; gene list will be included online). For both single examples of such active genes (CDKN2B in Fig. 1B and eight other genes in Supplementary Fig. S2), and for the average enrichment at all active CpG island gene promoters (Fig. 1C), we found, as in studies of others (9–13), that the active H3K4me2 mark is highly enriched in a 1- to 2-kb region centered around the transcription start site. In addition, as previously reported (15), we observed the sharp dip for this mark directly over the transcription start site (Fig. 1B and C; Supplementary Fig. S2) in most of the genes, which marks the presence of a nucleosome-free region at actively transcribed promoters (14–17). Concomitantly, we observe a virtual lack of the repressive H3K27me3 mark throughout a 7- to 8-kb region spanning the entire proximal promoter region with values generally of 0 to 1 enrichment especially toward the annotated transcriptional start site (Fig. 1B and C; Supplementary Fig. S2).

In marked contrast to the above pattern for the active genes, we observe starkly reduced levels of H3K4me2, now with only a minimal peak directly over the transcription start site of all previously verified (4, 6) DNA-hypermethylated and fully silent genes (individual examples, GATA4 in Fig. 1D and eight other genes in Supplementary Figs. S3 and S4, and the average distribution for all the genes in Fig. 1E). Also in contrast to the active genes, the chromatin of the 42 verified DNA-hypermethylated genes contained a low but broad level (general enrichment of 1.0 and ranging to 5.0) of H3K27me3 enrichment over the entire promoter region (Fig. 1D and E; Supplementary Figs. S3 and S4). These results match well with our previous studies linking, in local ChIP analyses (matches of such local ChIP data with patterns of ChIP-chip results for five individual genes shown in Supplementary Fig. S3), the association of H3K27me3 enrichment with CpG island DNA-hypermethylated genes in cancer (4).

Interestingly, we found, however, that the H3K27me3 enrichment at DNA-hypermethylated genes is actually intermediate when compared with other PcG-marked genes. This is illustrated by a comparison of these DNA-hypermethylated genes to a group of 815 genes with very high H3K27me3 enrichment (Fig. 1F, for average patterns, and for individual genes see Supplementary Fig. S5—gene list will be included online). The difference in H3K27me3 enrichment at hypermethylated genes is highly statistically significant when compared with levels at either active or high PcG genes (P < 2e−16, Wilcoxon rank sum test). Interestingly, this high H3K27me3 enrichment selects for those genes lacking CpG islands (P < 2.2e−6, Fisher’s exact test). Of the 815 genes selected for this highest H3K27me3 enrichment, 698 (86%) lack such islands in the promoter region. However, the pattern of broad distribution around the annotated start sites of the genes seems to be similar to that for the DNA-hypermethylated genes (Supplementary Fig. S5).

Having defined the above gene expression and chromatin parameters, we sought to characterize essentially all of the candidate promoter CpG island DNA-hypermethylated genes in HCT116 colon cancer cells. As shown schematically in Fig. 2A, our previous discovery approach using microarray expression analysis has discovered hypermethylated genes by defining a zone of increased expression following either genetic disruption of two DNA methyltransferases, DNMT1 and 3b (double-knockout cells), or treatment of the cells with the DNMT inhibitor 5-aza-2'-deoxycytidine (5-aza-dC; ref. 6). Further, we excluded those genes, which increased expression following treatment of the HCT116 cells with trichostatin A (TSA), an inhibitor of histone deacetylases (6). Such refractoriness to TSA is a well-defined feature of genes with densely DNA-methylated promoter CpG islands (18, 19). Finally, only genes with low basal expression on the expression microarray (previously using a cutoff that would include genes currently characterized in both the intermediate and very low expression ranges shown in Fig. 1A) were considered as possible candidates for silencing by DNA hypermethylation.
Figure 1. Chromatin signature of active and DNA-hypermethylated genes in colorectal cancer cells. A, box plot showing ranges of expression for key gene groups in HCT116 and double-knockout (DKO) cells. Y axis, log of normalized single channel signal from the Agilent 44K expression microarray. Average expression range is shown for all genes in HCT116 cells; 4,500 actively expressed genes with CpG islands in HCT116 cells; 42 verified DNA-hypermethylated genes in HCT116 cells; all top tier and next tier candidate DNA-hypermethylated genes (Fig. 2A); the 42 verified DNA-methylated genes in double-knockout cells; and the top tier and next tier genes in double-knockout cells. Orange lines, limits separating active, intermediate, and silent expression zones. TT, top tier; NT, next tier. B, plot showing normalized enrichment (bound/input) of histone marks (Y axis) as a function of distance (X axis, with 0 indicating annotated transcription start site) at the active CDKN2B gene promoter in HCT116 cells. C, composite graph showing average normalized enrichment of histone marks at >4,500 active genes with CpG island-containing promoters in HCT116 cells. D, plot showing normalized enrichment of histone marks at the hypermethylated and silent GATA4 gene promoter in HCT116 cells. E, composite graph showing average normalized enrichment of histone marks at 42 verified DNA-hypermethylated and silent genes in HCT116 cells. F, box plot showing range of enrichment for H3K27me3 at active, DNA-hypermethylated and silent, and genes with highest H3K27 enrichment.
We now show, through our refined expression state classification in Fig. 1A, that many of the genes that ended up as being false positives for DNA hypermethylation in our previous study (6) actually exhibit chromatin characteristics of active genes. We examined a group of more than 1,792 genes in HCT116 cells (gene list will be included online) whose expression behavior on microarray analysis (increase with 5-aza-dC and no increase with TSA, regardless of basal expression) placed them in the top tier and next tier candidate zone (zones 1 and 2 in Fig. 2A). Surprisingly, the chromatin of 515 genes in this tier, which had intermediate expression on the microarray (Fig. 1A), have distinctly more H3K4me2 enrichment around the transcription start site when compared with a group of 610 genes within the very low expression region occupied by the verified DNA-hypermethylated genes (Fig. 2B). Thus, these genes with intermediate expression levels have an active chromatin pattern for this mark (Fig. 2B).

To place these findings into perspective for the purpose of enriching the efficiency of our expression array paradigm to randomly identify CpG island DNA-hypermethylated genes, we examined the H3K4me2 mark for 23 genes previously identified in our recent study (6) as being false positives for promoter CpG island DNA methylation. The leading characteristic of these false-positive genes in our previous study was that they had a basal level of expression by RT-PCR rather than the absent mRNA signal that characterized all of the genes verified to have dense promoter DNA hypermethylation (6). Indeed, we now find that these 23 genes have an intermediate level of expression on microarray analysis (Supplementary Fig. S1). Moreover, by ChIP-chip analyses, these false-positive genes have a H3K4me2 pattern similar to those for active genes (Fig. 2C).

Strikingly, having now eliminated the above 515 genes with intermediate basal expression and enriched H3K4me2 from the candidate list, the chromatin pattern for the remaining 610 genes...
within the silent zone (gene list in Supplementary Table S2) shows an identical chromatin pattern to that of the 42 verified DNA-hypermethylated genes. Individual patterns for selected genes from the 610 are shown in Supplementary Fig. S6. The pattern includes a broad distribution of H3K27me3 around the gene promoters and a low H3K4me2 peak positioned directly over the transcription start site without the dip for the nucleosome-free region (Fig. 2D). Thus, the level of H3K4me2, matched with a very stringent level for low basal transcription, can potentially eliminate false positives, eliminating some 28.75% of genes in the top tier (zone 1 in Fig. 2A) and 51.09% of the next tier (zone 2 in Fig. 2A) in the no TSA response zone that are not truly DNA methylated. We have previously identified that verification of genes as DNA methylated and silenced in the cell lines on which the microarray approach is used is ~80% in the top tier and 50% in the next tier (6). Thus, knowledge of the chromatin status of the genes in cancer cells can markedly increase the efficiency of our expression array approach for identification of genes with promoter CpG island DNA methylation and complete transcriptional silencing.

Our previous chromatin studies at a limited number of DNA-hypermethylated genes suggested that even when these genes are induced to reexpress by DNA demethylation, they do not return to a highly active state but rather to one of low, poised transcription with exhibition of the bivalent chromatin pattern (4). In stem/progenitor cells, this pattern is thought to hold a key group of genes in a poised, low transcription state, necessary to maintain pluripotency (5). We were now in a position to test whether this is truly a universal property of CpG island DNA-hypermethylated genes in colon cancer cells. We examined this by comparing HCT116 cells to their isogenic derivative double-knockout cells, in which all tested DNA-hypermethylated genes are demethylated and reexpressed (3, 20).

We first find that the expression of the 42 verified DNA-hypermethylated genes is distinctly increased to an intermediate expression range in double-knockout cells (Fig. 1A), verifying our RT-PCR results and microarray analysis in our previous study (6). We then examined the chromatin profile of the genes in the double-knockout cells, shown for example for the individual SFRP1...
gene promoter in Fig. 3A and B; for other individual genes in Supplementary Figs. S3, S4, and S6; and for the composite profile of the validated hypermethylated genes in Fig. 3C (matching of previous local ChIP data to ChIP-chip data for five individual DNA-hypermethylated genes is shown in Supplementary Fig. S3). These promoters distinctly adopt a bivalent chromatin pattern on DNA demethylation, characterized by an increase in H3K27me3 enrichment near the gene start sites ($P = 0.009873$, Wilcoxon rank sum test; Fig. 1F), with a simultaneous increase in H3K4me2 levels ($P = 5.865 \times 10^{-7}$, Wilcoxon rank sum test). We next examined the entire list of 610 best candidate DNA-hypermethylated genes from the TSA-negative zone (zones 1 and 2 in Fig. 2A), as defined by the stringent basal expression criteria above. These genes also assume an intermediate expression range in the double-knockout cells (Fig. 1A) and, again, adopt a bivalent pattern chromatin pattern (Fig. 3D). It is particularly interesting that the H3K4me2 mark not only increases but also redistributes the peak positions to either side of the direct transcription start site.

Lastly, we address a question that is key to the biology of promoter CpG island DNA methylation and its associated chromatin, in addition to being clinically important for the potential anticancer therapeutic strategy of reexpressing aberrantly silenced genes (2). An important concept underlying ongoing trials of epigenetic therapy for cancer is that promoter DNA methylation is dominant over histone deacetylation. Thus, if a low dose of a DNA-demethylating agent is first administered to cancer cells, a histone deacetylase inhibitor, which does not alone cause reexpression of aberrantly densely DNA-methylated genes, is then able to exert synergistic effects with the DNA demethylation for reexpression of such genes (18, 19). To examine chromatin patterns that may help explain the above relationships between DNA methylation and the effects of histone deacetylase inhibitors, we examined the chromatin of genes whose expression is augmented by treatment with TSA alone (zones 3 and 4 in Fig. 2A; gene list will be included online). We first found that the average basal expression of these genes is generally higher than for the DNA-hypermethylated genes (Fig. 4A). Most importantly, the chromatin revealed a striking difference, showing that genes stimulated by TSA alone have much higher basal enrichment of H3K4me2, and the characteristic dip that occurs directly at the transcription start site for active genes (Fig. 4B and C). Whether examining genes that were stimulated only by TSA (zone 3 in Fig. 2A) or genes that increased expression after either 5-aza-dC or TSA treatment alone (zone 4 in Fig. 2A), marked enrichment for the H3K4me2 active mark was found (Fig. 4B and C).

Thus, these global studies now provide more insight into why DNA-hypermethylated genes cannot be reexpressed by TSA alone, and why initial DNA demethylation is required before these genes are sensitized to TSA responsiveness. Our previous local ChIP studies of a delimited number of DNA-hypermethylated genes in 5-aza-dC–treated cells, or double-knockout cells (4), and now our current study of many such genes in the ChIP-chip analyses all suggest that removal of DNA methylation results in increases in promoter region H3K4me2. Thus, there is a resultant transformation of a fully silenced transcription state to one of a low-expression, poised transcription state. The enrichment of H3K4me2 at the transcription start site in this state now resembles that for genes that can be reexpressed by administration of the histone deacetylase inhibitor alone.

In summary, our global genome studies of two important histone modification marks in colon cancer cells have taught us much about...
a key component of abnormal epigenetic gene regulation in cancer cells, gene promoter CpG island DNA methylation, and associated tight transcriptional silencing. Our findings firmly suggest that DNA methylation is superimposed on a bivalent chromatin state, which has been best associated with the CpG islands of an important subset of genes in embryonic stem and committed progenitor cells. These global findings further link the origins of aberrantly DNA-hypermethylated genes in cancer cells to an underlying chromatin pattern similar to that of an important group of CpG island-containing, low-expression genes in stem/progenitor cells. We can now provide additional support for these findings by the fact that 48.69% ($P < 2.2e^{-16}$, Fisher’s exact test) of the 610 best candidate DNA-hypermethylated genes, as we now define them in this study, are listed as PcG marked in published tiling studies of embryonic stem cells, embryonic fibroblasts, mouse embryonic fibroblasts, or neural progenitor cells (21–23). This accentuates the likelihood that this precursor cell state, which normally helps hold genes in a low-transcription poised state, may predispose such genes to DNA methylation and conversion of this state to very tight, heritable, transcriptional silencing in adult cancers. As we have discussed, this chromatin pattern may reflect the early cell compartments from which cancers arise (24). Loss of function of many of the genes involved may help cancers retain properties of these early cells at the expense of normal differentiation. Finally, our chromatin findings help explain how CpG island DNA methylation must be relieved to allow for reexpression via histone deacetylase inhibition. This finding may help refine the concepts underlying ongoing clinical trials for combined use of DNA-demethylating and histone deacetylase inhibitors in treating cancer.

**Disclosure of Potential Conflicts of Interest**

S.B. Baylin and J.G. Herman: commercial research grant, speakers bureau/honoraria from OncoMethylome Sciences.

**Acknowledgments**

Received 2/26/2008; revised 4/10/2008; accepted 5/16/2008.

**Grant support:** National Institute of Environmental Health Sciences grant ES11858 and NIH grant CA116160.

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

**References**


# Defining a Chromatin Pattern that Characterizes DNA-Hypermethylated Genes in Colon Cancer Cells

Kelly M. McGarvey, Leander Van Neste, Leslie Cope, et al.


## Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/14/5753

## Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2008/08/11/68.14.5753.DC1

## Cited articles
This article cites 24 articles, 4 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/14/5753.full.html#ref-list-1

## Citing articles
This article has been cited by 32 HighWire-hosted articles. Access the articles at:
/content/68/14/5753.full.html#related-urls

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

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

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