**GENOME-WIDE ANALYSIS OF DNA METHYLATION IN COLORECTAL CANCER**

Sarah Bazzocco1,2\*, M. Carmen Ruiz de Villa3, Paulo Rodrigues1,2, Elena Andretta1,2, Hafid Alazzouzi1,2, Georgia Corner4, Simo Schwartz Jr2,5, Alex Sánchez3, John M. Mariadason4, Diego Arango1,2

*1Group of Molecular Oncology, CIBBIM-Nanomedicine, Vall d’Hebron University Hospital, Research Institute (VHIR), Universitat Autònoma de Barcelona, Passeig Vall d’Hebron, 119-129, 08035 Barcelona, Spain; 2CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN); Spain; 3Departament d'Estadística, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain; 4Ludwig Institute for Cancer Research, Melbourne Centre for Clinical Sciences, Austin Health, Heidelberg, Victoria, Australia; 5Group of Drug Delivery and Targeting, CIBBIM-Nanomedicine, Vall d’Hebron University Hospital, Research Institute (VHIR), Universitat Autònoma de Barcelona, Passeig Vall d’Hebron, 119-129, 08035 Barcelona, Spain.*

**RUNNING TITLE:** CRC methylome.

**CONTACT INFORMATION:** Correspondence and requests for materials should be addressed to DA: email: diego.arango@vhir.org; Tel: +34-93-274-6739; Fax: +34-93-489-3893.

**ABSTRACT**

Methylation of CpG dinucleotides in the promoter of genes involved in the oncogenic process has been shown to be a key process contributing to tumor initiation and/or progression. This epigenetic process occurring throughout the genome is believed to significantly contribute to the profound expression reprograming of colorectal tumor cells. However, further investigation is required to identify the genes regulated by DNA methylation during tumorigenesis. Using genome-wide analysis of the levels of methylation of >14,000 promoters and the levels of mRNA expression in a panel of 45 cell lines derived from colorectal tumors, we found a subset of 677 genes whose levels of methylation were significantly associated with their expression levels. An enrichment of zinc finger proteins was observed among these 677 genes and higher overall methylation levels were associated with absence of APC mutations, faster growth and a microsatellite instable phenotype. Moreover, a group of cell lines with significantly higher methylation levels was observed, supporting the notion that there is a group of colorectal tumors with a CpG methylator phenotype (CIMP+). In addition, we found that approximately one fourth of the genes regulated by CpG methylation are not associated with CpG islands, indicating that DNA methylation outside these CpG rich regions is an important mechanism regulating gene expression and significantly contribute to tumor progression. These results shed new light into the role of the epigenetic mechanisms driving the tumorigenic process in colorectal cancer.

**INTRODUCTION**

Colorectal cancer progression is associated with the accumulation of genetic and epigenetic defects that drive the transition from normal intestinal epithelial cells to full-blown metastatic colorectal tumors. The genetic defects driving the tumorigenic process in colorectal cancer have been studied for several decades and include mutations in *APC, KRAS, TP53* and the loss of heterozygosity (LOH) in chromosome 18q ([1-5](#_ENREF_1)). Epigenetic changes, such as DNA methylation or histone acetylation, modulate gene expression without affecting the DNA nucleotide sequence. Methylation of cytosines in CpG dinucleotides is common in the human genome and a general hypomethylation has been observed in human tumors ([6](#_ENREF_6)). However, CpG hypermethylation of regions rich in CpG dinucleotides (CpG islands) in the promoter regions of tumor suppressor genes is an important mechanism of gene silencing during tumor progression ([7](#_ENREF_7)).

Chromosomal instability is a common feature of approximately 80% of colorectal tumors and results in large chromosomal gains and losses. An alternative tumorigenic mechanism is observed in 10-20% of the tumors, and seems to be driven by the epigenetic defects causing widespread hypermethylation of CpG islands throughout the genome. The promoter of the mismatch repair gene MLH1 is frequently methylated in these tumors, accounting for the microsatellite instable (MSI) phenotype observed in a large proportion of these so called ‘CpG island methylator phenotype’ (CIMP). Both MSI+ and CIMP+ tumors are associated with proximal tumor location, poor differentiation, and KRAS/BRAF mutations ([8](#_ENREF_8), [9](#_ENREF_9)) and may arise from hyperplastic polyps and serrated adenomas ([10](#_ENREF_10), [11](#_ENREF_11)).

While the genetic events that drive the tumorigenic process are relatively well characterized for colorectal cancer, the epigenetic events and their impact on the transcriptional reprogramming observed in colorectal tumors have not been extensively characterized. Although recent genome-wide studies have analyzed the genomic distribution of hypermethylated CpGs in a small number of colorectal tumors ([12](#_ENREF_12)), a detailed analysis of the subset of these events that are important for gene expression regulation is currently lacking.

In this study we used a panel of 45 colorectal cancer cells and CpG methylation arrays to investigate the genome-wide distribution of these epigenetic changes in colorectal cancer cells. In addition, we used gene expression microarrays of a subset of 30 colorectal cancer cell lines to identify the genes that are frequently regulated by promoter methylation in colon cancer.

**MATERIALS AND METHODS**

**Cell lines and primary tumors.** A total of 45 colorectal cancer cell lines were used in this study: ALA, CACO2, CO115, COLO201, COLO205, COLO320, DLD1, GP5D, HCA7, HCC2998, HCT116, HCT15, HDC108, HDC9, HT29, IS1, IS2, IS3, KM12, LIM1215, LIM2405, LOVO, LS1034, LS174T, LS513, RKO, RW2982, RW7213, SKCO1, SW1116, SW403, SW48, SW620, SW837, SW948, T84, TC71, V9P, VACO5, FET, HDC111, HDC114, HDC75, HDC87 and HDC15. All cell lines were grown in DMEM medium supplemented with 10% FBS and 1x antibiotic antimycotic (Invitrogen). All lines were tested to be negative for mycoplasma contamination (PCR Mycoplasma Detection Set, Takara). Cell lines were cultured in 100mm plates until reaching 70% confluence and the medium was changed 8h before harvesting the cultures. Possible cell line cross-contamination was investigated by clustering analysis of using the mRNA expression and promoter methylation data, confirming the identity of the tumor lines used.

**DNA methylation.** We assessed the quantitative levels of methylation at the single nucleotide resolution level for a total of 27,578 highly informative CpG sites using HumanMethylation27 Beadchips (Illumina). These chips target CpG sites located within the proximal promoter regions of transcription start sites of 14,475 consensus coding sequencing (CCDS) in the NCBI Database (Genome Build 36). On average, two CpG sites were selected per CCDS gene and from 3-20 CpG sites for >200 cancer-related and imprinted genes. The levels of DNA methylation were studied in 45 different lines. In addition, one of the lines (SW48) was hybridized twice. Also, an *in vitro* methylated control (CpG Methylated Jurkat Genomic DNA, New England BioLabs) and an unmethylated control (Jurkat DNA amplified in vitro with illustra GenomiPhi HY DNA Amplification Kit from GE Healthcare) were included in the experimental design. The DNA from all samples was extracted using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich) and then bisulfite treated and hybridized following manufacturer’s recommendations at the Spanish National genotyping Center (CeGen-CRG Genotyping Unit). The levels of methylation were calculated using GenomeStudio software (Illumina). In some analysis, we sued the levels of methylation of primary tumors and normal colonic mucosa samples using HumanMethylation27 beadchips (Illumina) as previously reported in the Gene Expression Omnibus repository (GSE17648) or The Cancer Genome Atlas ([13](#_ENREF_13)). Promoter methylation microarrays data (HumanMethylation27, Illumina) for HCT116 cells where DNA methyltransferase activity has been inactivated either pharmacologically (5-Aza-2'-deoxycytidine treatment) or genetically (*DNMT1* and *DNMT3b* double knockout) can be found in the ArrayExpress (E-MTAB-210) or Gene Expression Omnibus (GSE26990) repositories, respectively ([14](#_ENREF_14)).

**mRNA expression.** The levels of expression of more than 47,000 transcripts and variants, including more than 38,500 well characterized genes and UniGenes, were investigated using GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix) in a subset of 30 colorrectal cancer cell lines (CACO2, CO115, CCOLO201, COLO205, COLO320, DLD1, HCC2998, HCT116, HCT15, HT29, IS1, IS2, KM12, LIM1215, LIM2405, LOVO, LS174T, RKO, RW2982, RW7213, SKCO1, SW1116, SW403, SW48, SW620, SW837, SW948, T84, TC71 and Vaco5). Total RNA was extracted with TRIzol Reagent (Invitrogen) and then labeled and hybridized to HG-U133 Plus 2.0 chips at the Ludwig Institute for Cancer Research, Melbourne. The relative mRNA levels were calculated after RMA (Robust Multichip Average) normalization as described ([15](#_ENREF_15)). The mRNA expression microarray analysis (GeneChip Human Genome U133 Plus 2.0 Array, Affymetrix) of HCT116 colon cancer cells treated with the demethylating agent 5-Aza-2'-deoxycytidine (GSE5816) or HCT116 that are double knockout for the DNA methyltransferases DNMT1 and DNMT3B (GSE26018) are from Gene Expression Omnibus ([16](#_ENREF_16), [17](#_ENREF_17)).

**Associations between mRNA and promoter methylation levels.** Using the gene symbol as a common identifier, we found that expression data was available for a total of 11,858 (81.92%) of the 14475 promoters interrogated in the HumanMethylation27 arrays. If there was more than one probe for a given gene/promoter, we used the average value of expression/methylation. The Pearson correlation coefficient was used to identify significant negative correlations (r>-0.355; p<0.05) between gene expression and promoter methylation in 30 colorectal cancer cell lines. This identified 1409 genes. To eliminate from this list genes whose correlation is heavily dependent on a small number of cell lines, the graphs of the expression (Y-axis) and methylation (X-axis) of all 30 cell lines was divided in four quadrants using the median expression/methylation values. Only genes with >5 data points in the upper left and lower right quadrants were further considered (see examples in **Figure 3A-D**). In addition, genes showing no significant Pearson’s correlation between methylation and expression, but that showed high expression levels in at least a subset of the cell lines with low methylation levels and low/absent expression in cell lines with high methylation levels (i.e., ‘L-shaped’ in the scattered plots; see **Figure 3E-H**) were also included. **Please, Mamen, can you briefly describe the maths behind this?**

**Functional group enrichment analysis.** To investigate whether there was a significant enrichment in the number of genes with significant correlation between expression and methylation levels that belonged to different categories of functionally related genes, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID ) v6.7 ([18](#_ENREF_18)). This Functional Classification Tool generates a gene-to-gene similarity matrix based on shared functional annotation using over 75,000 terms from 14 functional annotation sources. A Fisher’s exact test was used to identify significantly enriched categories in the 1,409 genes regulated by methylation relative to the 11,858 genes investigated. The Benjamini-Hochberg procedure was used to correct for multiple hypothesis testing (p<0.05).

**RNA extraction and quantitative RT-PCR.** Cell cultures were harvested at 70% confluence and total RNA was extracted using the TRI Reagent (Molecular Research Center) according to the manufacturer’s instructions. Total RNA (500ng) was reverse transcribed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems), and relative mRNA levels of ITGA9, KLHL3, and PP1R14D, respectively, were assessed by Real-Time PCR using SYBR Green Master Mix (Applied Biosystems, Branchburg, NJ). We used 18S rRNA (Taqman Master Mix) as a standardization control for the 2-ΔΔCt method as described before ([18](https://webmail.vhir.org/zimbra/public/blank.html" \l "_ENREF_18" \t "_blank" \o "Arango, 2005 #21)). The primers used were KLHL3-qPCR-forward 5’-AGT ACT GGC CTA GCA TCG GT-3’, KLHL3-qPCR-reverse 5’-CGG GAA GCT CCA TCA TAA C-3’, PPP1R14D-qPCR-forward 5’-AGA CTC AGC TGG AGG CCA T-3’, PPP1R14D-qPCR-reverse 5’-CAG TGC TGA GGC TGC TAA AG-3’, ITGA9-qPCR-forward 5’-GTT GGT GGG AAT CCT CAT CT-3’, ITGA9-qPCR-R 5’-AAA GAA GCC CAT CTT CCA GA-3’, 18S rRNA-forward 5'-AGT CCC TGC CCT TTG TAC ACA-3'; 18S-rRNA-reverse 5'-GAT CCG AGG GCC TCA CTA AAC-3', 18S Probe 5'-FAM-CGC CCG TCG CTA CCG ATT GG-TAMRA-3'.

**Bisulfite sequencing.** Genomic DNA was isolated from each line using the GenElute™Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO) and bisulfite treated with EZ DNA Methylation-Gold Kit (Zymo Research) both, following the manufacturer’s instructions and then PCR-amplified with the following conditions: 95ºC for 5 minutes, then 40 cycles of 94ºC for 30 seconds, 59.1ºC (for ITGA9 and PPP1R14D) or 64.1 ºC (for KLHL3) for 45 seconds and 72ºC for 30 seconds, and then the last elongation at 72ºC for 10 minutes. Primers for DNA amplification and sequencing were designed using MethPrimer software ([19](#_ENREF_19)) and were ITGA9-forward 5’-GGG ATT TGA GGA TTT GTA TTT TTT-3’, ITGA9-reverse 5’-CCT CTA CTC CTT CAC CCA ATT ATA A-3’, PPP1R14D-forward 5’- AGG TTA GGT TGA TAG TAG TTT ATA TT-3’, PPP1R14D-reverse 5’-CCT CTA TAT CCC ACC TTC CTA AAA C-3’, KLHL3-forward 5’-AAGT TGG AAA GGT GGT AGT GTA TTT-3’, KLHL3-reverse 5’-CCA ACA AAC CAA TAA AAA ATC TAA TC-3’. The amplified regions were sequenced using the Macrogen standard sequencing facility (Macrogen Inc, Korea).

**RESULTS**

**Genome-wide analysis of CpG methylation in colorectal cancer cell lines.** To investigate the levels of CpG methylation throughout the genome of colorectal cancer cells we used a microarray approach with single CpG dinucleotide resolution to quantify the levels of methylation of the promoters of 14,475 consensus coding sequences (CCDS) in a panel of 45 different colorectal cancer cell lines. The study design included an unmethylated control (mean methylation 5%), an in vitro methylated control (mean methylation 86%; See **Figure 1A**) and a biological replicate showing excellent correlation between independent methylation measurements (**Figure 1B**; Pearson r=0.998, p<0.0001; **Alex, please, calculate exact p values n=27578**).

The average methylation per sample for all the >27,000 individual CpGs interrogated ranged from 22% to 43% with a mean of 31±5% (**Figure 1A**; mean±SD). The average methylation per CpG for all the 45 lines in the study showed a bimodal distribution. Most (>50%) of the >27,000 CpGs had levels of methylation below 20%. However, a significant number of CpGs (>30%) showed methylation levels above 50% (**Figure 1C**). In agreement with the reported association between MSI and CIMP+, the overall levels of methylation were significantly higher in MSI lines compared to MSS lines (Student’s T-test p=0.001; **Figure 1D**). In addition, the average levels of methylation across the >14,000 promoters investigated was significantly lower in the cell lines with *APC* mutations (Student’s T-test p=0.043; **Figure 1E**). Moreover, there was a significant association between higher methylation levels and faster growth in this panel of 45 colorectal cancer cell lines (**Figure 1F**; Pearson’s r=-0.39; p=0.010), suggesting that higher levels of promoter CpG methylation may contribute to the uncontrolled proliferation characteristic of tumor cells. No associations were found between the overall methylation levels of these 45 cell lines and mutations in *TP53*, *KRAS*, *BRAF* or *PIC3CA* (Student’s T-test p>0.25; **Supplementary Figure 1**).

We used bisulfite sequencing to validate the results of the methylation microarrays using an independent technique. We randomly selected 3 genes showing wide differences in methylation levels in this set of cell lines, and after bisulfite treatment, a genomic region containing the CpG interrogated by the methylation arrays was directly PCR amplified and sequenced. Bisulfite sequencing results were in perfect agreement with the methylation levels observed with the methylation microarrays (**Figure 2A-D**).

**CpG island methylator phenotype.** It has been proposed that widespread CpG island hypermethylation defines a subset of colorectal tumors, known as ‘CpG island methylator phenotype’ positive (CIMP+) ([9](#_ENREF_9), [20](#_ENREF_20)). However, the existence of CIMP+ tumors has been challenged in some studies because colorectal tumors could not be clearly dichotomized as CIMP+ and CIMP- since a continues gradient was observed in the overall levels of promoter methylation ([21](#_ENREF_21), [22](#_ENREF_22)). In good agreement, here we found that if the methylation levels of all the >27,000 CpGs interrogated or a subset 1,516 of them with the highest methylation variability among samples (1.0<SD/Mean<1000) is considered, the distribution of the number of cell lines with different average methylation levels did not show any evidence of bimodality (not shown). The original definition of CpG island methylator phenotype used a set of promoters that do not show high methylation levels in normal colonic samples –‘type C’ promoters ([20](#_ENREF_20)). Here, we found that the average methylation of a selection of 643 ‘Type C’ promoters in this panel of 45 colon cancer cell lines, showed a bimodal distribution and could be used to classify a group of 18 cell lines as CIMP+ (**Figure 3A**). Moreover, clustering analysis of the levels of methylation of these 643 ‘Type C’ promoters demonstrated that CIMP+ cell lines form a homogeneous branch of the dendrogram generated (**Figure 3B**). Therefore, a threshold methylation value of 0.6 could be used in this series to classify tumor samples as CIMP+ or CIMP-.

**Genes transcriptionally regulated by promoter methylation in colorectal cancer.** Although it is widely accepted that aberrant CpG methylation is a frequent event in colorectal cancer and other cancer types and that it is important in the transcriptional regulation of many genes, a detailed list of genes whose transcription is frequently regulated by CpG methylation is currently lacking. Here we used microarray analysis (Affymetrix U133 Plus 2.0) to assess the levels of expression of >47,000 transcripts. Expression data was available for a total of 11,858 (81.92%) of the 14,475 promoters interrogated in the HumanMethylation27 arrays. To validate the expression data obtained with the mRNA microarrays with an independent technique, we used Real-Time RT-PCR quantification of the levels of expression of 3 randomly selected genes. The quantitative RT-PCR results obtained are in perfect agreement with the mRNA microarray results (**Figure 2E-G**).

We found a significant inverse correlation (Pearson r<-0.355) between the levels of mRNA expression and the levels of promoter methylation of 1,409 genes (9.7%). To eliminate genes whose correlation between expression and methylation levels was driven by a small number of samples (see examples in **Supplementary Figure 2**), we used the criteria described in the ‘Materials and Methods’ section and identified a subset of 652 candidate genes whose expression is regulated by promoter methylation (**Figure 4A-D**). Moreover, we identified an additional set of 25 genes whose expression was not significantly correlated with the levels of methylation (Pearson’s r>-0.355), but are likely to be regulated by promoter methylation (**Figure 4E-H**). We therefore identified a total of 677 of the 11,858 genes investigated (5.7%) as candidate genes whose expression is regulated by promoter methylation (**Table 1**). Because high levels of promoter methylation are expected to result in low levels of gene expression (i.e., negative correlations), to assess the robustness of the analysis, we searched for genes whose expression was positively correlated with the levels of methylation and found 325 genes. As expected, we observed a strong bias towards negatively correlated genes (1409 of all 1734; 81.25% of all significantly correlated genes; Binomial test p=1.6x10-156 **--- Fisher’s Exact test: p<0.0001: Alex, please calculate exact p value**), convincingly showing that there are underlying biological reasons for the observed associations (namely, promoter methylation negatively regulates gene expression) and providing a good estimate of the proportion of false positive genes in our analysis.

To further assess whether these genes were regulated by CpG methylation of their promoter regions, we used mRNA expression microarray analysis of HCT116 colon cancer cells and HCT116 cells treated with the demethylating agent 5-Aza-2'-deoxycytidine or HCT116 that are double knockout for the DNA methyltransferases DNMT1 and DNMT3B. When looking at the genes that had methylation levels >50% in parental HCT116 cells (n=260 genes), we found that the majority of the genes (69.6% and 65.2%) had higher levels of expression in the HCT116+AZA or HCT116 DKO cells compared to the parental HCT116 cells (binomial p= 6.3x10-11 and 3.5x10-7, respectively). Therefore, the combined analysis of mRNA expression and promoter methylation levels identified a core set of 667 genes that are frequently regulated at the transcriptional level by aberrant CpG promoter methylation in colorectal tumors (**Table 1**).

The analysis of genes regulated by promoter methylation in primary colorectal tumors is complicated by the contamination of tumor samples with normal tissue. Tumor samples contain a significant proportion of normal cells from infiltrating lymphocytes, stromal cells, lymph/blood vessels, etc, that can significantly interfere with the quantification of the levels of methylation and expression observed in tumor samples. However, to further investigate the regulatory effects of promoter methylation on gene expression we used a series of 222 primary colorrectal tumors. Methylation and mRNA expression information was available for 12,719 genes in this series of primary tumors. As observed with the cell lines, there was a strong bias for negative correlations, as 3,542/12,719 (27.85%) showed a significant negative correlation between methylation and mRNA levels (Pearson’s r<-0.13) and only 1,039/12,719 genes showed significant positive correlations (8.16%; Binomial p<7.4x10-305). Of the 677 genes identified as regulated by promoter methylation in the tumor lines, expression and methylation data in the primary tumors was available from 669 genes. A negative correlation (Pearson’s r<-0.13) was observed between expression and methylation for 432/669 genes (64.57%; show **Table 1**), while only 10 of these 669 genes (1.49%) had a significant positive correlation between mRNA and methylation levels. Therefore, despite the interference of the contamination with DNA derived from the normal tissue in primary tumor samples, evidence of epigenetic silencing could be observed for the majority (>64%) of the genes initially identified in colorectal cancer cell lines.

**CpG methylation outside of CpG islands can regulate gene expression.** CpG islands are regions of increased density of CpG dinucleotides ([23](#_ENREF_23), [24](#_ENREF_24)). CpG methylation within CpG islands has long been known to regulate transcriptional activity. Surprisingly, it has recently been reported that CpG methylation in the ‘shore’ of these dense CpG island regions may be even more relevant for the transcriptional silencing of the associated promoter regions. However, the role of CpG methylation in regions not associated with CpG islands in transcriptional regulation remains unclear. Here we used genome-wide methylation and mRNA expression data on 30 cell lines to gain further insight into this question. Of the 677 candidate genes whose expression was regulated by CpG methylation, 497 (73.4%) were associated with a CpG island as defined by Gardiner-Garden and Frommer ([24](#_ENREF_24)) (UCSC Genome Browser, GRCh37/hq19; **Table 3**). However, the remaining 180 genes (26.6%) with significant associations between mRNA and methylation levels, were not associated with a conventionally defined CpG island. Moreover, using less stringent criteria for defining CpG islands throughout the human genome using a Hidden Markov model-based approach ([25](#_ENREF_25)), 123 genes (18.16%) showing good correlations between expression and methylation levels in colorectal cancer cells, were not associated with any CpG islands. Messenger RNA expression data before and after treatment with 5-Aza-2'-deoxycytidine in HCT116 colon cancer cells was available for 675 of the 677 genes identified as being epigenetically regulated. Of these, 260 genes showed >50% methylation in the associated CpG dinucleotides investigated and 92 of these (35.38%) had no CpG islands associated. Importantly, 71 of these 92 genes (77.17%; Binomial test 5.8x10-8) showed elevated expression levels after 5-Aza-2'-deoxycytidine treatment. Collectively these data indicate that CpG methylation outside CpG islands can regulate the transcriptional activity of a significant number of genes in colorectal cancer.

**Functional group enrichment analysis of genes regulated by CpG promoter methylation.** To gain a better understanding of the biological function of the genes regulated by CpG methylation, we used functional group enrichment analysis. Of all the categories analyzed (see methods), we identified 17 functional groups significantly enriched (average fold enrichment >2) in the number of genes regulated by CpG promoter methylation (Fisher’s Exact test, Benjamini correction, p<0.025; **Table 2**). Strikingly, all these overlapping categories are related to zinc finger proteins. The largest of these categories (Interpro IPR013087: Zinc finger, C2H2-type/integrase, DNA-binding) contains a total of 38 zinc finger domain proteins, including a large proportion of transcription factors (**Table 2** and **Supplementary** **Table 1**).

**DISCUSSION**

Cytosine methylation in CpG dinucleotides is important for normal development and cell differentiation in higher organisms. These methylation marks lead to chromatin condensation and gene silencing and CpG methylation is inherited by the daughter cells after cell division. Aberrant methylation of CpG dinucleotides in the 5’ regulatory regions of genes that are involved in the oncogenic process has emerged as an important mechanism leading to the initiation and/or progression of different tumor types, including tumors of the colon and rectum. However, the detailed transcriptional reprograming resulting from aberrant methylation in colorectal tumors remains to be thoroughly characterized. The advent of high throughput technologies to investigate genome-wide levels of both mRNA expression and CpG methylation allows the systematic analysis of the genes regulated by CpG methylation throughout the genome in colorectal tumors.

Although CpG promoter hypermethylation was found throughout the genome of colorectal cancer cell lines, only the expression of a subset of the genes showing frequent promoter methylation were shown to be regulated through aberrant methylation. This study describes for the first time details of the genome-wide transcriptional reprogramming resulting from aberrant methylation in colorectal tumors at the level of individual genes. We found that the expression of 677 of the 11,858 genes investigated (5,6%) was silenced by promoter methylation. These genes included important tumor suppressor genes known to be frequently methylated in colorectal tumors such as E-cadherin (CDH1) (REFS), CA9 (Carbonic anhydrase IX) ([26-28](#_ENREF_26)), the serine/threonine protein kinase DAPK2 ([29](#_ENREF_29)) and mismatch repair gene MLH1 (REFS). In addition, other important tumor suppressor genes and oncogenes that have not previously been reported to be methylated in colorectal tumors were also identified. These included F11R (junctional adhesion molecule A) and CLD1 (Claudin 1), two important regulators of tight junction assembly in epithelial sheets ([30-32](#_ENREF_30)), the proapoptotic Bcl2-family member BOK (BCL2-related ovarian killer)([33](#_ENREF_33), [34](#_ENREF_34)) and the Ephrin ligand EFNB2 (Ephrin B2) ([35](#_ENREF_35), [36](#_ENREF_36)).

A significant enrichment of zinc finger proteins was observed in the number of genes regulated by methylation in colorectal tumors. Zinc finger proteins contain a structural domain that has one or more zinc ions stabilizing the protein folding. Zinc finger proteins are DNA binding proteins, conferring sequence specificity to multiple transcription factors. A group of 38 zinc finger transcription factors that are regulated by CpG methylation was identified and are likely to be important in amplifying the transcriptional reprogramming imposed by CpG methylation in colorectal cancer cells.

Recently, there has been some controversy regarding the existence of a new tumorigenic mechanism resulting from the widespread hypermethylation of CpG islands (CIMP+ tumors) ([21](#_ENREF_21), [22](#_ENREF_22)). Although the important role of CpG methylation in tumor progression is well documented, the existence of the CpG methylator phenotype (CIMP) critically depends on the existence of a bimodal distribution in the levels of methylation, rather than a continuous gradient in the levels of methylation in colorectal tumors. The quantitative analysis of >27,000 CpGs in a panel of 45 cell lines derived from colorectal tumors allows the genome-wide analysis of the levels of CpG methylation in large sample sets. Moreover, unlike studies with primary tumors, the interpretation of the results of methylation analysis of cell liens is not complicated by the presence of a variable proportion of ‘contaminating’ normal cells (stromal, infiltrating lymphocytes, blood/lymph vessels, etc). Although using the methylation status of all >27,000 CpGs interrogated by the HumanMethylation27 chips, unsupervised cluster analysis of the 45 cell lines in this study identified a group of cell lines with higher methylation levels, the distribution of cell lines with high/low methylation was not bimodal. However, the initial definition of the CpG methylator phenotype was based in the methylation status of cancer-related promoters that are not methylated in normal colonic mucosa samples (Type C promoters) ([20](#_ENREF_20)). Using the 643 Type C promoters that do not show methylation in a series of 22 normal colonic samples, or the 5 CIMP markers that have recently been used to define the CIMP phenotype, not only clustering analysis identified an extended set of samples showing significantly higher methylation levels, but also the distribution of cell lines with high/low methylation was bimodal. Although the molecular mechanisms underlying this tumorigenic pathway remain to be fully determined, these results support the existence of a type of colorectal tumors driven by the higher incidence of promoter methylation. Moreover, as previously reported ([20](#_ENREF_20)), CIMP+ cases were found to include all the *BRAF* mutant lines as well as the majority (9/17; 53%) of MSI lines. In addition, cell lines with higher methylation levels showed significantly faster growth than cell lines with lower levels of CpG methylation. Although from this analysis it is not possible to establish a causal effect, this is consistent with the observation that several genes that inhibit cell cycle progression (such as, RPRM, PLAGL1 and p16) are frequently methylated and their expression is regulated by these epigenetic defects.

Although the role of cytosine methylation in the context of CpG islands in the regulation of gene expression has been known for a long time, the possible functional effects of aberrant CpG methylation events outside CpG islands is poorly characterized. A recent genome-wide study reported that approximately 24% of the genes that are regulated by methylation outside a conventionally defined CpG island, although the detailed list of these genes was not reported. Here, we found that 26% of the 677 genes that are regulated by CpG methylation are not associated with a CpG island. These included multiple genes known to be important during colorectal cancer progression such as MYO1A, a brush border myosin that we have recently shown to have important tumor suppressor effects in this organ ([37](#_ENREF_37), [38](#_ENREF_38)). Other genes known to be important in the oncogenic process are Cadherin 17 ([39](#_ENREF_39), [40](#_ENREF_40)), selenium binding protein 1 ([41](#_ENREF_41)), TNF, BCL2-like 14 (apoptosis facilitator).

In summary, in this study we used genome-wide microarray analysis of CpG methylation and mRNA expression in a panel of 45 colorectal cancer cell lines to identify the subset of genes whose expression is regulated by epigenetic silencing. A significant enrichment of zinc finger proteins was found, suggesting that deregulation of this family of transcriptional factors important for colorectal tumorigenesis. Moreover, we found that 26% of these genes are not associated with CpG islands. These results significantly contribute towards a deeper understanding of the role of DNA methylation in colorectal cancer and provide the first genome-wide comprehensive catalog of genes regulated by promoter methylation in colorectal tumors.

**Zinc finger proteins:**

[**http://www.nature.com/onc/journal/v32/n3/abs/onc201254a.html**](http://www.nature.com/onc/journal/v32/n3/abs/onc201254a.html)

[**http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0041060**](http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0041060)

[**http://cancerres.aacrjournals.org/content/63/14/3877.full**](http://cancerres.aacrjournals.org/content/63/14/3877.full)

**ACKNOWLEDGEMENTS**

This study was partially funded by grants of the Spanish Ministry for Economy and Competitiveness (CP05/00256, TRA2009-0093, SAF2008-00789, PI12/03112 and PI12/01095), the Association for International Cancer Research (AICR13-0245) and Agència de Gestió d'Ajuts Universitaris i de Recerca (SGR 157) to Diego Arango.

**REFERENCES**

1. Bardelli, A., Parsons, D. W., Silliman, N., Ptak, J., Szabo, S., Saha, S., Markowitz, S., Willson, J. K., Parmigiani, G., Kinzler, K. W., Vogelstein, B., and Velculescu, V. E. (2003) Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300**, 949

2. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274

3. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E., and Vogelstein, B. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113

4. Jen, J., Kim, H., Piantadosi, S., Liu, Z. F., Levitt, R. C., Sistonen, P., Kinzler, K. W., Vogelstein, B., and Hamilton, S. R. (1994) Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N Engl J Med* **331**, 213-221

5. Muzny, D., Bainbridge, M., Chang, K., Dinh, H., Drummond, J., Fowler, G., Kovar, C., Lewis, L., Morgan, M., Newsham, I., Reid, J., Santibanez, J., Shinbrot, E., Trevino, L., Wu, Y., Wang, M., Gunaratne, P., Donehower, L., Creighton, C., Wheeler, D., Gibbs, R., Lawrence, M., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E., Gabriel, S., Getz, G., Ding, L., Fulton, R., Koboldt, D., Wylie, T., Walker, J., Dooling, D., Fulton, L., Delehaunty, K., Fronick, C., Demeter, R., Mardis, E., Wilson, R., Chu, A., Chun, H., Mungall, A., Pleasance, E., Robertson, A., Stoll, D., Balasundaram, M., Birol, I., Butterfield, Y., Chuah, E., Coope, R., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R., Lee, D., Li, H., Mayo, M., Moore, R., Schein, J., Slobodan, J., Tam, A., Thiessen, N., Varhol, R., Zeng, T., Zhao, Y., Jones, S., Marra, M., Bass, A., Ramos, A., Saksena, G., Cherniack, A., Schumacher, S., Tabak, B., Carter, S., Pho, N., Nguyen, H., Onofrio, R., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Meyerson, M., Protopopov, A., Zhang, J., Hadjipanayis, A., Lee, E., Xi, R., Yang, L., Ren, X., Zhang, H., Sathiamoorthy, N., Shukla, S., Chen, P., Haseley, P., Xiao, Y., Lee, S., Seidman, J., Chin, L., Park, P., Kucherlapati, R., Auman, J., Hoadley, K., Du, Y., Wilkerson, M., Shi, Y., Liquori, C., Meng, S., Li, L., Turman, Y., Topal, M., Tan, D., Waring, S., Buda, E., Walsh, J., Jones, C., Mieczkowski, P., Singh, D., Wu, J., Gulabani, A., Dolina, P., Bodenheimer, T., Hoyle, A., Simons, J., Soloway, M., Mose, L., Jefferys, S., Balu, S., O'Connor, B., Prins, J., Chiang, D., Hayes, D., Perou, C., Hinoue, T., Weisenberger, D., Maglinte, D., Pan, F., Berman, B., Van, D., Berg, DJ., Shen, H., Triche, T., Jr., Baylin, S., Laird, P., Getz, G., Noble, M., Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C., Liu, S., Shukla, S., Lawrence, M., Zhou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Park, R., Nazaire, M., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P., Chin, L., Thorsson, V., Reynolds, S., Bernard, B., Kreisberg, R., Lin, J., Iype, L., Bressler, R., Erkkilä, T., Gundapuneni, M., Liu, Y., Norberg, A., Robinson, T., Yang, D., Zhang, W., Shmulevich, I., de, R., JJ., Schultz, N., Cerami, E., Ciriello, G., Goldberg, A., Gross, B., Jacobsen, A., Gao, J., Kaczkowski, B., Sinha, R., Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B., Chan, T., Ladanyi, M., Sander, C., Akbani, R., Zhang, N., Broom, B., Casasent, T., Unruh, A., Wakefield, C., Hamilton, S., Cason, R., Baggerly, K., Weinstein, J., Haussler, D., Benz, C., Stuart, J., Benz, S., Sanborn, J., Vaske, C., Zhu, J., Szeto, C., Scott, G., Yau, C., Ng, S., Goldstein, T., Ellrott, K., Collisson, E., Cozen, A., Zerbino, D., Wilks, C., Craft, B., Spellman, P., Penny, R., Shelton, T., Hatfield, M., Morris, S., Yena, P., Shelton, C., Sherman, M., Paulauskis, J., Gastier-Foster, J., Bowen, J., Ramirez, N., Black, A., Pyatt, R., Wise, L., White, P., Bertagnolli, M., Brown, J., Chan, T., Chu, G., Czerwinski, C., Denstman, F., Dhir, R., Dörner, A., Fuchs, C., Guillem, J., Iacocca, M., Juhl, H., Kaufman, A., Kohl, B., 3rd., Van, L., X., Mariano, M., Medina, E., Meyers, M., Nash, G., Paty, P., Petrelli, N., Rabeno, B., Richards, W., Solit, D., Swanson, P., Temple, L., Tepper, J., Thorp, R., Vakiani, E., Weiser, M., Willis, J., Witkin, G., Zeng, Z., Zinner, M., Zornig, C., Jensen, M., Sfeir, R., Kahn, A., Chu, A., Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Pihl, T., Ayala, B., Backus, M., Walton, J., Whitmore, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta, S., Greene, J., Pot, D., Shaw, K., Dillon, L., Buetow, K., Davidsen, T., Demchok, J., Eley, G., Ferguson, M., Fielding, P., Schaefer, C., Sheth, M., Yang, L., Guyer, M., Ozenberger, B., Palchik, J., Peterson, J., Sofia, H., and Thomson, E. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337

6. Feinberg, A. P., and Vogelstein, B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92

7. Jones, P. A., and Baylin, S. B. (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-428

8. van Rijnsoever, M., Grieu, F., Elsaleh, H., Joseph, D., and Iacopetta, B. (2002) Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. *Gut* **51**, 797-802

9. Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., Kang, G. H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A. J., Thibodeau, S. N., Jass, J., Haile, R., and Laird, P. W. (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793

10. Jass, J. R. (2001) Serrated route to colorectal cancer: back street or super highway? *J Pathol* **193**, 283-285

11. Hawkins, N. J., and Ward, R. L. (2001) Sporadic colorectal cancers with microsatellite instability and their possible origin in hyperplastic polyps and serrated adenomas. *J Natl Cancer Inst* **93**, 1307-1313

12. Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., and Feinberg, A. P. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186

13. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337

14. Sproul, D., Nestor, C., Culley, J., Dickson, J. H., Dixon, J. M., Harrison, D. J., Meehan, R. R., Sims, A. H., and Ramsahoye, B. H. (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A* **108**, 4364-4369

15. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264

16. Shames, D. S., Girard, L., Gao, B., Sato, M., Lewis, C. M., Shivapurkar, N., Jiang, A., Perou, C. M., Kim, Y. H., Pollack, J. R., Fong, K. M., Lam, C. L., Wong, M., Shyr, Y., Nanda, R., Olopade, O. I., Gerald, W., Euhus, D. M., Shay, J. W., Gazdar, A. F., and Minna, J. D. (2006) A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med* **3**, e486

17. Hahn, M. A., Wu, X., Li, A. X., Hahn, T., and Pfeifer, G. P. (2011) Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS ONE* **6**, e18844

18. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57

19. Li, L. C., and Dahiya, R. (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427-1431

20. Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., and Issa, J. P. (1999) CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686

21. Anacleto, C., Leopoldino, A. M., Rossi, B., Soares, F. A., Lopes, A., Rocha, J. C., Caballero, O., Camargo, A. A., Simpson, A. J., and Pena, S. D. (2005) Colorectal cancer "methylator phenotype": fact or artifact? *Neoplasia* **7**, 331-335

22. Yamashita, K., Dai, T., Dai, Y., Yamamoto, F., and Perucho, M. (2003) Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* **4**, 121-131

23. Takai, D., and Jones, P. A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740-3745

24. Gardiner-Garden, M., and Frommer, M. (1987) CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-282

25. Irizarry, R. A., Wu, H., and Feinberg, A. P. (2009) A species-generalized probabilistic model-based definition of CpG islands. *Mamm Genome* **20**, 674-680

26. Nakamura, J., Kitajima, Y., Kai, K., Hashiguchi, K., Hiraki, M., Noshiro, H., and Miyazaki, K. (2011) Expression of hypoxic marker CA IX is regulated by site-specific DNA methylation and is associated with the histology of gastric cancer. *Am J Pathol* **178**, 515-524

27. Jakubickova, L., Biesova, Z., Pastorekova, S., Kettmann, R., and Pastorek, J. (2005) Methylation of the CA9 promoter can modulate expression of the tumor-associated carbonic anhydrase IX in dense carcinoma cell lines. *Int J Oncol* **26**, 1121-1127

28. Zatovicova, M., Jelenska, L., Hulikova, A., Csaderova, L., Ditte, Z., Ditte, P., Goliasova, T., Pastorek, J., and Pastorekova, S. (2010) Carbonic anhydrase IX as an anticancer therapy target: preclinical evaluation of internalizing monoclonal antibody directed to catalytic domain. *Curr Pharm Des* **16**, 3255-3263

29. Mittag, F., Kuester, D., Vieth, M., Peters, B., Stolte, B., Roessner, A., and Schneider-Stock, R. (2006) DAPK promotor methylation is an early event in colorectal carcinogenesis. *Cancer Lett* **240**, 69-75

30. Singh, A. B., Sharma, A., Smith, J. J., Krishnan, M., Chen, X., Eschrich, S., Washington, M. K., Yeatman, T. J., Beauchamp, R. D., and Dhawan, P. (2011) Claudin-1 Upregulates the Repressor ZEB-1 to Inhibit E-Cadherin Expression in Colon Cancer Cells. *Gastroenterology*

31. Nakagawa, S., Miyoshi, N., Ishii, H., Mimori, K., Tanaka, F., Sekimoto, M., Doki, Y., and Mori, M. (2011) Expression of CLDN1 in colorectal cancer: a novel marker for prognosis. *Int J Oncol* **39**, 791-796

32. Konopka, G., Tekiela, J., Iverson, M., Wells, C., and Duncan, S. A. (2007) Junctional adhesion molecule-A is critical for the formation of pseudocanaliculi and modulates E-cadherin expression in hepatic cells. *J Biol Chem* **282**, 28137-28148

33. Zeilstra, J., Joosten, S. P., Dokter, M., Verwiel, E., Spaargaren, M., and Pals, S. T. (2008) Deletion of the WNT target and cancer stem cell marker CD44 in Apc(Min/+) mice attenuates intestinal tumorigenesis. *Cancer Res* **68**, 3655-3661

34. Zeilstra, J., Joosten, S. P., Wensveen, F. M., Dessing, M. C., Schutze, D. M., Eldering, E., Spaargaren, M., and Pals, S. T. (2011) WNT signaling controls expression of pro-apoptotic BOK and BAX in intestinal cancer. *Biochem Biophys Res Commun* **406**, 1-6

35. Liu, W., Jung, Y. D., Ahmad, S. A., McCarty, M. F., Stoeltzing, O., Reinmuth, N., Fan, F., and Ellis, L. M. (2004) Effects of overexpression of ephrin-B2 on tumour growth in human colorectal cancer. *Br J Cancer* **90**, 1620-1626

36. Batlle, E., Bacani, J., Begthel, H., Jonkheer, S., Gregorieff, A., van de Born, M., Malats, N., Sancho, E., Boon, E., Pawson, T., Gallinger, S., Pals, S., and Clevers, H. (2005) EphB receptor activity suppresses colorectal cancer progression. *Nature* **435**, 1126-1130

37. Mazzolini, R., Dopeso, H., Mateo-Lozano, S., Chang, W., Rodrigues, P., Bazzocco, S., Alazzouzi, H., Landolfi, S., Hernandez-Losa, J., Andretta, E., Alhopuro, P., Espin, E., Armengol, M., Tabernero, J., Ramon, Y. C. S., Kloor, M., Gebert, J., Mariadason, J. M., Schwartz, S., Jr., Aaltonen, L. A., Mooseker, M. S., and Arango, D. (2012) Brush border Myosin Ia has tumor suppressor activity in the intestine. *Proc Natl Acad Sci U S A* **109**, 1530-1535

38. Mazzolini, R., Rodrigues, P., Bazzocco, S., Dopeso, H., Ferreira, A. M., Mateo-Lozano, S., Andretta, E., Woerner, S. M., Alazzouzi, H., Landolfi, S., Hernandez-Losa, J., Macaya, I., Suzuki, H., Ramon, Y. C. S., Mooseker, M. S., Mariadason, J. M., Gebert, J., Hofstra, R. M., Reventos, J., Yamamoto, H., Schwartz, S., Jr., and Arango, D. (2012) Brush border myosin Ia inactivation in gastric but not endometrial tumors. *Int J Cancer*

39. Lee, N. P., Poon, R. T., Shek, F. H., Ng, I. O., and Luk, J. M. (2010) Role of cadherin-17 in oncogenesis and potential therapeutic implications in hepatocellular carcinoma. *Biochim Biophys Acta* **1806**, 138-145

40. Su, M.-C., Yuan, R.-H., Lin, C.-Y., and Jeng, Y.-M. (2008) Cadherin-17 is a useful diagnostic marker for adenocarcinomas of the digestive system. *Mod Pathol* **21**, 1379-1386

41. Li, T., Yang, W., Li, M., Byun, D. S., Tong, C., Nasser, S., Zhuang, M., Arango, D., Mariadason, J. M., and Augenlicht, L. H. (2008) Expression of selenium-binding protein 1 characterizes intestinal cell maturation and predicts survival for patients with colorectal cancer. *Mol Nutr Food Res* **52**, 1289-1299

**FIGURE LEGENDS**

**Figure 1: CpG methylation in colorectal cancer cell lines.** A) The average methylation of all the >27000 CpGs interrogated, representing >14,400 genes was variable and ranged from 0.22 to 0.43 with an overall average in all CpGs in all 45 cell lines of 0.31±0.05 (mean±SD). The unmethylated control (NoMet: whole genome amplification with GenomiPhi, GE Healthcare) and the methylated control (MET: in vitro methylated with SssI methylase; New England Biolabs) controls have an average methylation of 0.05 and 0.86, respectively. CRC cell lines range from 0.22 to 0.43. B) Correlation between the levels of methylation of >27,000 CpGs in replicate determinations for the cell line SW48 (Pearson’s r=0.99; p<0.0001). C) The distribution of the average methylation in all 45 cell lines investigated for each one of the >27,000 CpGs shows a bimodal distribution (**Mamen: can you please test this**). The majority of the CpGs investigated (51.2%) show average methylation below 20%, while 30.6% of the CpGs have average methylation levels greater than 50% in these 45 colorectal cancer cell lines. B) The overall levels of methylation were significantly (Student’s t-test p=0.0037) higher in MSI lines (n=17; 0.35±0.05) than in MSS lines (n=29; 0.3±0.04). E) There was a significant correlation (Pearson’s R=-0.39; p=0.01) between the average levels of methylation and the doubling time of 45 colon cancer cell lines. Student’s T-Test: \*p<0.05; \*\*p<0.001.

**Figure 2: Independent validation of the of the methylation and mRNA microarrays.** The levels of promoter methylation observed in the genes ITGA9 (A), KLH3 (B) and PPP1R14D (C) by bisulfite sequencing (open circle: unmethylated; filled circles: methylated) and by Illumina’s HumanMethylation27 microarrays (numbers under the BS sequencing data) are shown for the colorectal cancer cell lines indicated. D) Representative example of bisulfite sequencing assessment of methylation, showing the CpG interrogated by the arrays (asterisk) and a second CpG dinucleotide in fully methylated HCT116 and unmethylated T84 cells. A total of 8, 5 or 3 CpG sites were analyzed for the promoter of ITGA9, KLH3 and PPP1R14D, respectively, showing a consistent methylation status for these promoters. The comparison of mRNA expression levels observed by microarray and by quantitative Real-Time RT-PCR is shown for ITGA9 (E), KLH3 (F) and PPP1R14D (G).

**Supplementary Figure 1:** Association of the levels of the overall levels of methylation with frequent mutations in colorectal tumors. No associations were observed between the levels of overall methylation in >27,000 CpG sites and the mutations status of KRAS (A), BRAF (B), TP53 (C) and PIC3CA (D).

**Figure 3: CpG methylator phenotype in colorectal cancer cell lines.** A) Clustering analysis of the 45 cell lines investigated using all 27,000 CpGs interrogated identified a subgroup of cell lines with significantly higher methylation levels (lines in blue box; mean 0.35±0.04 vs. 0.31±0.05; p=0.014). B) However, no evidence of bimodality was observed in the overall methylation levels of the >27,000 CpGs for the cell lines investigated. C) Using the 643 Type C promoters (e.i., never methylated in the normal colonic mucosa) a group of cell lines with higher methylation levels was found (lines in blue box; mean 0.35±0.04 vs. 0.30±0.04; p=7.3x10-4) and panel D) shows the bimodal distribution of the levels of methylation of these 45 the cell lines .

**Supplementary Figure 2: Examples of genes with significant correlations between mRNA and methylation levels and excluded from the analysis.** The genes with significant correlations between mRNA and methylation levels that were driven by a small number of samples were exluded as described in the ‘Materials and Methods’ section. **Mamen: what are the blue dashed lines?**

**Figure 4: Representative examples of genes showing significant correlations between mRNA and methylation levels.** Panels A-D show examples of genes with significant negative correlations between mRNA expression and promoter methylation levels. A significant association was observed between high promoter methylation and low levels of expression. Panels E-H are representative examples of genes with no significant correlations between mRNA and promoter methylation levels, but that show the expected ‘L-shaped’ profile characteristic of genes whose expression is regulated by promoter methylation in this type of plots.