UNIVERSITAT DE BARCELONA

Alteracions epigenètiques en càncer colorectal:

canvis globals i identificació de noves dianes

Memòria presentada per

Jairo Rodríguez Lumbiarres

per optar al grau de

Doctor en Biologia

Tesi realitzada sota la direcció del Dr. Miquel Àngel Peinado

a l'IDIBELL-Institut de Recerca Oncològica.

Tesi adscrita al Departament de Genètica de la Facultat de Biologia

Universitat de Barcelona.

Programa de Genètica, bienni 2003-2005.

Director Tutor

Miquel À. Peinado Ricard Albalat Jairo Rodríguez

Barcelona, Novembre de 2007

OBJECTIUS

Objectius

1. Determinar la relació entre canvis globals de la metilació genòmica amb alteracions genètiques en tumors primaris humans.

2. Identificació de noves regions hipometilades en tumors colorectals amb una especial atenció sobre aquelles que continguin elements repetitius Alu. En aquest objectiu s'hi inclou tant el mapatge com la quantificació d'elements Alu desmetilats en cèl·lules normals i tumorals.

3. Identificació de noves regions hipermetilades en tumors colorectals i caracterització dels canvis epigenètics associats al silenciament.

RESULTATS

CAPÍTOL 1

Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers

Jairo Rodríguez, Jordi Frigola, Elisenda Vendrell, Rosa-Ana Risques, Mario F. Fraga, Cristina Morales, Víctor Moreno, Manel Esteller, Gabriel Capellà, Miguel A. Peinado.

Cancer Research, 2006, Vol.66 (17): 8462-8468.

Els dos primers autors han contribuït per igual en aquest treball.

Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers

Aquest treball descriptiu es centra bàsicament en l'anàlisi de la relació existent entre alteracions genètiques i epigenètiques en tumors de CCR humans. Si bé les evidències assenyalaven que podia existir una relació causal entre hipometilació i inestabilitat genètica en diferents models, aquesta relació no s'havia pogut veure de forma clara en tumors primaris humans. Un únic treball previ publicat l'any 2005 indica que existeix en tumors primaris humans de CCR una correlació positiva entre nivells d'hipometilació mesurats en elements repetitius LINE i pèrdua d'heterozigositat en determinats loci. Tot i aquests resultats, limitacions tècniques en la mesura del grau de hipometilació (els autors mesuren el nivell d'hipometilació amb una única determinació per PCR) així com la determinació d'alteracions cromosòmiques en un nombre reduït de loci limiten el impacte dels resultats obtinguts.

En aquest treball, es presenten de forma conjunta els resultats obtinguts per Jordi Frigola sobre una sèrie de 83 tumors colorectals i per mi mateix sobre una segona sèrie de 50 tumors colorectals. En la primera sèrie, l'anàlisi del grau de metilació de les mostres es realitza per AIMS. Sobre la mateixa sèrie de mostres es va calcular dany genètic per mitjà de AP-PCR. Aquesta tècnica de *fingerprinting* de DNA permet avaluar de forma ràpida i senzilla alteracions genètiques al llarg de tot el genoma amb la particularitat que no podem distingir entre mutacions puntuals, pèrdues o guanys de material genètic (fragments o cromosomes sencers) i reorganitzacions cromosòmiques. Els resultats obtinguts en aquesta primera sèrie indiquen que existeix una correlació positiva entre els nivells d'hipometilació de les mostres i el nombre d'alteracions genètiques observades per AP-PCR.

En la segona sèrie, si bé l'anàlisi del grau d'hipometilació dels tumors va ser idèntic per mitjà de AIMS, la quantificació d'alteracions genètiques va ser duta a terme per mitjà de la tècnica de *Comparative Genomic Hibridization* o CGH. Aquesta tècnica permet veure alteracions cromosòmiques relativament grans de tipus numèric, com poden ser els guanys i pèrdues de fragments cromosòmics o cromosomes sencers, així com delecions o amplificacions de fragments cromosòmics relativament grans. En aquest cas però, no podem veure alteracions estructurals balancejades, es a dir, aquelles que no impliquen guanys o pèrdues netes de material genètic. En aquest cas també vàrem observar una correlació positiva entre el grau de hipometilació en tumor i el nombre d'alteracions cromosòmiques, si bé la correlació era molt més bona que la observada en la primera sèrie de mostres. Aquest fet és indicatiu que el tipus d'inestabilitat associada a desmetilació és bàsicament de tipus cromosòmic, donat que quan mesurem les alteracions genètiques d'origen exclusivament cromosòmic la correlació observada amb el grau d'hipometilació és significativament millor que la obtinguda entre hipometilació i alteracions genètiques de divers origen mesurades per AP-PCR. La disposició de dades clínico-patològiques de la sèrie de 50 tumors colorectals va permetre establir que la correlació entre hipometilació i alteracions cromosòmiques és independent de mutacions en altres gens importants per al manteniment de la integritat del genoma com p53, fet que indica, però no demostra, que la hipometilació pot ser causa del dany cromosòmic en tumors primaris humans.

Research Article

Chromosomal Instability Correlates with Genome-wide DNA Demethylation in Human Primary Colorectal Cancers

Jairo Rodriguez,[1] Jordi Frigola,[1] Elisenda Vendrell,[1] Rosa-Ana Risques,[1] Mario F. Fraga,[3] Cristina Morales,[1] Victor Moreno,[2] Manel Esteller,[3] Gabriel Capellà,[2] Maria Ribas,[1] and Miguel A. Peinado[1]

[1]Institut d'Investigació Biomèdica de Bellvitge and [2]Institut Català d'Oncologia, L'Hospitalet, Barcelona, Catalonia, Spain; and [3]Cancer Epigenetics Laboratory, Spanish National Cancer Centre, Madrid, Spain

Abstract

DNA hypomethylation is a common trait of colorectal cancer. Studies in tumor cell lines and animal models indicate that genome-wide demethylation may cause genetic instability and hence facilitate or accelerate tumor progression. Recent studies have shown that DNA hypomethylation precedes genomic damage in human gastrointestinal cancer, but the nature of this damage has not been clearly established. Here, we show a thorough analysis of DNA methylation and genetic alterations in two series of colorectal carcinomas. The extent of DNA demethylation but not of hypermethylation (both analyzed by amplification of intermethylated sites in near 200 independent sequences arbitrarily selected) correlated with the cumulated genomic damage assessed by two different techniques (arbitrarily primed PCR and comparative genomic hybridization). DNA hypomethylation–related instability was mainly of chromosomal nature and could be explained by a genome-wide effect rather than by the concurrence of the most prevalent genetic and epigenetic alterations. Moreover, the association of *p53* mutations with genomic instability was secondary to DNA hypomethylation and the correlation between DNA hypomethylation and genomic instability was observed in tumors with and without mutation in the *p53* gene. Our data support a direct link between genome-wide demethylation and chromosomal instability in human colorectal carcinogenesis and are consistent with the studies in model systems demonstrating a role of DNA demethylation in inducing chromosomal instability. (Cancer Res 2006; 66(17): 8462-8)

Introduction

Early studies in the 1980s already identified a depletion of the methyl-cytosine content as a landmark in colorectal cancer and other types of tumors (1, 2). DNA hypomethylation has been shown to promote tumorigenesis in murine colon and liver (3) and was included as an early event in the classic Vogelstein's model for colorectal tumorigenesis (4). Different investigations sustain a causal link between DNA hypomethylation and genetic instability (reviewed in refs. 5, 6), reporting an association between defects in DNA methylation and aneuploidy in human colorectal cancer cell lines (7), increased chromosomal rearrangements in hypomethylated centromeric regions in mitogen-stimulated cells from individuals affected with immunodeficiency, centromere instability and facial anomalies (ICF syndrome; ref. 8), and an increased mutation rate owing to DNMT1 deficiency in murine embryonic stem cells (9) and in murine somatic cells (10, 11). Moreover, it has been shown that DNMT1 deficiency also results in constitutive chromosomal instability in a human colon cancer cell line (12). Fewer studies have investigated the association of DNA hypomethylation with genetic alterations in human primary cancers. Matsuzaki et al. (13) reported that decreased methylation levels in LINE sequences correlate with losses of heterozygosity on discrete chromosomal loci in a series of colorectal carcinomas. A very recent study has shown that DNA hypomethylation precedes genomic damage in gastrointestinal cancer (14). Besides these evidences supporting a close relationship between genome-wide hypomethylation and genomic damage in carcinogenesis, the nature of the genomic damage associated with DNA hypomethylation and its possible interaction with other molecular and clinicopathologic variables is still unclear.

To shed some light into this issue, we have analyzed genome-wide hypomethylation profiles in two series of colorectal carcinomas and studied its association with genomic instability. The extent of DNA hypomethylation was analyzed by amplification of intermethylated sites (AIMS; ref. 15). AIMS profiles represent unique sequences flanked by two methylated CpG sites. Differences in the display of AIMS-amplified bands between paired normal and tumor tissue correspond to changes in the methylation status of specific sequences. AIMS allows the concurrent, but independent, analysis of hypermethylations and hypomethylations in multiple samples; therefore, separate estimates of the two opposite alterations may be obtained and compared (15). Application of AIMS has been instrumental to detect relevant epigenetic changes associated with carcinogenesis (16–19) and aging (20). In a first setting, genomic instability was determined in the form of cumulated genetic alterations by arbitrarily primed PCR (AP-PCR). AP-PCR fingerprints generated from normal and tumor DNA may be easily compared, allowing the detection and characterization of multiple genetic differences (21–23) and the obtainment of unbiased estimates of global genomic disruption (24, 25). Moreover, the application of AP-PCR was instrumental in the discovery of ubiquitous microsatellite instability in a subset of colorectal tumors (26). In a second setting, comparative genomic hybridization (CGH) was used to score genomic damage at chromosomal

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

J. Rodriguez and J. Frigola contributed equally to this work. J. Rodriguez and R-A. Risques were fellows of the Comissió Interdepartamental de Recerca i Innovació Tecnològica. J. Frigola was a fellow of the Formación de Profesorado Universitario at the Universitat Autònoma de Barcelona. E. Vendrell was a fellow of the Fondo de Investigaciones Santarias. C. Morales was a fellow of Institut d'Investigació Biomèdica de Bellvitge.

Current address for Rosa-Ana Risques: Department of Pathology, University of Washington, Seattle, WA 98195-7705.

Requests for reprints: Miguel A. Peinado, Institut d'Investigació Biomèdica de Bellvitge, Granvia km 2.7, 08907 L'Hospitalet, Barcelona, Catalonia, Spain. Phone: 34-93-2607464; Fax: 34-93-2607466; E-mail: mpeinado@iro.es.

doi:10.1158/0008-5472.CAN-06-0293

level and to investigate the association of DNA hypomethylation with chromosomal profiles.

Materials and Methods

Patients and tumor samples. Two series of fresh-frozen (−80°C) colorectal carcinomas, including the paired normal tissue, were used in this study. Series HSP included 83 patients diagnosed with colorectal cancer at the Hospital de la Santa Creu i Sant Pau (Barcelona, Catalonia, Spain). Series HUB included 50 patients diagnosed of Dukes B or Dukes C colorectal cancer at the Hospital Universitari de Bellvitge (L'Hospitalet, Catalonia, Spain). These cases were part of two larger series of patients preoperatively diagnosed with colorectal cancer, and prospectively included in a study designed to evaluate the prognostic value of specific genetic and epigenetic alterations. Sample collection, patient selection criteria, and main clinicopathologic characteristics of the patients are described in Supplementary Data. Transformed cell content was >75% in most tumor specimens as assessed by histologic examination. DNA amenable for genetic and epigenetic analysis was obtained by using standard procedures.

Quantification of the degree of hypermethylation and hypomethylation. Comparative fingerprints representing the methylome of the tumor and the normal tissue were obtained by AIMS (15). AIMS bands correspond to selected genomic sequences flanked by two methylated *Sma*I sites (CCCGGG). Lack of methylation at either site prevents amplification of the band. Three independent experiments were done, resulting in the reproducible display of 208 bands that were initially considered for analysis. Assay conditions and technical validation of the approach have been described before (15, 17). Bands with age-related display in the normal tissue were excluded in normal-tumor comparisons. A total of 193 sequences were scored for differential methylation between the normal and the tumor tissue. The index of hypomethylation was calculated as the number of hypomethylated sequences (bands with a decreased intensity in the tumor compared with the normal tissue) divided by the total number of bands analyzed. Simultaneously, an index of hypermethylation (bands with an increased intensity in the tumor compared with the normal tissue) was calculated in the same way. For each AIMS-tagged band, a hypomethylation rate was calculated as the fraction of tumors showing hypomethylation with respect to the total number of informative cases.

Estimation of the genomic damage fraction. Genomic alterations at chromosomal and subchromosomal level were assessed using the DNA fingerprinting technique AP-PCR. AP-PCR is based on the amplification by PCR of genomic DNA using primers of arbitrarily chosen sequence and initial cycles of low stringency. Because the primer anneals to multiple sites, many PCR products are generated and result in a reproducible fingerprint when analyzed by gel electrophoresis. Three independent experiments were done, generating information on a total of 181 independent loci distributed genome wide (although not all of them were informative in all the samples). An index of cumulated alterations [genomic damage fraction (GDF)] was calculated as the number of bands with differential display in the tumor with regard to the paired normal tissue divided by the total number of bands visualized. Assay conditions and main associations of the GDF with genetic and clinical features of the tumors have been described in detail elsewhere (24).

Analysis of chromosomal alterations by CGH. Normal human genomic DNA (control) and tumor DNA (test) were labeled with Spectrum Green-dUTP and Spectrum Red-dUTP, respectively, by nick translation (Vysis, Downers Grove, IL). Equal amounts of control and test-labeled probes were coprecipitated and dissolved in 12 μL hybridization buffer, denatured, and hybridized in a moist chamber for 48 to 72 hours on normal metaphase spreads (Vysis). Slides were washed according to the protocol of the manufacturer. Chromosomes were counterstained with 4,6-diamino-2-phenylindole (Sigma, St. Louis, MO) and analyzed using Cytovision Ultra workstation (Applied Imaging, Sunderland, United Kingdom). Ratio values obtained from at least 10 metaphases were averaged. Ratio values >1.25 and <0.75 were considered to represent chromosomal gains and loses, respectively.

Competitive hybridization of AIMS products to metaphase chromosomes. Unrestrained AIMS were done as described (15) in 10 normal tumor pairs. Products were purified using Concert Rapid Purification System (Life Technologies, Carlsbad, CA) and labeled with SpectrumRed dUTP or SpectrumGreen dUTP (Vysis) using a nick translation kit (Vysis). Both normal and tumor probes from the same patient were cohybridized to metaphase chromosomes. Procedures and image analysis were done as described for conventional CGH (27).

High-performance capillary electrophoresis. 5-Methylcytosine content was determined by high-performance capillary in 16 tumors exhibiting a wide range of hypomethylation index electrophoresis and their paired normal tissue, as previously described (28). Two cases were excluded due to technical failure; therefore, 14 cases remained available for analysis. Briefly, genomic DNA samples were boiled and treated with nuclease P1 (Sigma) for 16 hours at 37°C and with alkaline phosphatase (Sigma) for an additional 2 hours at 37°C. After hydrolysis, total cytosine and 5-methylcytosine content was measured by capillary electrophoresis using a P/ACE MDQ system (Beckman-Coulter, Fullerton, CA) with UV detection. Relative 5-methylcytosine content was expressed as a percentage with respect to the total cytosine content (methylated and nonmethylated).

Statistical analysis. Results are expressed as a mean ± SD, except where indicated. Statistical differences between variables were analyzed with unpaired/paired *t* or nonparametric tests as appropriate. Correlation

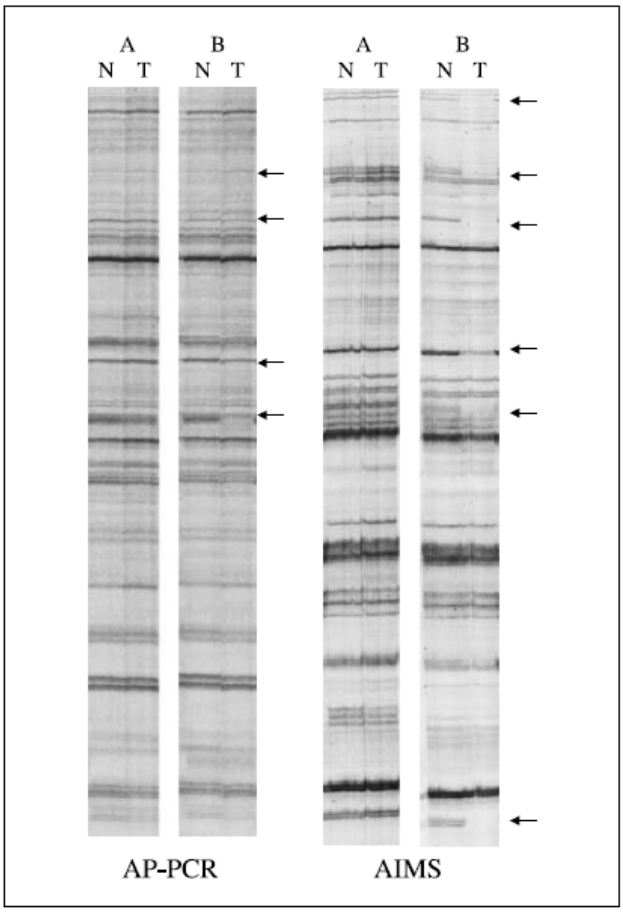


Figure 1. Illustrative examples of genomic damage assessment by AP-PCR (*left*) and DNA methylation changes by AIMS (*right*). Fingerprints from two matched normal (*N*)-tumor (*T*) pairs (*patients A and B*). Differences of intensity in the tumor with regard to the paired normal tissue (*arrowheads*) were scored as gains/losses (in AP-PCR) or hypermethylations/hypomethylations (in AIMS).

Table 1. Genomic damage and hypomethylation index in colorectal cancer

	Series HSP			Series HUB		
	n	GDF	Hypomethylation index	*n*	No. alterations	Hypomethylation Index
All tumors	83	0.172 ± 0.085 (0.015-0.389)	0.149 ± 0.082 (0.005-0.356)	50	6.8 ± 4.8 (0-16)	0.122 ± 0.062 (0.017-0.272)
p53 status						
Wild-type	46	0.150 ± 0.082	0.142 ± 0.076	16	3.7 ± 4.1	0.102 ± 0.058
Mutated	33	0.202 ± 0.080	0.164 ± 0.091	32	8.4 ± 4.3	0.128 ± 0.061
		$P = 0.007$	$P = 0.245$		$P = 0.001$	$P = 0.162$

NOTE: Values are expressed as mean ± SD. Numbers in parentheses indicate range. *P* values are estimated from two-tailed *t* test.

between quantitative variables was calculated with Pearson's coefficient and linear regression models were used to estimate the effects. Nonparametric Spearman's correlation coefficient was also used, but only Pearson's is reported because the results were essentially similar. Normality of the data was tested with Kolmogorov's test. All *P* values are calculated from two-sided statistical tests.

The comparison of the correlation coefficients between the hypomethylation index and GDF with that of the hypomethylation index and the number of chromosomal alterations was done with resampling techniques, quantifying the proportion of bootstrap samples in which the first correlation was greater than the second, taking into account that different tumors were used for each study. This method accounts for potential outliers in the data and nonnormality.

Results and Discussion

Cumulated genomic damage correlates with genome-wide demethylation. In a first setting, we analyzed abnormalities at DNA methylation and genomic level using two DNA fingerprinting methods. DNA hypermethylation and hypomethylation profiles were analyzed by AIMS in 83 colorectal carcinomas and their paired normal tissues (referred here as series HSP). The extent of genomic alterations was assessed by AP-PCR (24). Illustrative examples of the AIMS and AP-PCR fingerprints are shown in Fig. 1. HSP series did not include any case with microsatellite instability to avoid the contribution of microsatellite-related instability in the calculation of the genomic damage index.

DNA methylation profiles generated by AIMS rendered an average of 185 informative bands per case. In AP-PCR experiments, an average of 141 ± 27 informative loci were scored per case. Indices of hypomethylation and hypermethylation and the index of genomic alterations (GDF) were calculated as the fraction of bands with differential display between paired normal and tumor according to AIMS (DNA methylation changes) and AP-PCR fingerprints (GDF). DNA hypomethylation indices and GDF are shown in Table 1. Tumors with *p53* mutations exhibited higher levels of GDF but not of DNA hypomethylation (Table 1). No differences in the GDF and hypomethylation indices were observed when tumors were classified according to clinicopathologic features (Supplementary Table S1), except for sex: Men showed higher hypomethylation levels than women as previously reported (17).

A correlation between GDF and the hypomethylation index ($r = 0.250$, $P = 0.022$; Fig. 2*A*), but not the hypermethylation index ($r = 0.020$, $P = 0.856$; Supplementary Fig. 1*A*), was observed. Therefore, it can be concluded that demethylation, but not *de novo* DNA methylation, correlates with the number of genetic alterations in colorectal cancer. In a recent study in gastrointestinal cancer, Suzuki et al. (14) have reached similar conclusions demonstrating the dominant role of hypomethylation over hypermethylation on genomic damage.

Classification of the cases according to clinicopathologic variables revealed that the correlation was strengthen in younger patients (≤67 years) and men (Supplementary Table S2). Because age-related hypomethylation and hypermethylation have been reported in normal mucosa (14, 29), it can be speculated that the epigenetic "noise" associated with aging is lower in younger patients, bearing an undisturbed and enhanced correlation between the epigenetic and genetic damage.

DNA hypomethylation–related genomic instability is of chromosomal nature. The genetic alterations revealed by AP-PCR may be of multiple and different nature (structural and numerical chromosomal alterations, point mutations, short insertions, and deletions; reviewed in ref. 30). In the first study, we excluded tumors exhibiting microsatellite instability. Microsatellite instability is also detectable in AP-PCR fingerprints (21, 26) and the mechanisms underlying this type of instability are well known. Therefore, we knew that the genomic damage detected by AP-PCR did not include microsatellite instability. However, we wondered if the genomic instability associated with hypomethylation was the reflection of multiple types of genomic damage or rather of a specific type of alteration. Hence, we analyzed DNA methylation profiles by AIMS and genomic alterations by CGH in a second series of 50 colorectal carcinomas (series HUB). The indices of DNA methylation and genomic instability are shown in Table 1 and Supplementary Table S1. Noteworthy, the correlation between the hypomethylation index and the number of chromosomal alterations was highly significant ($r = 0.514$, $P < 0.001$; Fig. 2*B*). We also investigated if the type of chromosomal alteration (structural, numerical, gains, and losses) was differentially associated with the hypomethylation index. In all cases, the correlation was highly significant, as it was among the measures of each category of chromosomal alteration (Supplementary Table S3), suggesting that chromosomal instability is manifested in different types of alterations concurrently and that DNA hypomethylation does not associate preferentially with one specific kind of chromosomal damage.

The enhancement in the correlation between DNA hypomethylation and cumulated genomic damage determined by CGH (compared with that obtained from AP-PCR data) suggests that the main type of genomic instability resulting directly from DNA hypomethylation is most probably of chromosomal nature. By using resampling techniques, we calculated that the correlation hypomethylation index/number of chromosome alterations is

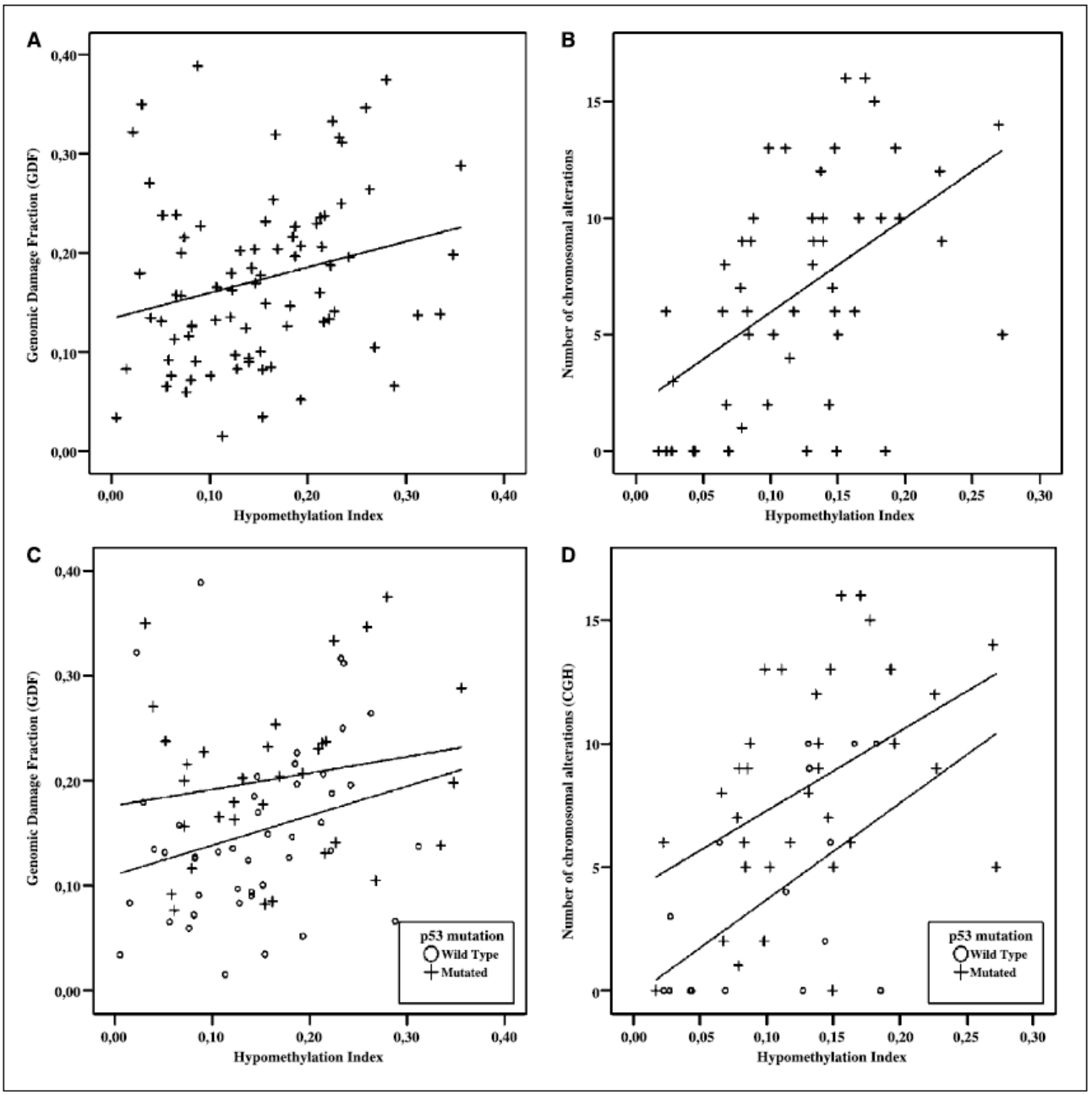


Figure 2. Scatterplot of the distribution of the hypomethylation index and the cumulated genomic damage determined by two different techniques in colorectal carcinomas. *Box A,* GDF detected by AP-PCR in HSP series ($n = 83$). *Box B,* number of chromosomal alterations detected by CGH in the HUB series ($n = 50$). *Boxes C and D,* multiple regression analysis after categorization of tumors by the *p53* mutational status. *p53* mutation showed an additive effect in the number of genetic alterations (GDF, *box C*; number of chromosomal alterations, *box D*) to the hypomethylation index. *Top regression line, p53* mutated tumors (*crosses*); *bottom regression line,* wild-type *p53* tumors (*open circles*).

greater than the correlation hypomethylation index/GDF with a 95% confidence. The method used to calculate this value was based only on the observed data and took into account that different samples of patients were used to analyze each measure of genomic damage. Even if we exclude three cases of the HSP series that show very high values of GDF and low DNA hypomethylation index, the probability that the correlation is greater for the number of chromosome alterations is 80%. The positive correlation was maintained when tumors were classified by clinicopathologic features, although some differences could be observed according to sex and tumor location (Supplementary Table S2). Similar to what we observed in series HSP, the hypermethylation index showed no correlation with cumulated genomic alterations (Supplementary Fig. S1*B*).

Genomic extent of hypomethylation related instability. To determine which chromosomes were more affected by DNA

demethylation, we compared, for each chromosome alteration, the hypomethylation index between tumors with and without any given alteration. For most of the chromosomes, the hypomethylation index was higher in tumors exhibiting the specific alteration than tumors without (Fig. 3). Representation of the individual data illustrated that alterations for any given chromosome tended to occur more frequently in tumors with high hypomethylation levels (Fig. 3). This result suggests that DNA hypomethylation is likely to affect the stability of all the chromosomes, and that the differences in the frequency of alterations among chromosomes are rather a consequence of differential selective pressure (31).

Finally, to exclude a possible effect of chromosomal alterations on the determination of differential methylation indexes, we analyzed gross DNA methylation changes at chromosomal level by competitive hybridization of AIMS products to metaphase chromosomes in 10 tumors. This approach allows the detection of hypomethylations and hypermethylations at a genomic scale (16, 20). The differential methylation profiles were compared with the chromosomal profiles of losses and gains determined by CGH and no consistent parallelism between genetic and epigenetic alterations (gains with hypermethylation and losses with hypomethylation) was observed (Supplementary Figs. S2 and S3), indicating that AIMS changes are unaffected by chromosomal copy number alterations.

Demethylation variables that correlate with chromosomal instability. Hypomethylations detected by AIMS are broadly distributed throughout the genome (15, 17) and its computation represents a natural measure of demethylating events occurring in the tumor cells compared with its paired normal tissue. This measure, when transformed in an index, represents an estimation of genome-wide hypomethylations. Because AIMS screening through the genome is uneven, with a more intense representation of euchromatin and gene-rich regions (15, 17), the hypomethylation index calculated from AIMS fingerprints cannot be considered a representation of global hypomethylation, which also affects heterochromatin (32, 33). To get insights into the DNA methylation variables that are more likely to play a role in chromosomal instability, we did global and specific analyses.

At global level, we determined 5-methylcytosine content in normal and tumor tissue in a short subset of cases. 5-Methylcytosine content was lower in tumors (normal 3.3 $\pm$ 1.0; tumor 2.7 $\pm$ 0.4; Wilcoxon test $P = 0.026$), but neither the relative loss of 5-methylcytosine (ratio of 5-methylcytosine content in the tumor with respect to the paired normal tissue) nor the tumor 5-methylcytosine showed statistically significant correlation with the hypomethylation index determined by AIMS (Supplementary Fig. S4).

Interestingly, the number of chromosomal alterations, although showing a good correlation with the hypomethylation index, showed no correlation with any of the 5-methylcytosine variable considered: global loss of 5-methylcytosine (T/N ratio), normal 5-methylcytosine content, and tumor 5-methylcytosine content (Supplementary Fig. S5). We think that these results are not in contradiction with previous studies done in model systems defective in DNMT1 and demonstrating a relationship between global hypomethylation and genomic instability (3, 10–12). The content of 5-methylcytosine in the tumor is consequence of the initial 5-methylcytosine content (that varies in a wide range from

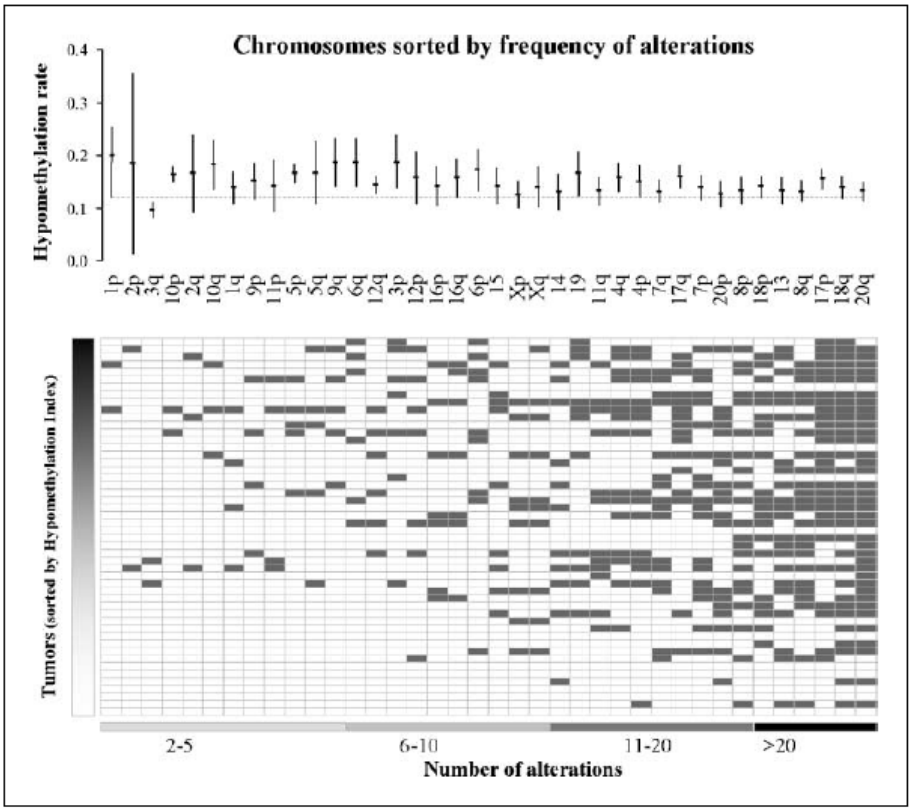


Figure 3. Effect of DNA hypomethylation levels on individual chromosome instability. *Top*, level of hypomethylation in tumors showing alterations in each chromosome or chromosome arm as analyzed by CGH. *Points*, mean; *bars*, 95% confidence interval. Chromosomes have been arranged by frequency of alterations (*left→right*). *Horizontal dashed line*, mean of the hypomethylation level in all tumors. *Bottom*, distribution of chromosomal alterations in the 50 tumors arranged by the hypomethylation index (*down→up*). Chromosomes have been arranged by frequency of alterations from left (*low*) to right (*high*).

individual to individual, from 2.5 to 6.3%) and two opposite and independent processes: demethylation and *de novo* methylation. Therefore, the lack of correlation in our assessment was predictable.

Because AIMS bands showed a wide distribution in the rate of hypomethylation (0% to 78%), we wondered if the effect of hypomethylation on chromosomal instability could be attributed to a fraction of the sequences represented in the AIMS fingerprints, or, alternatively, it was a cooperative effect. We compared the number of chromosomal alterations in tumors with hypomethylation versus no hypomethylation for each one of the AIMS bands showing recurrent hypomethylation (in six or more tumors, 12%). No trends were observed between the hypomethylation rate (fraction of tumors showing hypomethylation for any given band) and the number of chromosomal alterations (Supplementary Fig. S6). This implies that the association of hypomethylation with chromosomal instability results from the cumulated effect of multiple independent loci distributed genome wide (Supplementary Fig. S7). Therefore, it can be concluded that the observed correlations are explained by a genome-wide cumulative effect rather than by the concurrence of the most prevalent genetic and epigenetic alterations.

It should be noted that previous studies have attributed the global DNA hypomethylation observed in cancer cells to a demethylation of repetitive elements and heterochromatin (32). Nevertheless, most of the hypomethylated sequences isolated from AIMS fingerprints are of heterogeneous nature and map in gene-rich regions (15, 17),[4] supporting the notion that DNA hypomethylation also involves unique sequences and euchromatin.

DNA hypomethylation and *p53* mutations define the degree of chromosomal instability. It is well known that inactivation of the tumor suppressor gene *p53* also plays a pivotal role in genetic instability (34). In agreement with previous studies, we observed increased rates of genomic alterations in tumors displaying mutations in the *p53* gene (Table 1) in both series of cases. On the other hand, *p53* mutations were not associated with a higher hypomethylation index. To fully dissect the possible interaction, we did regression analysis of genomic damage and hypomethylation index, including the *p53* status in the model. The presence of *p53* mutations resulted in a net increase of the genomic instability (0.047 ± 0.018 GDF units or 3.5 ± 1.2 chromosomal alterations) compared with tumors with wild-type *p53*. In both cases, the correlation between DNA hypomethylation and genomic instability was maintained (Fig. 2*C-D*; Supplementary Table S2). The *P* value for the differences in the linear relationship according to *p53* was 0.86 for GDF and 0.69 for CGH.

Whereas we do not show a causal relationship between DNA hypomethylation and genomic instability, an issue that has already been addressed in cell lines and murine models (see above), our results support the hypothesis that chromosomal instability is a direct outcome of DNA hypomethylation. Hence, p53 inactivation does not seem as the primary cause of genomic instability (35), but would contribute by releasing instability repression, which is manifested as an increase in the number of genetic alterations. This scenario is compatible with the early appearance of DNA demethylation in tumor progression (17, 33, 36).

Implications. This is a comprehensive study assessing the effect of genome-wide epigenetic deregulation on genomic instability in human colorectal carcinogenesis. Our results are consistent with a critical role of genome-wide hypomethylation in carcinogenesis. Its early contribution to genomic instabilization is supported by previous studies in model systems and hereditary cancer. DNA hypomethylation–induced chromosomal instability boosts the rate of losses of heterozygosity at the *APC* locus (the second hit in Knudson's model) in APC(Min/+) DNMT1 hypomorphic mice (with reduced genomic methylation), which results in an increased number of microadenomas (3). Moreover, a recent study suggests that chromosomal instability caused by another early alteration, *APC* mutation, is likely to be subtle and that chromosomal instability may even precede adenoma onset in polyposis patients (37).

Cumulated data have outlined tumor progression as alternative genetic pathways in colorectal cancer (24, 38, 39). Each pathway displays particular molecular profiles and confers differentiated biological and clinical behavior. Based on our results, we hypothesize that DNA hypomethylation is likely to induce a cascade effect with direct implications in the determination of the progression pathway, and hence the patient's outcome. Future studies should dissect the interactions and the cause-consequence order of the different factors involved.

[4] Unpublished data.

Acknowledgments

Received 1/24/2006; revised 5/26/2006; accepted 6/26/2006.

Grant support: Ministry of Education and Science (SAF03/5821).

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.

We thank Gemma Aiza and Mar Muñoz for technical support.

References

1. Gama-Sosa MA, Slagel VA, Trewyn RW, et al. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 1983;11:6883–94.
2. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301:89–92.
3. Yamada Y, Jackson-Grusby L, Linhart H, et al. Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. Proc Natl Acad Sci U S A 2005;102:13580–5.
4. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759–67.
5. Costello JF, Plass C. Methylation matters. J Med Genet 2001;38:285–303.
6. Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer 2004;4:143–53.
7. Lengauer C, Kinzler KW, Vogelstein B. DNA methylation and genetic instability in colorectal cancer cells. Proc Natl Acad Sci U S A 1997;94:2545–50.
8. Xu GL, Bestor TH, Bourchis D, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 1999;402:187–91.
9. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature 1998;395:89–93.
10. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science 2003;300:455.
11. Gaudet F, Graeme JG, Eden A, et al. Induction of tumors in mice by genomic hypomethylation. Science 2003;300:489–92.
12. Karpf AR, Matsui S. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer Res 2005;65:8635–9.
13. Matsuzaki K, Deng G, Tanaka H, et al. The Relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. Clin Cancer Res 2005;11:8564–9.
14. Suzuki K, Suzuki I, Leodolter A, et al. Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. Cancer Cell 2006;9:199–207.

15. Frigola J, Ribas M, Risques RA, Peinado MA. Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). Nucleic Acids Res 2002; 30:e28.
16. Paz MF, Wei S, Cigudosa JC, et al. Genetic unmasking of epigenetically silenced tumor suppressor genes in colon cancer cells deficient in DNA methyltransferases. Hum Mol Genet 2003;12:2209–19.
17. Frigola J, Sole X, Paz MF, et al. Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer. Hum Mol Genet 2005;14:319–26.
18. Frigola J, Muñoz M, Clark SJ, et al. Hypermethylation of the prostacyclin synthase (PTGIS) promoter is a frequent event in colorectal cancer and associated with aneuploidy. Oncogene 2005;24:7320–6.
19. Frigola J, Song J, Stirzaker C, et al. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat Genet 2006;38:540–9.
20. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 2005;11:11.
21. Peinado MA, Malkhosyan S, Velazquez A, Perucho M. Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. Proc Natl Acad Sci U S A 1992;89:10065–9.
22. Arribas R, Risques RA, Gonzalez-Garcia I, et al. Tracking recurrent quantitative genomic alterations in colorectal cancer: allelic losses in chromosome 4 correlate with tumor aggressiveness. Lab Invest 1999; 79:111–22.
23. Piao Z, Lee KS, Kim H, Perucho M, Malkhosyan S. Identification of novel deletion regions on chromosome arms 2q and 6p in breast carcinomas by amplotype analysis. Genes Chromosomes Cancer 2001;30:113–22.
24. Risques RA, Moreno V, Ribas M, et al. Genetic pathways and genome-wide determinants of clinical outcome in colorectal cancer. Cancer Res 2003;63:7206–14.
25. Suzuki K, Ohnami S, Tanabe C, et al. The genomic damage estimated by arbitrarily primed PCR DNA fingerprinting is useful for the prognosis of gastric cancer. Gastroenterology 2003;125:1330–40.
26. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993;363:558–61.
27. Masramon L, Ribas M, Cifuentes P, et al. Cytogenetic characterization of two colon cell lines by using conventional G-banding, comparative genomic hybridization, and whole chromosome painting. Cancer Genet Cytogenet 2000;121:17–21.
28. Fraga MF, Uriol E, Borja Diego L, et al. High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. Electrophoresis 2002;23:1677–81.
29. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 1998;58:5489–94.
30. Risques RA, Ribas M, Peinado MA. Assessment of cumulated genetic alterations in colorectal cancer. Histol Histopathol 2003;18:1289–99.
31. Gorringe KL, Chin SF, Pharoah P, et al. Evidence that both genetic instability and selection contribute to the accumulation of chromosome alterations in cancer. Carcinogenesis 2005;26:923–30.
32. Plass C. Cancer epigenomics. Hum Mol Genet 2002; 11:2479–88.
33. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene 2002;21:5400–13.
34. Janus F, Albrechtsen N, Dornreiter I, et al. The dual role model for p53 in maintaining genomic integrity. Cell Mol Life Sci 1999;55:12–27.
35. Duensing A, Duensing S. Guilt by association? p53 and the development of aneuploidy in cancer. Biochem Biophys Res Commun 2005;331:694–700.
36. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science 1985;228:187–90.
37. Cardoso J, Molenaar L, de Menezes RX, et al. Chromosomal instability in MYH- and APC-mutant adenomatous polyps. Cancer Res 2006;66:2514–9.
38. Olschwang S, Hamelin R, Laurent-Puig P, et al. Alternative genetic pathways in colorectal carcinogenesis. Proc Natl Acad Sci U S A 1997;94:12122–7.
39. Hermsen M, Postma C, Baak J, et al. Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. Gastroenterology 2002;123:1109–19.

Supplemental Material

PATIENTS

Two series of patients preoperatively diagnosed of colorectal cancer and operated upon with curative or palliative intention were used in this study. HSP series consisted of 83 patients from the Hospital de Sant Pau (Barcelona, Spain). HUB series consisted of 50 patients from the Ciutat Sanitària i Universitària de Bellvitge (L'Hospitalet, Barcelona, Spain). Patients were prospectively included in a study specifically designed to evaluate the prognostic value of genetic alterations in colorectal cancer and to evaluate the relationships between diet and molecular alterations. No chemo- or radiotherapy was given prior to surgery in these patients. Inclusion in the study did not influence the adjuvant treatment given. The study protocol was approved by the Ethics Committee.

Inclusion criteria in both series were: (a) electively resected primary adenocarcinomas; (b) the obtention of fresh paired normal mucosa-tumor samples within two hours after tumor removal; (c) availability of high quality DNA from paired normal and tumor tissue; and (d) no postoperative death. Additionally, HUB series only included Dukes B and C stage tumors and series HSP only included tumors negative for microsatellite instability (83 from an original collection of 93 cases) to avoid the detection of microsatellite related instability in the determination of genomic damage by Arbitrarily Primed PCR (AP-PCR) fingerprints. In series HUB it was not necessary to exclude tumors with microsatellite instabity (n=3) because it is not detected by Comparative Genomic Hybridization.

Cases were pathologically staged using Astler-Coller modification of Dukes' classification system. Surgical specimens were collected at the operating room and immediately taken to the Pathology Department in ice. Carcinomas and paired normal samples were snap frozen within two hours after removal and then stored at -80ºC.

Supplemental Table 1. Indices of Genomic Damage and DNA hypomethylation in colorectal carcinomas classified by clinicopathological features[a].

	Series HSP			Series HUB		
	n	**GDF**	**Hypomethylation Index**	**n**	**No of Alterations**	**Hypomethylation Index**
All tumors	83	0.172 ± 0.085	0.149 ± 0.082	50	6.8 ± 4.8	0.122 ± 0.062
Age						
≤ 67	36	0.169 ± 0.084	0.162 ± 0.084	28	5.8 ± 5.0	0.116 ± 0.063
> 67	47	0.175 ± 0.086	0.138 ± 0.080	22	8.1 ± 4.3	0.129 ± 0.060
		p=0.744	p=0.200		p=0.455	p=0.093
Sex						
Female	37	0.175 ± 0.092	0.125 ± 0.083	24	6.0 ± 4.5	0.122 ± 0.064
Male	46	0.168 ± 0.077	0.168 ± 0.077	26	7.6 ± 5.0	0.122 ± 0.061
		p=0.814	**p=0.016**		p=0.241	p=0.978
Dukes' stage						
A-B	44	0.178 ± 0.088	0.157 ± 0.079	29	6.4 ± 5.2	0.122 ± 0.062
C-D	39	0.165 ± 0.081	0.139 ± 0.085	21	7.4 ± 4.3	0.122 ± 0.063
		p=0.494	p=0.336		p=0.506	p=0.999
Localization						
Right	20	0.145 ± 0.076	0.139 ± 0.056	15	6.5 ± 4.6	0.125 ± 0.057
Left	63	0.181 ± 0.086	0.152 ± 0.089	35	7.5 ± 5.5	0.120 ± 0.064
		p=0.102	p=0.554		p=0.512	p=0.788

Supplemental Table 2. Correlation between genomic damage and the hypomethylation index in colorectal cancers classified according to clinicopathological features

	Series HSP			Series HUB		
Variable	**n**	**Correlation[a]**	**p**	**n**	**Correlation[a]**	**p**
All tumors	83	0.250	**0.022**	50	0.514	**<0.001**
Age						
≤ 67	36	0.519	**0.001**	23	0.530	**0.009**
> 67	47	0.057	0.715	27	0.538	**0.004**
Sex						
Female	37	0.241	0.151	24	0.342	0.102
Male	46	0.293	**0.048**	26	0.681	**<0.001**
Dukes' stage						
A-B	44	0.185	0.228	29	0.443	**0.016**
C-D	39	0.313	0.052	21	0.647	**0.002**
Localization						
Right	20	0.399	0.081	15	0.288	0.297
Left	63	0.218	0.086	35	0.618	**<0.001**
p53 status						
wild type	46	0.263	0.077	17	0.644	**0.005**
mutated	33	0.177	0.323	33	0.425	**0.014**

[a] Pearson's correlation.

Supplemental Table 3. Correlation[a] among the Hypomethylation index and the number of different types of chromosmal alterations

	Losses	Gains	Numerical	Estructural
Hypomethylation index	0.530 **<0.001**	0.427 **0.002**	0.3610 **0.010**	0.510 **<0.001**
Losses		0.748 **<0.001**	0.743 **<0.001**	0.845 **<0.001**
Gains			0.886 **<0.001**	0.726 **<0.001**
Numerical				0.489 **<0.001**

[a] Pearson's coefficient (p value is shown bold) (n=50).

CAPÍTOL 2

Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells

Jairo Rodríguez, Laura Vives, Mireia Jordà, Cristina Morales, Mar Muñoz, Elisenda Vendrell, Miguel A. Peinado

Article enviat a la revista Nucleic Acids Research.

Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells

Durant dècades els elements repetitius de DNA, considerats els paràsits més senzills, han fascinat als investigadors, els quals han intentat desxifrar quin pot ser el paper d'aquests elements en altres temps considerats DNA "escombraries" (junk DNA). Tot i que la seqüenciació del genoma humà ens ha proporcionat dades molt precises del seu nombre i distribució, desconeixem en gran mesura l'aportació d'aquestes seqüències a la regulació epigenètica en un context fisiològic normal i més important, com la seva desregulació epigenètica contribueix a la desestabilització del genoma i l'aparició de malalties com el càncer. Basant-nos en mètodes preexistents, hem desenvolupat una nova metodologia que permet tant la quantificació com la identificació i el mapatge d'elements Alu desmetilats en el genoma humà.

L'aplicació d'aquesta metodologia a una sèrie de tumors colorectals i les seves parelles de teixit normal ens ha permès obtenir dades del nombre d'Alu's desmetilades en teixit normal i tumoral així com la identificació d'un gran nombre de regions recurrentment hipermetilades i hipometilades en tumors colorectals. Les dades obtingudes indiquen que el nombre d'elements Alu que de forma normal tenen la seva diana SmaI desmetilada està al voltant dels 25000 elements, mentre que aquesta figura arriba a quasi doblar-se en teixit tumoral, indicant que la hipometilació global que observem en tumors té un fort impacte sobre l'estat de metilació dels elements Alu, com prèviament estimes globals havien indicat. El sol fet que hi hagi un nombre tant elevat d'elements Alu desmetilats és un resultat sorprenent, donat que aquests sempre han estat considerats com els elements repetits dispersos més fortament metilats del genoma. L'obtenció de patrons reproduïbles ens indica que hi ha en la cèl·lula normal un nombre elevat d'elements repetitius l'estat desmetilat dels quals està consistentment mantingut, fet que reforça la seva singularitat dins del conjunt d'elements repetitius del genoma.

L'estudi en profunditat d'una de les regions hipometilades, la qual conté dos elements repetitius, indica que els canvis de la metilació que afecten a aquests elements durant el procés tumoral estan associats a canvis en les histones que podrien, al seu torn, afectar l'estabilitat de la regió hipometilada. Aquestes dades són consistents amb la hipòtesi que els elements repetitius desmetilats proporcionen un mecanisme inductor d'inestabilitat genètica a través de canvis en l'estructura de la cromatina.

El gran nombre d'alteracions de la metilació detectats per mitjà d'aquest mètode suggereix una gran plasticitat dels patrons de metilació d'elements Alu durant el procés tumoral, fet que requerirà investigacions futures dirigides a aclarir el paper d'aquests i altres elements repetitius en l'alteració dels patrons de metilació en la cèl·lula tumoral, la inducció d'inestabilitat genètica i el silenciament transcripcional.

Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells

Jairo Rodriguez[1], Laura Vives[2,3], Mireia Jordà[1], Cristina Morales[1], Mar Muñoz[2,3], Elisenda Vendrell[1] and Miguel A. Peinado[1,2]

[1] Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet, Catalonia, Spain
[2] Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC), Badalona, Catalonia, Spain
[3] Institut Català d'Oncologia (ICO), L'Hospitalet, Catalonia, Spain

Running title: Tracking of unmethylated Alu
Keywords: interspersed repeats, DNA methylation, colorectal cancer

Correspondence: Miguel A. Peinado
Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC)
Cr. Can Ruti, Camí de les Escoles s/n
08916 Badalona, Catalonia, Spain
email: map@imppc.org
Tel 34-934978693
Fax 34-934978697

ABSTRACT

Methylation of the cytosine is the most frequent epigenetic modification of DNA in mammalian cells. In humans, most of the methylated cytosines are found in CpG-rich sequences within tandem and interspersed repeats that make up to 45% of the human genome, being Alu repeats the most common family. Demethylation of Alu elements occurs in aging and cancer processes and has been associated with gene reactivation and genomic instability. By targeting the unmethylated SmaI site within the Alu sequence as a surrogate marker we have quantified and identified unmethylated Alu elements on the genomic scale. Normal colon epithelial cells contain in average 25486±10157 unmethylated Alu's per haploid genome, while in tumor cells this figure is 41995±17187 (p=0.004). There is an inverse relationship in Alu families with respect to their age and methylation status: the youngest elements exhibit the highest prevalence of the SmaI site (AluY: 42%; AluS: 18%, AluJ: 5%) but the lower rates of unmethylation (AluY: 1.65%; AluS: 3.1%, AluJ: 12%). Data are consistent with a stronger silencing pressure on the youngest repetitive elements, which are closer to genes. Further insights into the functional implications of atypical unmethylation states in Alu elements will surely contribute to decipher genomic organization and gene regulation in complex organisms.

Funding

This work was supported by a grant from the Ministry of Education and Science (SAF2006/351) and the Consolider-Ingenio 2010 Program (CSD2006-49).

INTRODUCTION

Progress in large-scale sequencing projects is critical to identify and decipher gene organization and regulation in many species including human. Nevertheless, cumulated evidences indicate that the complexity of living organisms is not just a direct outcome of the number of coding sequences and that the presence of multiple regulatory mechanisms accounts for a significant part of biological complexity (1,2). Among these mechanisms, repetitive elements may play a key role in gene regulation and genomic structure. Active transposable elements are involved in genome rearrangement and illegitimate recombination and can also influence gene expression by altering splicing or by acting as enhancers or promoters (3-7). Advances in the understanding of epigenetic mechanisms that regulate this repetitive elements may contribute to elucidate their specific participation in biological processes (8).

Silenced regions in mammals and other vertebrates are differentiated, although not exclusively, by the presence of DNA methylation (reviewed in (9)). Methylation of the cytosine is an epigenetic modification of DNA that plays an important role in the control of gene expression and chromosome structure in mammalian cells (reviewed in (10-13)). Most of the 5-methylCytosines are found in CpG-rich sequences within tandem and interspersed repeats (9,12) which previous estimates indicate that constitute up to 45% of the human genome (14). Among these repeats, Alu's, with more than 1 million copies per haploid genome, are considered the most successful family (15). Interestingly, Alu's are not randomly distributed within the human genome as they tend to accumulate in gene rich regions (14,16,17). Previous works have estimated that Alu elements harbor up to 33% of the total number of CpG sites in the genome (18) and have been reported to be highly methylated in most somatic tissues (18-20). Methylation represents the primary mechanism of transposon suppression and active transposons are demethylated in mammalian genomes (12). It has been proposed that regions of the genome containing repetitive elements might be masked by compartmentalization of the chromatin, resulting in a reduction of the effective size of the genome (21).

Noteworthy, even though a vast number of CpG dinucleotides are provided by the collection of repetitive sequences in the human genome, this dinucleotide is greatly under-represented throughout the genome, but it can be found at close to its expected frequency in small genomic regions (200bp to a few kb), known as CpG islands (22). These areas are "protected" from methylation and are located in the proximal promoter regions of 75% of human genes (12,13,22). Methylated CpG islands are strongly and hereditably repressed (12). Hence DNA methylation is usually considered as a sign of long-term inactivation (9,10,12).

Cancer cells are characterized by the accumulation of both genetic and epigenetic changes. Widespread genomic hypomethylation is an early alteration in carcinogenesis and has been associated with genomic disruption and genetic instability (23-27). Repeats unmasked by demethylation are likely to facilitate rearrangements due to mitotic recombination and unwanted transcription (28-30). Alternatively, aberrant *de novo* methylation of CpG islands is a hallmark of human cancers and is associated with epigenetic silencing of multiple tumor suppressor genes (31-37). Therefore, the screening for differentially methylated sequences in

tumors appears as a key tool to further understand the molecular mechanisms underlying malignant transformation of cells. Although, the repertoire of methylation screening methodologies has expanded widely (37-39), and different approaches have been used to make bulk estimates of methylation in repetitive elements (40,41), there is still a lack of screening strategies that specifically allow a feasible identification of DNA methylation alterations in repetitive elements (21).

Here we report two variants of a novel methodology to quantify and identify unmethylated Alu sequences. The CpG site within the consensus Alu sequence AACCCGGG is used as a surrogate reporter of methylation. Unmethylated sites are cut with the methylation-sensitive restriction endonuclease SmaI (CCCGGG) and an adaptor is ligated to the DNA ends. Quantification of UnMethylated Alus (QUMA) is performed by real-time amplification of the digested and adaptor-ligated DNA using an Alu consensus primer that anneals upstream of the SmaI site and an adaptor primer extended with the TT dinucleotide in its 3' end (Figure 1A). The product generated by this approach is completely inside the Alu element and hence it is not possible to make a unique identification. As an alternative approach, we have also performed restrained amplification of digested and adaptor-ligated DNA fragments that are flanked by two close SmaI sites. In this case, the same primer homologous to the adaptor with the additional TT nucleotides at the 3' end to enrich for Alu sequences is used in absence of the Alu consensus primer (Figure 1B). This second approach is named Amplification of UnMethylated Alu's (AUMA) and results in a complex representation of unique DNA sequences flanked by two unmethylated SmaI sites. When resolved by high resolution electrophoresis, the AUMA generated sequences appear as a fingerprint characteristic of each sample (Figure 2) and individual scoring and identification of each band can be performed. Because AUMA's stringency is based on a short sequence (AACCCGGG) that is found preferentially but not exclusively in Alu elements, other unmethylated sequences are also present in AUMA fingerprints.

Application of QUMA and AUMA to a series of colorectal carcinomas and their paired normal mucosa has offered global estimates of unmethylation of Alu elements in normal and cancer cells and has revealed a large collection of unique sequences that undergo highly recurrent hypomethylation and hypermethylation in colorectal tumors.

MATERIALS AND METHODS

Tissues and cell lines

Fifty colorectal carcinomas and their paired non-adjacent areas of normal colonic mucosa were included in this analysis. Samples were collected simultaneously as fresh specimens and snap-frozen within 2 h of removal and then stored at -80ºC. All samples were obtained from the Ciutat Sanitària i Universitària de Bellvitge (Barcelona, Spain). The study protocol was approved by the Ethics Committee. Human colon cancer cell lines (HT29, SW480, HCT116, LoVo, DLD-1, CaCo-2 and LS174T) were obtained from the American Type Culture Collection (ATCC; Manassas, VA). KM12C and KM12SM cells were generously provided by A. Fabra. DNA from

tumor-normal pairs was obtained by conventional organic extraction and ethanol precipitation. DNA purity and quality was checked in a 0.8% agarose gel electrophoresis. RNA from cell lines was obtained by phenol-chloroform extraction and ethanol precipitation, following standard procedures.

Bioinformatic analysis

The distribution of SmaI sites, putative amplification hits, PCR homologies, CpG islands and repetitive elements was assessed using the human genome assembly 36.1 from NCBI. Data were obtained from the Repbase (http://www.girinst.org/repbase/index.html) and the Genome Browser Databases (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/). Only assembled chromosome fragments were considered. A Perl routine was used to score all positions containing the target sequences in all chromosomes (available from the authors upon request). Data were analyzed using Excel spreadsheets.

To calculate the proportion of unmethylated Alu elements at the genomic level, the number of AUMA hits identified in bioinformatic analysis was corrected according to the distribution of experimentally generated AUMA products performing Monte Carlo simulations. One thousand Monte Carlo simulations were performed using an Excel Add-in (available at www.wabash.edu/econometrics). In Monte Carlo simulations it was assumed that 80-100% of SmaI sites at CpG islands are unmethylated and that 50-100% of SmaI sites in other genomic regions different from Alu's and CpG islands are unmethylated.

Quantification of UnMethylated Alu (QUMA)

One microgram of DNA was digested with 20 U of the methylation sensitive restriction endonuclease SmaI (Roche Diagnostics GmbH, Mannheim, Germany) for 16h at 30ºC, leaving cleaved fragments with blunt ends (CCC/GGG). Adaptors were prepared incubating the oligonucleotides Blue (CCGAATTCGCAAAGCTCTGA) and the 5' phosphorilated MCF oligonucleotide (TCAGAGCTTTGCGAAT) at 65ºC for 2 min, and then cooling to room temperature for 30-60 min. One microgram of the digested DNA was ligated to 2nmol of adaptor using T4 DNA ligase (New England Biolabs, Beverly, MA). Subsequent digestion of the ligated products with the methylation insensitive restriction endonuclease XmaI (New England Biolabs) was performed to avoid amplifications from non-digested methylated Alu's. The products were purified using the GFX Kit (Amersham Biosciences, Buckinghamshire, UK) and eluted in 250 µl of sterile water.

Quantitative real time PCR was performed using 1ng (the equivalent of 333 genomes) of DNA in a LightCycler 480 real time PCR system with Fast Start Master SYBR Green I kit (Roche). For PCR conditions see Supplemental data. The downstream BAu-TT primer (constituted by the 3' end of Blue primer, and the GGGTT sequence including the GGG 3' side of the cut SmaI site and the Alu homologous TT dinucleotide, ATTCGCAAAGCTCTGAGGGTT) and the upstream primer was an Alu consensus sequence (CCGTCTCTACTAAAAATACA) (see Supplemental data). Magnitudes were expressed as number of unmethylated Alu's per haploid genome after DNA

input normalization. The number of haploid genomes present in the test tube was determined in the same multiwell plate by quantification of Alu sequences irrespectively of the methylation state. A real time PCR using Alu consensus primers upstream of the CCCGGG site was performed (see Supplemental data) and the number of genomes was calculated against a standard curve constructed with a reference genomic DNA measured by UV spectrophotometry.

To determine the efficiency of the assay and to perform absolute quantification, an external Alu product generated by PCR from a DNA fragment containing an AluSx element was used as standard (Supplemental methods). The number of copies of the external control was spectrophotometrically quantified and dilution curves were generated and treated as samples. Comparison of dilution curves before and after sample processing indicated that the mean recovery was 73%. DNA samples overdigested with the the methylation insensitive XmaI endonuclease were spiked with different amounts of the external standard and processed. The sensitivity of the QUMA detection was 100 unmethylated Alu's per haploid genome (Supplemental figure 1) using 1 ng of genomic DNA per PCR. A linear response was observed between 1000 and 100000 unmethylated Alu's per haploid genome (Supplemental Data).

Amplification of UnMethylated Alu (AUMA)

DNA digestion with SmaI enzyme and ligation to the linker was performed as described above for QUMA, except for the XmaI digestion which was skipped. The product was purified using the GFX Kit (Amersham Biosciences) and eluted in 250 µl of sterile water. Six different chimeric primers constituted by the 3' end of the Blue primer sequence (ATTCGCAAAGCTCTGA), the cut SmaI site (GGG) and two, four or seven additional nucleotides homologous to the Alu consensus sequence were used to enrich for Alu sequences (see supplemental Methods). Three primers were designed to amplify 'upstream' of the SmaI site (towards Alu promoter): BAu-TT, BAu-TTCA, BAu-TTCAAGC. Three other primers were designed to amplify 'downstream' (towards Alu poly-A): BAd-AG, BAd-AGGC, BAd-AGGCGGA. Letters after the dash correspond to the 3' sequence of the primer (see supplemental data). Data reported here were obtained by using the BAu-TT primer.

In each PCR reaction only one primer was used at a time. Products were resolved on denaturing sequencing gels. Although bands can be visualized by silver staining of the gels, radioactive AUMA's were performed for normal-tumor comparisons. A more detailed description of the PCR and the visualization of the bands is given as Supplemental data.

Only sharp bands that were reproducible and clearly distinguishable from the background were tagged and included in the analysis. Faint bands with inconsistent display due to small variations in gel electrophoresis resolution were not considered. Band reproducibility was assessed with the analysis of PCR duplicates of three independent sample digests from two different samples and PCR replicates from the same digest from four paired tumor-normal samples. AUMA fingerprints were visually checked for methylation differences between bands in the tumor with regard to its paired normal mucosa. Under these premises a given band was scored according to three possible behaviors: hypomethylation (increased intensity in the

tumor), hypermethylation (decreased intensity in the tumor) and no change (no substantial difference in intensity between normal and tumor samples) (Figure 2). Only those bands showing clear changes in their intensities in the fingerprint were considered to represent methylation changes. This is consistent with previous studies done using a related technique (42,43).

Competitive hybridization of AUMA products to metaphase chromosomes and BAC arrays

The origin and chromosomal distribution of sequences generated by AUMA was performed using procedures analogous to CGH. Briefly, an AUMA product obtained from a normal tissue DNA was purified using Jet quick PCR product purification kit (Genomed, Löhne, Germany) and labeled with SpectrumRed dUTP (Vysis, Downers Grove, IL) using a Nick Translation kit (Vysis). Similarly, genomic DNA of the same normal sample was labeled with SpectrumGreen dUTP (Vysis) and both probes were cohybridized to metaphase chromosomes. Procedures and image analysis were performed as described (44).

Differential normal-tumor representation of AUMA at the genomic scale was performed by competitive hybridization of AUMA products to BAC arrays. AUMA products from two normal-tumor pairs were purified using Jet quick PCR product purification kit (Genomed, Löhne, Germany) and 1 µg was labeled with dCTP-Cy3 or dCTP-Cy5 (Amersham Biosciences, UK) by use of the Bioprime DNA Labeling System (Invitrogen, Carlsbad, CA). Probes were hybridized to SpectralChip 2600 BAC arrays (Spectral Genomics, Houston, TX) following the manufacturer's instructions. Arrays were scanned with a ScanArray 4000 (GSI Lumonics, Watertown, MA) and processed with GenePix software (Axon Instruments, Union City, CA). The resulting data were processed to filter out low-quality spots based on spot area and similarity of readings between the two replicates of each BAC. Data manipulation was performed using Excel spreadsheets. Because AUMA products are not evenly distributed along chromosomes, only BACs with intensities above the 10% of maximum intensity in at least one of the two channels were considered for ratio calculations. The pattern of chromosomal alterations in these two tumors was determined by conventional CGH as described (44).

Isolation and cloning of AUMA tagged bands

DNA excised from gels was directly amplified with the same primer used in AUMA (BAu-TT) (Supplemental figure 2). The amplified product was cloned into plasmid vectors using the pGEM-T easy vector System I cloning kit (Promega, Madison, WI). Automated sequencing of multiple colonies was performed using the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA) to ascertain the unique identity of the isolated band. Sequence homologies were searched for using the Blat engine (http://genome.ucsc.edu/). Selected clones corresponding to AUMA isolated bands were radioactively labeled and used as a probe to confirm the identity of the excised band by hybridization to AUMA fingerprints as previously described (45).

Bisulfite genomic sequencing

Differential methylation observed in some AUMA tagged bands was confirmed by direct sequencing of bisulfite treated normal and tumor DNA as previously described (46). Prior to sequencing, DNA was amplified using a nested or semi-nested PCR approach, as appropriate. Three independent PCRs were done and products were pooled to ensure a representative sequencing. The sequence of PCR primers is described in Supplemental Data.

Histone modification analysis by Chromatin ImmunoPrecipitation (ChIP)

Briefly, $6x10^6$ cells were washed twice with PBS and cross-linked on the culture plate for 15 min at room temperature in the presence of 0,5% formaldehyde. Cross linking reaction was stopped by adding 0.125M glycine. All subsequent steps were carried out at 4°C. All buffers were pre-chilled and contained protease inhibitors (Complete Mini, Roche). Cells were washed twice with PBS and then scraped. Collected pellets were dissolved in 1ml lysis buffer (1% SDS, 5mM EDTA, 50mM Tris pH8) and were sonicated in a cold ethanol bath for 10 cycles at 100% amplitude using a UP50H sonicator (Hielscher, Teltow, Germany). Chromatin fragmentation was visualized in 1% agarose gel. Obtained fragments were in the 200 to 500 bp range. Soluble chromatin was obtained by centrifuging the sonicated samples at 14.000g for 10 min at 4°C. The soluble fraction was diluted 1/10 in dilution buffer (1% Triton X-100, 2mM EDTA, 20mM Tris pH8, 150mM NaCl) then aliquoted and stored at -80°C until use.

Immunoprecipitation was carried out at 4°C by adding 5 to 10 µg of the desired antibody to 1ml of chromatin. Chromatin-antibody complexes were immunoprecipitated with specific antibodies using a protein A/G 50% slurry (Upstate, Millipore, Billerica, MA) and subsequently washed and eluted according to the manufacturer's instructions. Antibodies against acetylated H3 K9/K14 (Upstate), dimethylated H3 K79 and trimethylated H3 K9 (Abcam, Cambridge, UK) were used. Enrichment for a given chromatin modification was quantified as a fold enrichment over the input using quantitative real time PCR (Roche). For every PCR, a standard curve was obtained to assess amplification efficiency. All quantifications were performed in duplicate.

RESULTS

Genomic estimation of the targets and evaluation of the adequacy of the approach by computational analysis.

The availability of the human genome map has allowed us to make a detailed estimation of the frequency and distribution of the sites targeted by our approaches on the genomic scale. A Perl routine was used to score all positions containing the target sequences in all chromosomes and was also applied to perform a virtual AUMA (see Material and Methods). Some of the most important data derived from the bioinformatic analysis are shown in table 1 and figure 3.

Because of the C to T mutational bias at CpG sites (47), any amplification method relying on the consensus sequence (see supplemental data) will only cover a fraction of all the Alu's. Therefore it is important to estimate the degree of representativity of the methods used here if genome-wide estimations are to be made. Alu repeats constitute 7.4% of the human genome but accumulate 40.7% of all SmaI sites (Table 1). Nearly 200000 Alu's (18% of all Alu's) contain a SmaI site and 155000 retain the AACCCGGG consensus sequence (Table 1 and Figure 3) and are therefore potential targets of QUMA and AUMA. While 38.0% of the youngest AluY elements contain this sequence, the proportions drop to 14.8% and 0.4% in AluS and AluJ families, respectively (Table 1 and Figure 3). These frequencies are consistent with a higher C to T transition trend at CpG sites in older Alu's (47).

The representativity of AUMA was analyzed by a virtual bioinformatic assay of the human genome sequence. A total of 168309 AACCCGGG (or CCCGGGTT) hits were identified throughout the genome, with 92.9% of all hits within Alu elements (Table 1 and Figure 3). This implies that 14.2% of all Alu elements contained the AACCCGGG sequence. Another 1.0% of the hits were in CpG islands and 6.8% in the rest of the genome (including unique sequences and other repeats) (Table 1 and Figure 3). As expected, Alu elements containing the target sequence mostly belonged to the AluS (2/3) and AluY families (1/3), with a minimal representation of the older family AluJ (less than 1%). Virtual AUMA determined the presence of 5498 putative products of less than 1Kb (the sequence AACCCGGG in the up strand and the sequence CCCGGGTT in the down strand at a distance of less than 1 kb). Although actual AUMA PCR products may reach 2 kb length (see below), we used the 1 kb limit to compare with AUMA isolated bands, which were shorter than 1 kb. Most virtual AUMA products contained an Alu element at at least one of the ends (93%).

Quantification of Unmethylated Alu in normal and tumor tissues.

The QUMA approach was applied to quantify unmethylated Alu's in a series of 18 colorectal carcinomas and their paired normal colonic mucosa. An external DNA fragment containing an AluSx element was used as a standard (see Material and Methods and Supplemental Data) in order to make an absolute quantification of the number of unmethylated Alu's. Replicates and dilution curves of the samples and standard were performed to assess reproducibility, sensitivity and accuracy (Supplemental Figure 1). Results were normalized by assessment of the number of haploid genomes per test tube (See Material and Methods). The average number of unmethylated Alu's per haploid genome was 25486±10157 in normal mucosa, and 41995±17187 in tumor samples ($p=0.004$, paired t-test) (Figure 4). At bioinformatic level we identified a total of 168309 Alu elements containing the AACCCGGG sequence (potential targets of QUMA) (Table 1). Therefore we estimate that 15.1% of Alu repeats with a AACCCGGG site are unmethylated in the average normal colonic mucosa cell, while this figure is 24.9% in the cancer cell. Considering that the human genome contains approximately 1.1 million Alu elements, these estimates indicate that unmethylated Alu's constitute the 2.3% and 3.8% of all Alu's in the normal and tumor tissues respectively.

Set-up and optimization of AUMA fingerprinting

Because QUMA products are fully contained within the Alu sequence, it is not possible to identify and position in the genome the unmethylated Alu elements. To achieve this it is necessary to amplify the targeted unmethylated Alu element together with an adjacent unique sequence. This was attained through the use of the second method, the Amplification of UnMethylated Alus (AUMA). The AUMA also targets the unmethylated AACCCGGG sequence, as in QUMA, but in this case a single primer is used in the PCR (BAu-TT) (see Figure 1 and Supplemental Data). Moreover, the product is resolved on a high resolution gel electrophoresis resulting in a band-rich fingerprint. AUMA bands correspond to sequences flanked by two unmethylated target sequences in opposite strands and sufficiently close to allow PCR amplification (Figure 1 and Supplemental Data). Since 92% of the AACCCGGG occurences in the human genome are in Alu's (Table 1), the approach is largely biased towards the amplification of unmethylated Alu's. The presence of non Alu sequences at one of the ends or between two repetitive elements allows the positioning within the genome map of all products.

AUMA products generated using the BAu-TT primer produced highly reproducible fingerprints consisting of bands ranging from ~100 to ~2000 bp when resolved in high resolution sequencing gels (Figure 2 and Supplemental Figure 2). Some well identifiable bands (up to 5 per experiment) showed random display in both intra-assay and inter-assay replicates (Supplemental figure 2) and were not considered for analysis. A subset of 110 bands with consistent display among all the experiments were tagged and selected for comparative analysis between samples (see Materials and Methods).

It should be noted that different fingerprints containing alternative representations may be obtained by AUMA just by using primers that either amplify from the SmaI site towards the Alu promoter (upstream Alu amplification) or towards the Alu poly-A tail (downstream Alu amplification). Also the stringency of the Alu selection may be increased by using longer primers containing additional nucleotides corresponding to the Alu consensus sequence (See Material and Methods). An illustrative example of AUMA fingerprints generated with different Alu-upstream and Alu-downstream primers is shown in supplemental figure 2. All the data reported in this paper regarding AUMA were obtained using the BAu-TT primer.

Chromosomal origin of AUMA products

Competitive hybridization between AUMA products and genomic DNA on metaphase chromosomes yielded a characteristic hybridization pattern demonstrating the unequal distribution of AUMA products along the human genome (Figure 5A). Competitive hybridization of AUMA products to BAC arrays showed profiles consistent with those obtained on metaphase chromosomes (Figure 5B). The highest AUMA signal was detected in whole chromosomes 16, 17 and 19 in contrast with chromosomes 2, 13, 18 and X which were mainly labelled by genomic DNA. Other chromosomes showed a discrete pattern of AUMA product hybridization, in which telomeric bands in chromosomes 1, 4, 5, 9, 12 and X, and interstitial bands in chromosomes 1, 3, 7, 11 and 12 are the most prominent examples.

Identification of AUMA amplified DNA products

To determine the identity of bands displayed by AUMA, 38 tagged bands were isolated and cloned. Multiple clones from each band were sequenced, resulting in a total of 49 different sequences due to the coincidence of more than one sequence in some bands. Characterized bands included bands displaying no changes in the normal-tumor comparisons and bands recurrently altered in the tumor. Table 2 summarizes the main features of a subset of the bands showing recurrent alterations. A list of all the sequences isolated from AUMA fingerprints is provided as Supplementary Table 1. All sequenced bands contained a region of non-repetitive sequence and matched with the BLAST reference sequence, allowing the assignment of a unique chromosomal localization. The BLAST reference sequence corresponding to the 49 sequences isolated from the AUMA fingerprint presented the target sequence CCCGGGTT including the SmaI at both ends. Southern blot analysis of selected cloned sequences showing coincidental size was performed to confirm its correspondence with the band displayed in AUMA fingerprints (Supplemental figure 4).

To obtain a more representative collection of AUMA bands, 200 clones obtained from normal tissue AUMA products were sequenced. The analysis revealed 88 additional sequences. This resulted in a total of 137 different loci represented in AUMA (Supplemental table 1). Most sequences obtained by random cloning were also flanked by two AACCCGGG sequences in opposite DNA strands. Nevertheless, in 27 sequences the AACCCGGG site was only present at one of the ends, with the other end showing high homology with the primer although it was not a perfect match. The presence of these sequences suggests that, in some instances, a single cut in the sequence may be enough to produce an amplifiable fragment. This is not considered an artifact since these bands still represent an unmethylated AACCCGGG site.

Genome-wide estimations of unmethylation in Alu's and distribution by subfamily

Of the 137 identified loci represented in AUMA, 114 were isolated from normal tissue DNA and 23 from tumor DNA. Half of the sequences contained an Alu sequence at one of the ends and two were flanked by two inverted Alu's. AUMA sequences isolated from tumor tissue and not present in normal tissue (this corresponds to a tumor specific hypomethylation) showed a higher proportion of Alu elements (16 out of 23, 70%), and included one band flanked by two inverted Alu's. Globally, 78 unmethylated Alu elements were identified and positioned in the human genome map.

To study the genomic distribution of unmethylated sequences in normal colon mucosa we only considered the 114 bands obtained from normal tissue. This resulted in a total of 201 unmethylated hits characterized throughout the genome. The nature of the sequences represented in actual AUMA showed striking differences with the distribution expected from the virtual AUMA analysis. The methylation status of the sequence is likely to be the main (if not the only) source of these differences because the virtual AUMA did not consider this state. Therefore we can use these differences to estimate the degree of unmethylation of the Alu repeats. Only 29.4% of the AUMA ends consisted of Alu's, as compared with the expected

92.9% resulting from the bioinformatic analysis. The highest downrepresentation corresponded to the youngest AluY family, which was present in 6.0% of the AUMA ends, while it was expected to add up to 33.1% in virtual AUMA. AluS representation in actual and virtual AUMA was 22.3% and 60%, respectively. Interestingly, AluJ representations, in both actual and virtual AUMA, were closer (1.0% and 0.7%, respectively) (figure 3 and table 1). These results suggest that there is a stronger pressure to methylate younger Alus. Alternatively, the hits corresponding to CpG islands were overrepresented in actual versus virtual AUMA by a factor of nearly 25 fold (27.4 versus 1.1%), consistently with the unmethylated status of most CpG islands (figure 3 and table 1). The rest of the hits were located in different types of repetitive elements (MIR, MER, LTR, LINE, etc.) and unique sequences (Supplemental table 1). The miscellaneous collection of sequences ("Rest") was over-represented by about 7 fold (observed hits: 43.3%; expected hits: 5.9%, table 1). The 46 AUMA hits represented by the 23 bands specific of tumor tissue showed a higher proportion of Alu's (41%, versus 29% in normal tissue) but similar distribution by Alu type (10 AluS, 5 AluY, 1 AluJ, 4 in CpG islands and 26 in other sequences).

To calculate the proportion and distribution of unmethylated Alu elements on a genomic scale, we performed Monte Carlo simulations taking into account the observed and expected distribution of hits in each Alu family and CpG islands and the rest of sequences (See Material and Methods). We estimate that at least 4104 Alu elements are unmethylated or partially unmethylated in normal colonic mucosa. This corresponds to 2.64% of all Alu elements containing the target sequence AACCCGGG (Table 1). Although AluS and AluY represent the majority of these sequences it should be noted that the methylation pressure is inverse to the conservation of the SmaI site. That is, the most conserved and younger AluY family shows the lowest relative rate of unmethylation; and the older and more degenerated AluJ family exhibits the highest unmethylation.

Application of AUMA to detect differential DNA methylation in colorectal carcinomas

In order to test the usefulness of the method for the detection of new altered methylation targets we applied AUMA to a series of 50 colorectal carcinomas and their paired normal mucosa. Two cases were excluded from the analysis due to recurring experimental failure of the normal or tumor tissue DNA. For the rest of 48 normal-tumor pairs, consistent and fully readable fingerprints were generated and evaluated for normal-tumor differential representation. A given case presented, on average, 107 ± 2.9 informative bands (range 98-110). The variation was due to polymorphic display or variable resolution power of gel electrophoresis.

In this study, only those bands showing clear intensity differences between normal and tumor tissue fingerprints (Figure 2) have been scored as methylation changes since they are more likely to reflect tumor-wide alterations. Because the fingerprints represent sequences flanked by two unmethylated sites, a decreased intensity in a given band in the tumor in regard to the paired normal tissue is indicative of hypermethylation, while an increased intensity corresponds to hypomethylation (Figure 2). All tumors displayed changes in regard to the paired normal

tissue. The average tumor showed 19 ± 7 (range 6-37) hypomethylations and 22 ± 10 (range 1-39) hypermethylations. It is of note that hypomethylations could either be seen as an increase in the intensity of a pre-existing band in the normal tissue or a *de novo* appearance of an non-existent band in the normal tissue. This contrasts with hypermethylation events, which rarely showed the complete loss of a band in the tumor sample, most likely due to the unavoidable contamination of normal tissue.

Virtually, all tagged bands (109 out of 110) were found to be altered in at least one tumor when compared to its normal paired mucosa. AUMA tagged bands presented a wide distribution in the hypomethylation/hypermethylation rates (proportion of tumors showing differential display compared to the paired normal tissue) (Figure 6). Hypomethylation and hypermethylation showed a strong negative correlation ($r=-0.55$ and $P<0.0001$), indicating that most bands tended to be either hypomethylated or hypermethylated. A large proportion of tagged bands (78 bands) were recurrently altered in over 25% of the cases included in this series.

In order to determine whether normal-tumor differences were limited to isolated independent loci or changes might affect larger chromosomal regions we compared the distribution of AUMA products generated from two paired normal and tumor tissues and hybridized to BAC arrays. Differential hybridization was observed in many BACs, suggesting that relatively large regions encompassing from several hundred Kbs to a few Mbs may undergo concurrent hypomethylation or hypermethylation. Telomeric regions of many chromosomes contained most of the differential display (Figure 5C). The differential methylation profiles were unaffected by chromosomal dosage as demonstrated by its independence of chromosomal losses and gains (as detected by Comparative Genomic Hybridization (CGH) (Supplemental figure 5).

Validation of methylation changes detected by AUMA

To confirm that the changes observed in AUMA fingerprints corresponded to actual changes in the methylation status of the sequence, eight different sequences obtained from AUMA fingerprints were analyzed in normal and tumor tissues by direct sequencing of sodium bisulfite treated DNA's (Table 2). AUMA detected hypomethylations and hypermethylations were confirmed in most cases. Moreover it was demonstrated that methylation changes affected not only the CpG in at least one of the two flanking SmaI sites (whose methylation prevents AUMA representation) but also neighboring CpG's (supplemental figure 6). In two samples, AUMA changes observed in the tumor could not be confirmed by bisulfite sequencing, suggesting that the change might affect only a small fraction of tumor cells and that both methods may exhibit different sensitivities. The presence of minor subpopulations can be detected using more sensitive techniques, i.e. the Methylation Specific PCR or by sequencing of multiples clones.

Functional implications of changes detected by AUMA

Next, we wondered if DNA methylation changes detected by AUMA may have any functional consequences. We chose one of the most recurrent hypomethylated AUMA sequences (Aq3)

and performed an insightful epigenetic characterization of the region in a series of normal-tumor pairs and in colon cancer cell lines.

Aq3 band is recurrently hypomethylated in tumors according to AUMA fingerprints (Figure 7A). It represents a sequence situated in the eighth intron of the MYOM2 gene (Table 2) and does not fall inside or close to any CpG island. The SmaI sites are located in a MLT1A repeat and an AluYd3 element. The methylation status of the two flanking regions of the AUMA band (465 bp and 213 bp long spanning 20 and 11 CpGs respectively) was analyzed by bisulfite direct sequencing (Figure 7B). Confirmation of AUMA data was performed in three normal-tumor pairs that displayed the *de novo* appearance of the Aq3 band in AUMA fingerprints (cases 17, 63 and 74) and two cases lacking this band in both normal and tumor pair (cases 53 and 99) (Figure 7A), as well as five cell lines (HCT116, DLD-1, LoVo, HT29 and CaCo2). All normal tissues as well as tumors 53 and 99 showed heavy methylation of this region (Figure 7C). In contrast to this and in agreement with AUMA results, tumors 17, 63 and 74 exhibited hypomethylation at most CpGs. Cell lines showed variable profiles of DNA methylation, with CaCo2 exhibiting unmethylation of the MLT1A element but heavy methylation of the AluYd3 element, which was also heavily methylated in HCT116 cells but not in the rest of the cell lines tested. MYOM2 expression levels analyzed by real time RT-PCR were not affected by the methylation status of this sequence (data not shown). Further 45 normal-tumor pairs were analyzed for methylation of the AluYd3 element by real-time dissociation analysis (Supplemental figure 7) and it was found hypomethylated in 26 tumors (58%).

Next, we wondered whether the DNA methylation status of the AluYd3 element was associated with alternative chromatin states. We performed Chromatin ImmunoPrecipitation (ChIP) analysis of histone 3 (H3) modifications indicative of active chromatin: acetylation of lysines 9 and 14 (AcH3K9/K14), and dimethylation of lysine 79 (2mH3K79); and silent chromatin: trimethylation of lysine 9 (3mH3K9). These histone marks were compared between cell lines HCT116 and LoVo (with 100% and 30% methylation of the AluY element, respectively). The silencing mark 3mH3K9 was 3.5 fold higher in HCT116 cells compared to the LoVo cell line (Figure 7D). No differences in active marks were observed and these were significantly lower than the silencing mark 3mH3K9. When HCT116 cells were treated with the demethylating agent 5-aza-2'-deoxycytidine (5AzaC) and the histone deacetylase inhibitor trichostatin A (TSA), a moderate decrease in the amount of the 3mH3K9 mark was observed (Figure 7E). As a whole, these data suggest that DNA methylation changes in this AluYd3 element are accompanied by altered chromatin states. The molecular consequences of such epigenetic changes remain to be identified.

DISCUSSION

Epigenetic states of Alu elements

Full genome sequencing has provided precise maps of repetitive elements, and several studies have investigated their distribution and relationship with genome structure (48-51). More

recently, a few studies have explored sequence-dependent associations between repetitive elements and the epigenetic landscape. There is a characteristic distribution of interspersed elements along methylated and unmethylated domains, with most elements in the methylated compartment of the genome (21). Nevertheless, SINEs, which include Alu elements, are the repetitive sequences most commonly found in unmethyated domains (21) and some Alu elements may contain discriminatory motifs associated with methylation-resistant CpG islands (52). Somatic cells show unstable epigenetic profiles in repetitive elements as demonstrated by global measurements of either DNA methylation (18,20,40,41) or histone modifications (53,54). Recent studies have revealed interindividual variability in DNA methylation profiles at specific Alu elements (55), and Fraga and colleagues detected epigenetic changes arising during the lifetime of monozygotic twins in Alu elements and other sequences (56).

Beyond these few studies, the extension and nature of the epigenetic state of interspersed elements is largely unknown. Global estimates of DNA methylation in repetitive elements have been obtained by Southern blot analyses (reviewed in (30)) and, more recently, by using approaches based on bisulfite conversion of the unmethylated cytosine (40,41,57). These studies have confirmed the global hypomethylation of most tumors but they do not provide detailed information on the nature and localization of the unmethylated elements. *In silico* analysis has revealed that a number of Alu elements close to CpG islands retain a high proportion of CpG sites, and this is presumed to be a sign of unmethylation (58), but no experimental proof has been provided. In our point of view, the lack of simple, specific and sensitive methodologies to screen for epigenetic changes in repetitive elements on the genomic scale has precluded a clearer understanding of the nature and implications of these sequences in cell biology.

Properties of QUMA and AUMA

Here we report a systematic screening of unmethylated Alus as a tool to determine the extent of DNA hypomethylation, to identify specifically unmethylated elements and to detect epigenetic alterations in cancer cells. The QUMA is a very simple and specific method and provides accurate relative estimates of the number of unmethylated elements. The QUMA is specially appropriate for comparative studies, but also provides a raw quantitation of the number of unmethylated elements per haploid genome, outlining the extent of hypomethylated Alu's in normal and pathologic cells. QUMA analysis indicates that about 1 out of 6 Alu elements containing the AACCCGGG site are unmethylated, while in tumors, this figure nearly doubles in agreement with previous studies (reviewed in (23)). Although these analyses are likely to generate good estimates at comparative level (between samples), absolute values should be treated with caution because the determination refers to a single CpG site within the Alu element.

To date there is still a lack of proper methodologies allowing genome-wide screenings for recurrent hypomethylated regions that may have some impact on tumor biology. Even though QUMA and other methodologies (41,59) allow quantitation of unmethylated repeats, they do not provide a straight forward approach to identify and map the amplified targets. At this point,

AUMA takes us a step further, allowing the undoubtful identification of hypomethylated sequences, in addition to hypermethylated targets. Although AUMA is specially suited to determine the nature of the unmethylated elements, it also allows the calculation of global unmethylation in Alu elements. Nevertheless, it should be taken into account that this is an indirect measure, because it relies in the extent of methylation in CpG islands and other sequences. Moreover, unmethylation of a second SmaI site near the Alu is also required to generate the AUMA band and hence to be detected.

Due to sequence degeneration, both QUMA and AUMA are more effective in screening for unmethylation in younger elements. This trend is more clearly seen in AUMA, were only 9% of the Alu elements of the old J subfamily containing the SmaI site retain the AA dinucleotide needed for their amplification, while this figure is 91% and 80% in the younger AluY and AluS subfamilies respectively (Table 1), making clear that younger Alu elements tend to retain the SmaI site nearly as much as they retain the AA dinucleotide required for their amplification. This bias is not a handicap, since unmethylated Alu sequences revealed by AUMA are likely to represent the most relevant events of this kind, because spurious unmethylation of old Alu elements retaining a single or a few CpG sites is expected to have less biological significance than unmethylation of younger Alu elements that are usually closer to active chromatin regions (21) and retain more CpG's. The stronger methylation pressure observed in the AluY class is consistent with this postulate.

AUMA was designed to amplify DNA fragments containing the target sequence (AACCCGGG), which is present in Alu and other repetitive elements. Because a single primer was used for PCR amplification, the target sequence must appear in both strands of the DNA at relatively nearby positions. As expected, Alu elements, with more than 1 million copies per human genome (15), were the most frequent repeat in AUMA bands (50% in sequences isolated from non-tumor tissue), but only two sequenced bands contained two inverted Alu repeats (Supplemental Table 1). This observation is in concordance with previous works reporting on the instability of this inverted repeats, which might have caused their exclusion from the human genome (60,61). More restrictive conditions to select for Alu, or any other repeat of interest, may be achieved by extending the 3' end of the primer specific sequence (see supplemental data); however, the number of sequences we obtained was considered appropriate to accomplish the original aim of the study which is to screen for differentially methylated repetitive elements in colorectal cancer.

It is worth noting that AUMA patterns are highly reproducible not only in replicates but also among different samples, which indicates that the unmethylated status of these repeats is tightly controlled, probably by the epigenetic status of nearby regions. This is strengthened by the confirmation that unmethylation extends many CpG sites beyond the SmaI cut site. Moreover, about 50% of the bands tagged in AUMA fingerprints exhibited variable display among normal tissues (data not shown), suggesting the usefulness of this technique to investigate epigenetic polymorphisms.

Alu's and other repetitive elements tend to be highly methylated in most somatic tissues (8,9,40,62). Here we have identified 78 "atypical" Alu elements exhibiting full or partial unmethylation in normal colonic mucosa cells. Different evidences underscore the adequacy of this approach to track changes with possible functional implications: (1) A significant portion of the characterized bands are located inside or nearby CpG islands and genes; (2) AUMA products show a characteristic distribution in R bands, coincidentally with the distribution of Alu sequences (15), indicating a bias toward the gene-richest portion of the genome known as the H3 isochore (63).

DNA methylation along Alu families

Alu families showed striking differences in their methylation level. Most of the Alu elements characterized here are from the younger families AluS and AluY (74% and 22%, respectively). Nevertheless this observation is mainly due to the depletion of CpG sites in older Alu elements. Hence, only 1 out 230 AluJ elements maintain the AUMA target site (AACCCGGG), while younger elements show higher rates of maintenance in accordance with their age (AluS: 1 out of 7; AluY: 2 out of 5) (Table 1). Interestingly, the rate of unmethylation is higher in older elements (AluJ: 12.2%; AluS: 3.1%; AluY: 1.6%) (Table 1). As noted by Rollins et al (21), the boundaries of unmethylated domains tend to be occupied by methylated Alu transposons of the younger AluS and AluY families. Other studies have noted that CpG island-associated Alu's retain a higher proportion of CpG sites, suggesting that these elements are unmethylated in the germ line (58). These unmethylated elements that can be easily revealed by AUMA are likely to play a regulatory role in a significant number of genes (64).

Application of AUMA to detect epigenetic changes in cancer cells

Cancer-related hypomethylation is well documented (reviewed in (23)) and different studies have demonstrated the demethylation of Alu's and other repetitive elements in different types of neoplasias (65). Our data are consistent with previous estimates and go one step forward in the characterization of unmethylated repeats. The AUMA approach was conceived as a straightforward DNA methylation screening strategy targeting specific interspersed repeats and suitable to be applied to large series of samples or experimental conditions as reported here.

The application of AUMA to a series of colorectal carcinomas and their paired matched normal tissue has revealed a high rate of alterations. This indicates the plasticity of epigenetic control of the elements screened by AUMA in colorectal carcinogenesis. Although some bands show bidirectional changes (hypomethylations and hypermethylations), which have been also reported in other sequences (42,66), most bands display an alternative trend either toward hypermethylations or hypomethylations. Some of these changes are highly recurrent (in more than 50% of tumors), suggesting that they may represent relevant alterations related to mechanisms frequently disturbed in colon cancer. Because the default status of repetitive elements is methylation, hypomethylations are readily detected as the emergence of a new band in the tumor AUMA fingerprint. Since all amplified bands include a unique sequence, it has been possible to identify all of the isolated bands and pinpoint them in the genomic map.

As an example, we have investigated the Aq3 sequence, one of the most recurrent hypomethylations in this study. Aq3 band is flanked by two repeats, a LTR and an AluY, which map within an intron between exons 8 and 9 of the MYOM2 gene at 8p23.3. Both elements are heavily methylated in normal tissue and partially to fully unmethylated in tumor tissue. Interestingly, we have found moderately high levels of the heterochromatin associated mark 3mH3K9 in the fully methylated AluYd3 element in the HCT116 cell line, while the levels were significantly lower (3.5 fold) in the partially demethylated LoVo cell line. Furthermore, none of the classical active marks AcH3K9/K14 and 2mH3K79, have been found enriched in the LoVo cell line. These data are in concordance with preliminary data showing that the hypomethylation does not affect the expression of MYOM2 (data not shown) but could rather affect chromatin structure in the region. In agreement with these observations, this genomic region undergoes frequent losses (67-69) and is rearranged in many different types of tumors (70,71), which hints at a role for DNA hypomethylation in genomic instability (24-27,72). The specific functional consequences of this hypomethylation deserve further investigations.

Another application of AUMA is the detection of genomic regions that have been silenced in cancer. Interspersed elements are concentrated in gene-rich regions and due to the intended selection of unmethylated repetitive elements in AUMA, it appears reasonable to postulate that normally unmethylated sequences are likely to pinpoint active genomic regions. In this context, AUMA provides a large collection of genomic regions undergoing hypermethylation, which are readily seen as bands recurrently loss in the fingerprints. DNA methylation associated epigenetic silencing is probably one of the most prevalent mechanisms of tumor suppression inactivation in cancer (33,37,73). Therefore, AUMA can also be used to screen for differential methylation not only in Alu elements but also in unique sequences and repetitive elements other than Alu.

In summary, the QUMA and AUMA methodologies are a simple and novel approach to explore and gain insights into the functional significance of interspersed genomic elements and neighboring sequences. Due to its distinctive features (bias for unmethylated elements in gene-rich regions and detection of both hypomethylation and hypermethylation) we think that these techniques constitute a new and unique tool that should complement global determinations and high resolution genome-wide scanning strategies. Beyond unmethylated repetitive elements, AUMA can be also used to detect recurrent epigenetic changes associated with tumorigenesis including gene epigenetic inactivation.

Acknowledgements

We thank Gemma Aiza for technical support and Jessica Halow for critical review of the manuscript. JR was a fellow of the Generalitat de Catalunya; EV was a fellow of the Fondo de Investigación Sanitaria (FIS).

REFERENCES

1. Mattick, J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays.*, **25**, 930-939.
2. Zuckerkandl, E. (2002) Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica.*, **115**, 105-129.
3. Kreahling, J. and Graveley, B.R. (2004) The origins and implications of Aluternative splicing. *Trends Genet.*, **20**, 1-4.
4. Jasinska, A. and Krzyzosiak, W.J. (2004) Repetitive sequences that shape the human transcriptome. *FEBS Lett.*, **567**, 136-141.
5. Jordan, I.K., Rogozin, I.B., Glazko, G.V. and Koonin, E.V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.*, **19**, 68-72.
6. Hasler, J. and Strub, K. (2006) Alu elements as regulators of gene expression. *Nucleic Acids Res.*, **34**, 5491-5497.
7. Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.*, **8**, 272-285.
8. Schmid, C.W. (1998) Does SINE evolution preclude Alu function? *Nucleic Acids Res.*, **26**, 4541-4550.
9. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335-340.
10. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33 Suppl**, 245-254.
11. Weissmann, F. and Lyko, F. (2003) Cooperative interactions between epigenetic modifications and their function in the regulation of chromosome architecture. *Bioessays.*, **25**, 792-797.
12. Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481-514.
13. Fazzari, M.J. and Greally, J.M. (2004) Epigenomics: beyond CpG islands. *Nat Rev Genet*, **5**, 446-455.
14. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
15. Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat Rev Genet*, **3**, 370-379.
16. Korenberg, J.R. and Rykowski, M.C. (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell*, **53**, 391-400.
17. Chen, C., Gentles, A.J., Jurka, J. and Karlin, S. (2002) Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 2930-2935.
18. Schmid, C.W. (1991) Human Alu subfamilies and their methylation revealed by blot hybridization. *Nucleic Acids Res.*, **19**, 5613-5617.
19. Gama-Sosa, M.A., Wang, R.Y., Kuo, K.C., Gehrke, C.W. and Ehrlich, M. (1983) The 5-methylcytosine content of highly repeated sequences in human DNA. *Nucleic Acids Res.*, **11**, 3087-3095.
20. Kochanek, S., Renz, D. and Doerfler, W. (1993) DNA methylation in the Alu sequences of diploid and haploid primary human cells. *EMBO J.*, **12**, 1141-1151.

21. Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J. and Bestor, T.H. (2006) Large-scale structure of genomic methylation patterns. *Genome Res.*, **16**, 157-163.

22. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6-21.

23. Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little. *Oncogene*, **21**, 5400-5413.

24. Eden, A., Gaudet, F., Waghmare, A. and Jaenisch, R. (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science*, **300**, 455.

25. Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998) DNA hypomethylation leads to elevated mutation rates. *Nature*, **395**, 89-93.

26. Suzuki, K., Suzuki, I., Leodolter, A., Alonso, S., Horiuchi, S., Yamashita, K. and Perucho, M. (2006) Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell.*, **9**, 199-207.

27. Rodriguez, J., Frigola, J., Vendrell, E., Risques, R.A., Fraga, M.F., Morales, C., Moreno, V., Esteller, M., Capella, G., Ribas, M. *et al.* (2006) Chromosomal Instability Correlates with Genome-wide DNA Demethylation in Human Primary Colorectal Cancers. *Cancer Res.*, **66**, 8462-9468.

28. Liu, W.M., Maraia, R.J., Rubin, C.M. and Schmid, C.W. (1994) Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.*, **22**, 1087-1095.

29. Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J.A., Navarro, G., Colomer, D. *et al.* (2005) Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. *Oncogene*, **24**, 7213-7223.

30. Ehrlich, M. (2006) Cancer-linked DNA hypomethylation and its relationship to hypermethylation. *Curr. Top. Microbiol. Immunol.*, **310**, 251-274.

31. Plass, C. (2002) Cancer epigenomics. *Hum. Mol. Genet.*, **11**, 2479-2488.

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

33. Herman, J.G. and Baylin, S.B. (2003) Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.*, **349**, 2042-2054.

34. Lund, A.H. and van Lohuizen, M. (2004) Epigenetics and cancer. *Genes Dev.*, **18**, 2315-2335.

35. Feinberg, A.P. and Tycko, B. (2004) The history of cancer epigenetics. *Nat. Rev. Cancer*, **4**, 143-153.

36. Laird, P.W. (2005) Cancer epigenetics. *Hum. Mol. Genet.*, **14**, R65-76.

37. Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, **6**, 6.

38. Laird, P.W. (2003) The power and the promise of DNA methylation markers. *Nat. Rev. Cancer*, **3**, 253-266.

39. Cottrell, S.E. (2004) Molecular diagnostic applications of DNA methylation technology. *Clin. Biochem.*, **37**, 595-604.

40. Yang, A.S., Estecio, M.R., Doshi, K., Kondo, Y., Tajara, E.H. and Issa, J.P. (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.*, **32**, e38.

41. Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M. and Laird, P.W. (2005) Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.*, **33**, 6823-6836. Print 2005.

42. Frigola, J., Sole, X., Paz, M.F., Moreno, V., Esteller, M., Capella, G. and Peinado, M.A. (2005) Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer. *Hum. Mol. Genet.*, **14**, 319-326.

43. Frigola, J., Ribas, M., Risques, R.A. and Peinado, M.A. (2002) Methylome profiling of cancer cells by amplification of inter- methylated sites (AIMS). *Nucleic Acids Res.*, **30**, e28.

44. Vendrell, E., Ribas, M., Valls, J., Sole, X., Grau, M., Moreno, V., Capella, G. and Peinado, M.A. (2007) Genomic and transcriptomic prognostic factors in R0 Dukes B and C colorectal cancer patients. *Int J Oncol.*, **30**, 1099-1107.

45. Perucho, M., Welsh, J., Peinado, M.A., Ionov, Y. and McClelland, M. (1995) Fingerprinting of DNA and RNA by arbitrarily primed polymerase chain reaction: applications in cancer research. *Methods Enzymol.*, **254**, 275-290.

46. Stirzaker, C., Song, J.Z., Davidson, B. and Clark, S.J. (2004) Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.*, **64**, 3871-3877.

47. Xing, J., Hedges, D.J., Han, K., Wang, H., Cordaux, R. and Batzer, M.A. (2004) Alu element mutation spectra: Molecular clocks and the effect of DNA methylation. *J Molec Biol*, **344**, 675-682.

48. Medstrand, P., van de Lagemaat, L.N. and Mager, D.L. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.*, **12**, 1483-1495.

49. Grover, D., Majumder, P.P., C, B.R., Brahmachari, S.K. and Mukerji, M. (2003) Nonrandom distribution of alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22. *Mol Biol Evol.*, **20**, 1420-1424.

50. Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K. and Brahmachari, S.K. (2004) Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics.*, **20**, 813-817.

51. Price, A.L., Eskin, E. and Pevzner, P.A. (2004) Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res.*, **14**, 2245-2252.

52. Feltus, F.A., Lee, E.K., Costello, J.F., Plass, C. and Vertino, P.M. (2006) DNA motifs associated with aberrant CpG island methylation. *Genomics*, **15**, 15.

53. Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K. *et al.* (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.*, **13**, 13.

54. Kondo, Y. and Issa, J.P. (2003) Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J Biol Chem.*, **278**, 27658-27662.

55. Sandovici, I., Kassovska-Bratinova, S., Loredo-Osti, J.C., Leppert, M., Suarez, A., Stewart, R., Bautista, F.D., Schiraldi, M. and Sapienza, C. (2005) Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum. Mol. Genet.*, **14**, 2135-2143.

56. Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 10604-10609.

57. Xiong, Z. and Laird, P.W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.*, **25**, 2532-2534.

58. Brohede, J. and Rand, K.N. (2006) Evolutionary evidence suggests that CpG island-associated Alus are frequently unmethylated in human germline. *Hum Genet.*, **119**, 457-458. Epub 2006 Mar 2004.

59. Choi, I.S., Estecio, M.R., Nagano, Y., Kim do, H., White, J.A., Yao, J.C., Issa, J.P. and Rashid, A. (2007) Hypomethylation of LINE-1 and Alu in well-differentiated neuroendocrine tumors (pancreatic endocrine tumors and carcinoid tumors). *Mod. Pathol.*, **20**, 802-810.

60. Lobachev, K.S., Stenger, J.E., Kozyreva, O.G., Jurka, J., Gordenin, D.A. and Resnick, M.A. (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.*, **19**, 3822-3830.

61. Stenger, J.E., Lobachev, K.S., Gordenin, D., Darden, T.A., Jurka, J. and Resnick, M.A. (2001) Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Res.*, **11**, 12-27.

62. Hellmann-Blumberg, U., Hintz, M.F., Gatewood, J.M. and Schmid, C.W. (1993) Developmental differences in methylation of human Alu repeats. *Mol Cell Biol.*, **13**, 4523-4530.

63. Saccone, S., Caccio, S., Kusuda, J., Andreozzi, L. and Bernardi, G. (1996) Identification of the gene-richest bands in human chromosomes. *Gene*, **174**, 85-94.

64. Oei, S.L., Babich, V.S., Kazakov, V.I., Usmanova, N.M., Kropotov, A.V. and Tomilin, N.V. (2004) Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics.*, **83**, 873-882.

65. Wilson, A.S., Power, B.E. and Molloy, P.L. (2007) DNA hypomethylation and human diseases. *Biochim Biophys Acta.*, **1775**, 138-162. Epub 2006 Sep 2001.

66. Nishiyama, R., Qi, L., Tsumagari, K., Weissbecker, K., Dubeau, L., Champagne, M., Sikka, S., Nagai, H. and Ehrlich, M. (2005) A DNA Repeat, NBL2, Is Hypermethylated in Some Cancers but Hypomethylated in Others. *Cancer Biol Ther*, **4**, 4.

67. Emi, M., Fujiwara, Y., Nakajima, T., Tsuchiya, E., Tsuda, H., Hirohashi, S., Maeda, Y., Tsuruta, K., Miyaki, M. and Nakamura, Y. (1992) Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. *Cancer Res.*, **52**, 5368-5372.

68. Sunwoo, J.B., Sun, P.C., Gupta, V.K., Schmidt, A.P., El-Mofty, S. and Scholnick, S.B. (1999) Localization of a putative tumor suppressor gene in the sub-telomeric region of chromosome 8p. *Oncogene.*, **18**, 2651-2655.

69. Muscheck, M., Sukosd, F., Pesti, T. and Kovacs, G. (2000) High density deletion mapping of bladder cancer localizes the putative tumor suppressor gene between loci D8S504 and D8S264 at chromosome 8p23.3. *Lab Invest.*, **80**, 1089-1093.

70. Wong, N., Wong, K.F., Chan, J.K. and Johnson, P.J. (2000) Chromosomal translocations are common in natural killer-cell lymphoma/leukemia as shown by spectral karyotyping. *Hum Pathol.*, **31**, 771-774.

71. Jin, Y., Jin, C., Wennerberg, J., Hoglund, M. and Mertens, F. (2001) Cytogenetic and fluorescence in situ hybridization characterization of chromosome 8 rearrangements in head and neck squamous cell carcinomas. *Cancer Genet Cytogenet.*, **130**, 111-117.

72. Karpf, A.R. and Matsui, S. (2005) Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res.*, **65**, 8635-8639.

73. Egger, G., Liang, G., Aparicio, A. and Jones, P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, **429**, 457-463.

Table 1. Content and distribution of QUMA and AUMA hits in the human genome

Sequence	Mb[1]	No of elements[2]	SmaI sites (CCCGGG)	AACCCGGG Hits[3]	Virtual AUMA hits[4]	AUMA hits[5]	Unmethylated Hits[6]	Unmethylated Hits (%)[7]
Total	3080.4	1118195	486835	168309	5498	201	14332 ± 2418	8.52 ± 1.4%
Alu (S+J+Y)	227.3	1091110	198201	155226	5109	59 (29.3%)	4104 ± 688	2.64 ± 0.44%
AluS	141.2	660415	122459	97951	3382	45 (22.4%)	3028 ± 510	3.09 ± 0.51%
AluJ	54.0	283104	14017	1235	38	2 (1.0%)	151 ± 25	12.25 ± 1.97%
AluY	32.1	147591	61725	56040	1689	12 (6.0%)	925 ± 156	1.65 ± 0.27%
CpG islands	16.2	27085	49430	1673	63	55 (27.4%)	1501 ± 97	90.5 ± 5.79%
Rest	2836.9	-	239204	11410	326	87 (43.3%)	8530 ± 1650	75.9 ± 14.63%

[1] Genome Mb represented by each type of element. Total number corresponds to the number of megabases analyzed for the presence of hits. Only assembled chromosome fragments were considered.
[2] Elements considered in the analysis as obtained from the Repbase and the Genome Browser Databases (see Material and Methods)
[3] Number of occurrences of the sequence AACCCGGG (or CCCGGGTT) within each type of element.
[4] Number of AUMA hits present in virtual PCR products of up to 1000bp.
[5] Hits of actual AUMA products. Only bands appearing in normal tissue were considered. Eighty seven bands contributed two hits each (174 hits) and 27 bands contributed only one due to poor sequence or incomplete homology with the NCBI Build 36.1 of the human genome (hg18 assembly, Mar. 2006). Twenty three additional bands were detected mainly in tumor tissue and were not considered to perform calculations.
[6] Estimated number of unmethylated sites using Monte Carlo simulations (See Material and Methods).
[7] Respect the total number of AACCCGGG (or CCCGGGTT) hits.

Table 2. A selection of characterized AUMA bands

Band ID	Size (bp)	% GC	Chromosome map (Location[a])	Gene	CpG island[b]	Repetitive elements in band ends (5'/3')	Methylation status in tumor[c]
Ai1 c3	509	55	17p11.2 (18206453 - 18206961)	SHMT1	Yes	Alu Sx / MIR	Hypermethylated
Aj2 c1	458	49	1q32.2 (206389082 - 206389539)	MGC29875	Yes	Alu Sq / None	Hypomethylated
Ao1 c4	365	50	19q13.32 (53550179 - 53550543)	AK001784	No	Alu Sx / MIRb	Hypomethylated
Ap1 c6	358	56	5q35.2 (175157321 - 175157674)	CPLX2	Yes	None / MIR	Hypermethylated
Aq3 c6	339	58	8p23.3 (2007343 - 2007682)	MYOM2	No	LTR / Alu Y	Hypomethylated
Ar3 c3	329	57	2q14.3 (127875178 - 127875506)	AF370412	Yes	None / MIRb	Hypomethylated
As3 c6	306	63	16p13.3 (3160477 - 3160782)	None	Yes	None / None	Hypermethylated
Au4 c1	268	57	16p13.3 (3162099 - 3162366)	None	No	tRNA / None	Hypermethylated

[a] Nucleotide position within the contig (strand +). NCBI Build 36.1 of the human genome.

[b] The whole sequence or a fragment of the sequence lays not further than 200 bp of a predicted CpG island.

[c] As compared to the paired normal tissue.

FIGURE LEGENDS

Figure 1. Schematic diagram of the QUMA and AUMA methods. DNA is depicted by a solid line, Alu elements are represented by dashed boxes. The QUMA and AUMA recognition sites (AACCCGGG) are represented by dashed/gray boxes. CpGs at SmaI sites are shown as full circles when methylated and as open circles when unmethylated. The methylation-sensitive restriction endonuclease SmaI can only digest unmethylated targets, leaving blunt ends to which adaptors can be ligated. (A) QUMA is performed by real time PCR of an inner Alu fragment using a primer complementary to the Alu consensus sequence upstream of the SmaI site and the primer complementary to the adaptor to which two Alu homologous nucleotides (TT) have been added. (B) In AUMA, sequences flanked by two ligated adaptors are amplified by PCR using a single primer, the same adaptor primer plus the TT nucleotides. When only a few nucleotides are added to the primer, i.e. TT, as illustrated here, other non Alu sequences may be amplified. This allows the amplification of a large number of sequences that typically range from 100 to 2000 bp.

Figure 2. AUMA of normal (N)-tumor (T) pairs of two different patients performed using primer BAu-TT. A highly reproducible band patterning is observed among the four replicates. Representative bands showing gains (hypomethylations) and losses (hypermethylations) are marked with up and down arrowheads respectively.

Figure 3. Relative distribution the Alu elements and sequence targets considered in bioinformatic and experimental QUMA and AUMA. Mb: Number of megabases occupied by each type of element; Elements: Number of elements considered ("Rest" has been set arbitrarily to 50%); SmaI site: CCCGGG sequence; vQUMA hits: AACCCGGG (or GGGCCCTT) sites in Alu elements; vAUMA hits: AACCCGGG (or GGGCCCTT) sites; vAUMA ends: vAUMA hits considering only putative AUMA products of less than 1 Kb (see Material and Methods); AUMA: elements at each one of the two ends of actual AUMA products.

Figure 4. Quantitation of unmethylated Alu's in 17 paired normal mucosa and colorectal carcinoma by QUMA. The values represent the estimated number of unmethylated Alu's per haploid genome. Most tumors exhibited a higher level of hypomethylation when compared with the respective normal.

Figure 5. (A) Chromosomal origin of AUMA products. A competitive hybridization of AUMA product obtained from normal tissue DNA (red) and genomic DNA (green) to metaphase chromosomes was performed. AUMA products showed an unequal distribution along chromosomes, displaying highest densities at most telomeric regions and some interstitial bands. Chromosomes 16, 17 and 19 yielded the highest AUMA density. (B) Intensity distribution of AUMA products hybridized to BAC arrays in selected chromosomes. The average intensity (X axis) of the two normal (blue) and tumor samples analyzed (red) for each BAC is shown. BACs are arranged along the Y axis according to its position in the chromosome. (C) Differential methylation profiles determined by competitive hybridization of AUMA products from normal and tumor tissue to BAC arrays. Illustrative examples are shown for chromosomes 7 and 8 from

the two cases analyzed (81 and 151). X axis indicates log2 ratio of Tumor/Normal intensities. Positive values (to the right) indicate hypomethylations, negative values (to the left) indicate hypermethylations. Additional examples are shown in Supplemental figure 5.

Figure 6. Distribution of hypermethylation and hypomethylation rates in the 110 AUMA tagged bands. Rates were obtained by comparison of the AUMA fingerprints obtained in 50 colorectal tumors as compared to their respective matched normal tissue.

Figure 7. (A) Detail of the AUMA fingerprints generated from 5 normal-tumor sample pairs. The presence of the Aq3 band is indicated by an asterisk under the three Aq3 positive cases. (B) The relative position of the AUMA Aq3 band, MLT1A and Alu Y repetitive elements, as well as MYOM2 ninth exon are shown. Each vertical line in the CpG distribution represents a CpG dinucleotide along the DNA sequence. Two different fragments were amplified for the bisulfite sequencing analysis (gray boxes). Sequence is oriented 5' to 3' in regard to MYOM2 3' end. (C) Methylation status of the CpG nucleotides in the two fragments amplified were ascertained by direct sequencing of bisulfite treated DNAs of 5 normal-tumor pairs and 5 colon cancer cell lines. (D) ChIP analysis of the AluY element frequently hypomethylated in cancer revealed loss of trimethylation in histone 3 lysine 9 residue (3mH3K9) in LoVo cells (unmethylated at DNA level) as compared to HCT116 (methylated at DNA level). Treatment of HCT116 cells with 5AzaC and TSA produced a moderate decrease in the levels of trimethylation in H3K9.

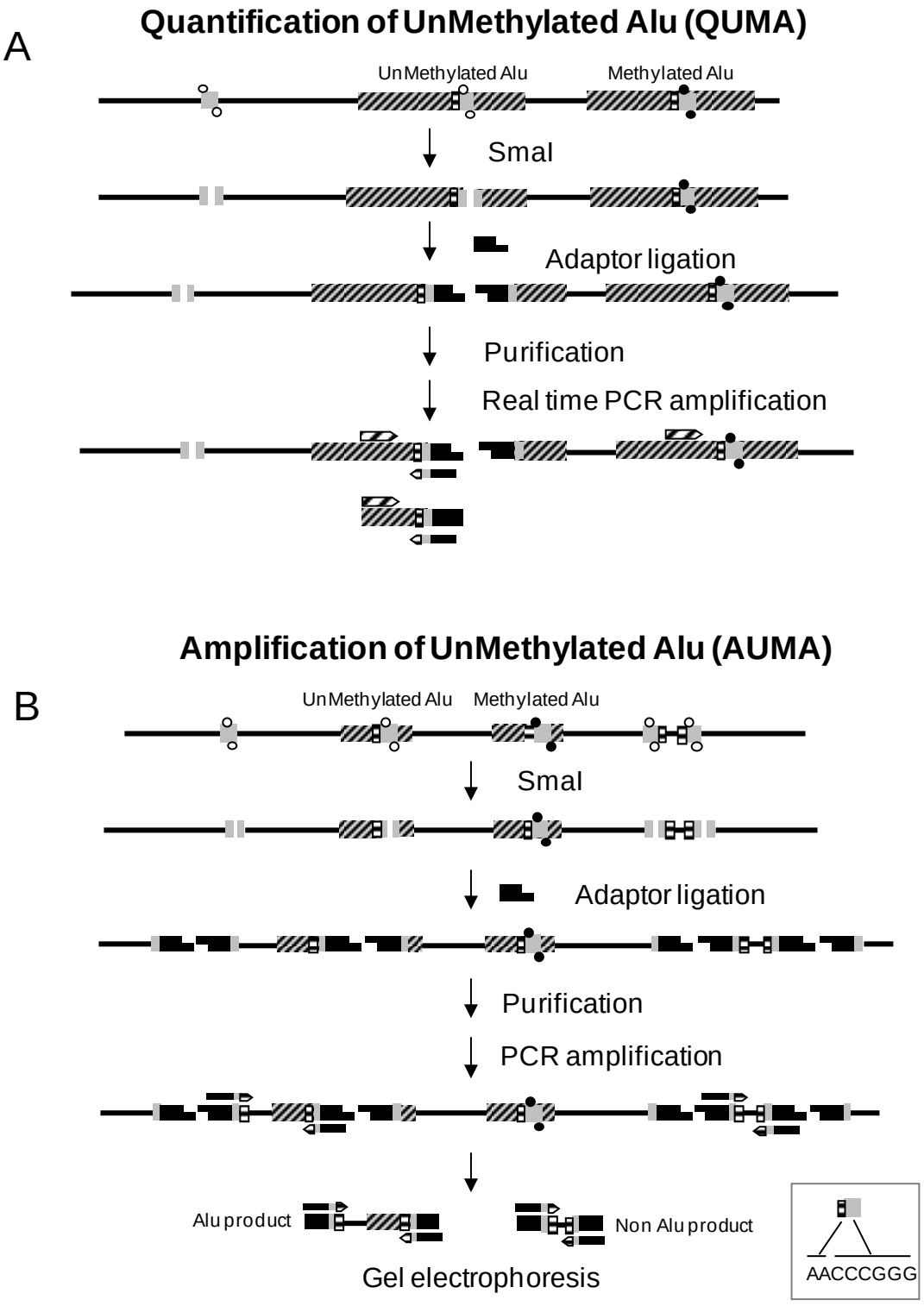

A
Quantification of UnMethylated Alu (QUMA)
UnMethylated Alu
Methylated Alu
SmaI
Adaptor ligation
Purification
Real time PCR amplification
B
Amplification of UnMethylated Alu (AUMA)
UnMethylated Alu
Methylated Alu
SmaI
Adaptor ligation
Purification
PCR amplification
Alu product
Non Alu product
Gel electrophoresis
AACCCGGG

FIGURE 1

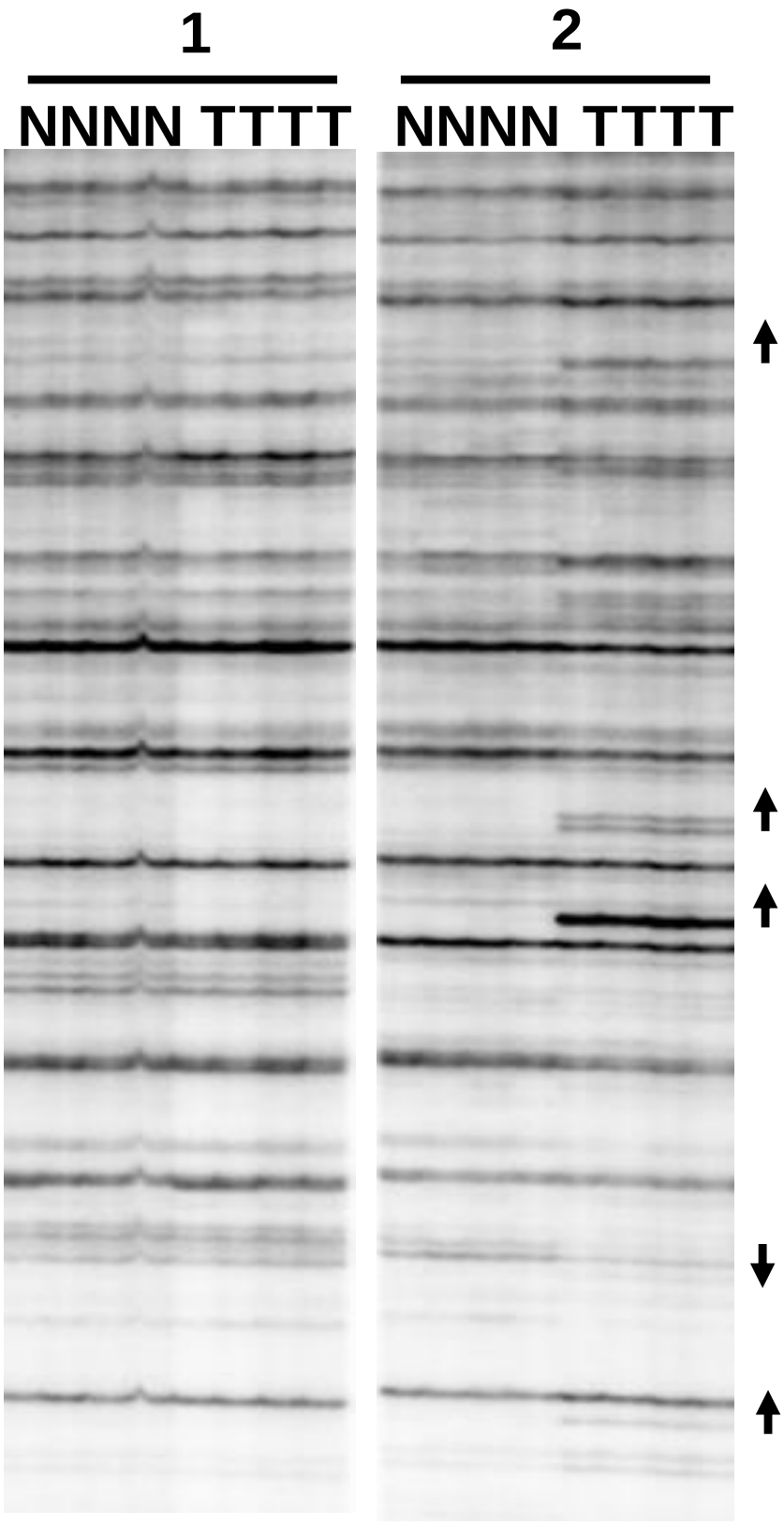


FIGURE 2

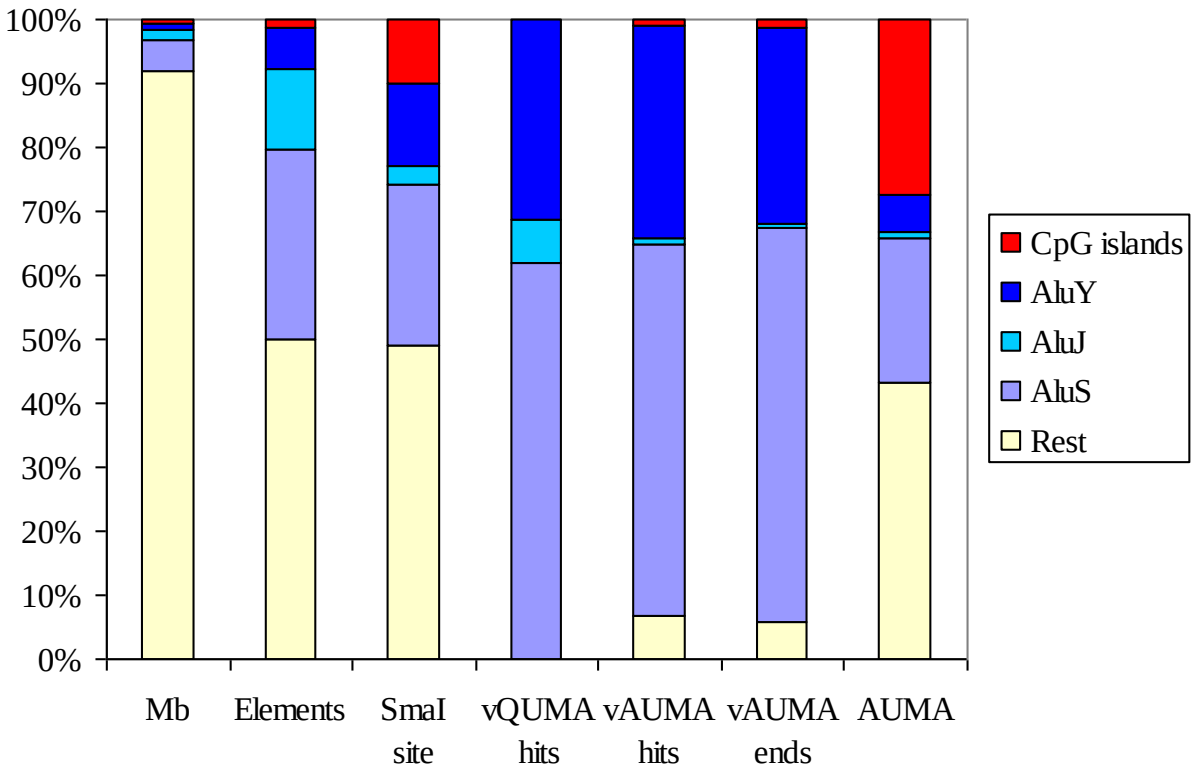

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Mb
Elements
SmaI site
vQUMA hits
vAUMA hits
vAUMA ends
AUMA
CpG islands
AluY
AluJ
AluS
Rest

FIGURE 3

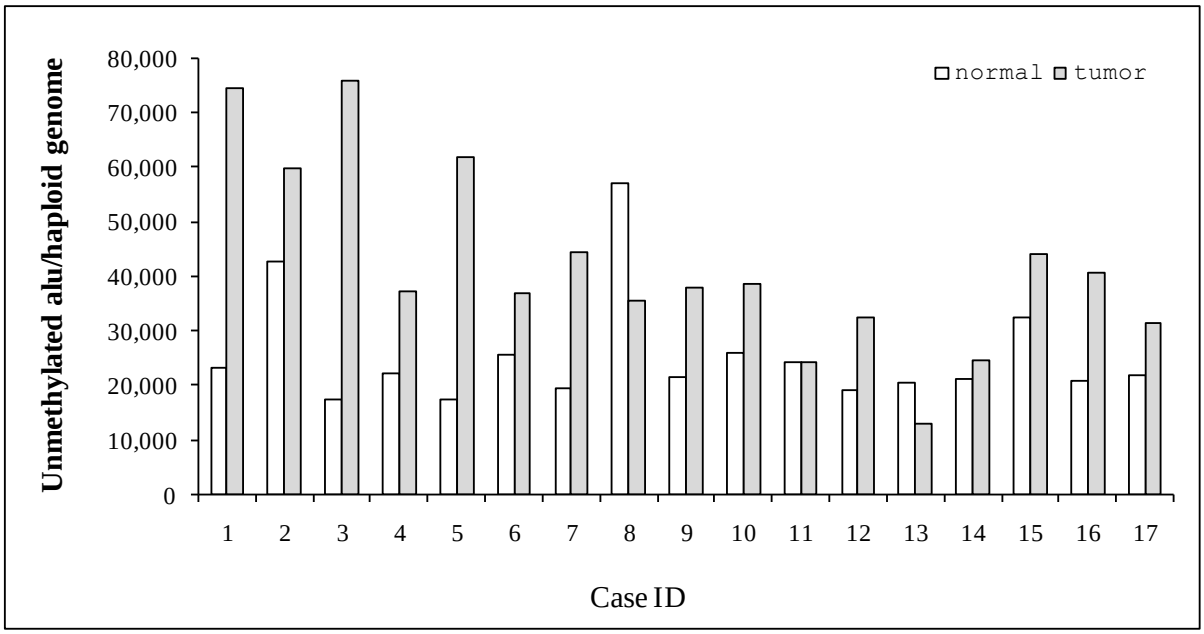

Unmethylated alu/haploid genome
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
normal
tumor
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Case ID

FIGURE 4

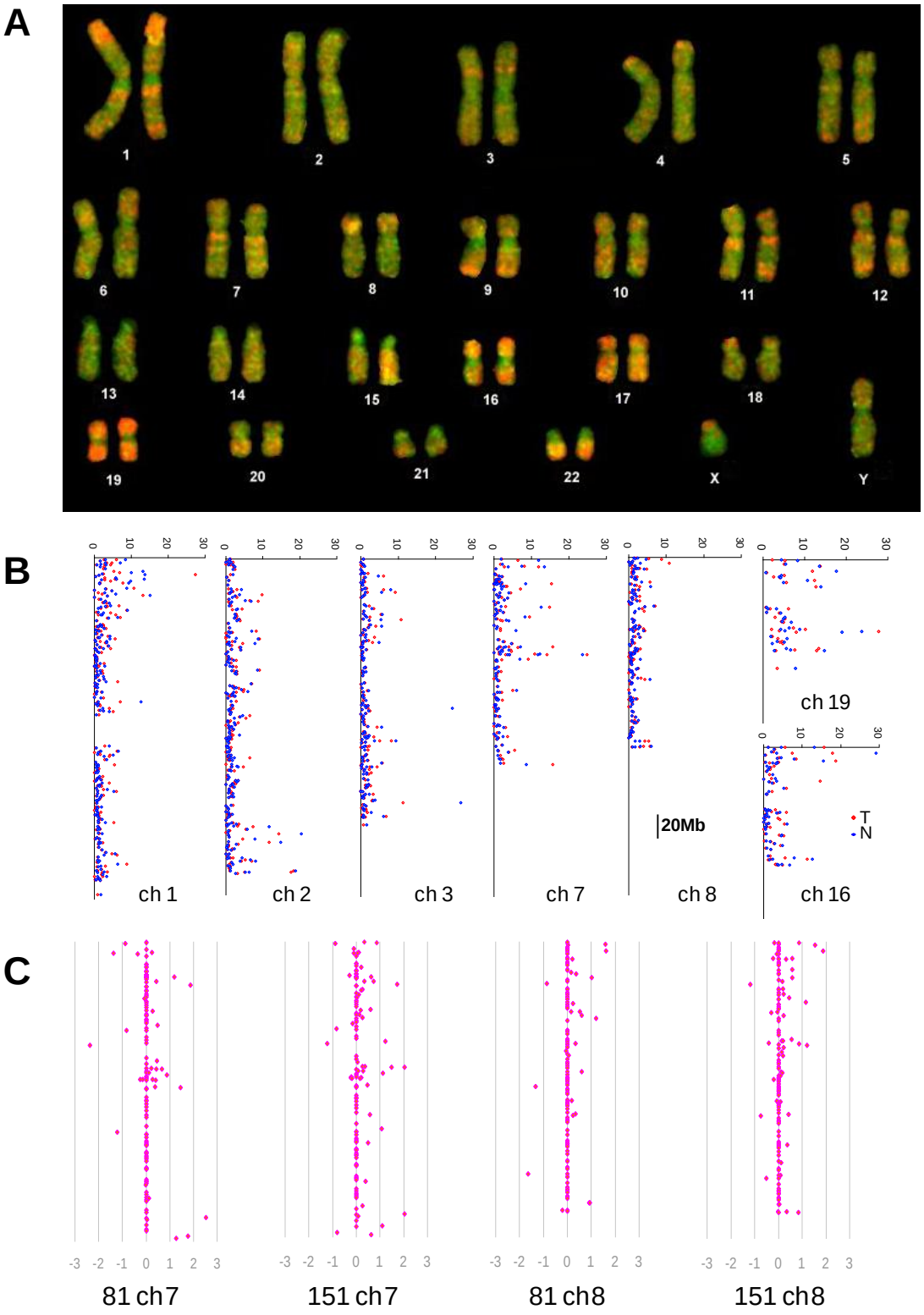

A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
B
ch 1
ch 2
ch 3
ch 7
ch 8
ch 16
ch 19
20Mb
T
N
C
81 ch7
151 ch7
81 ch8
151 ch8

FIGURE 5

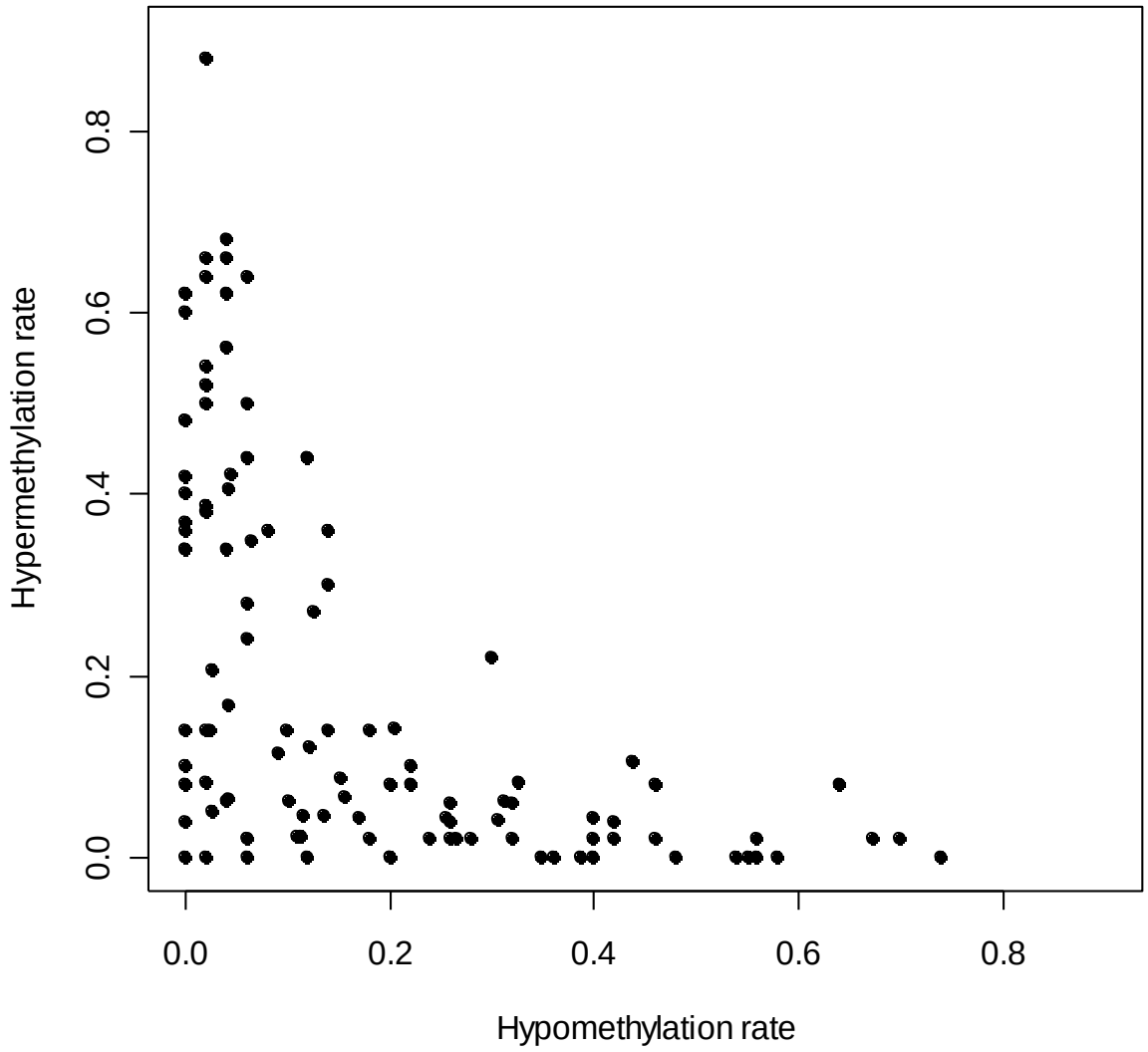

Hypermethylation rate
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
Hypomethylation rate

FIGURE 6

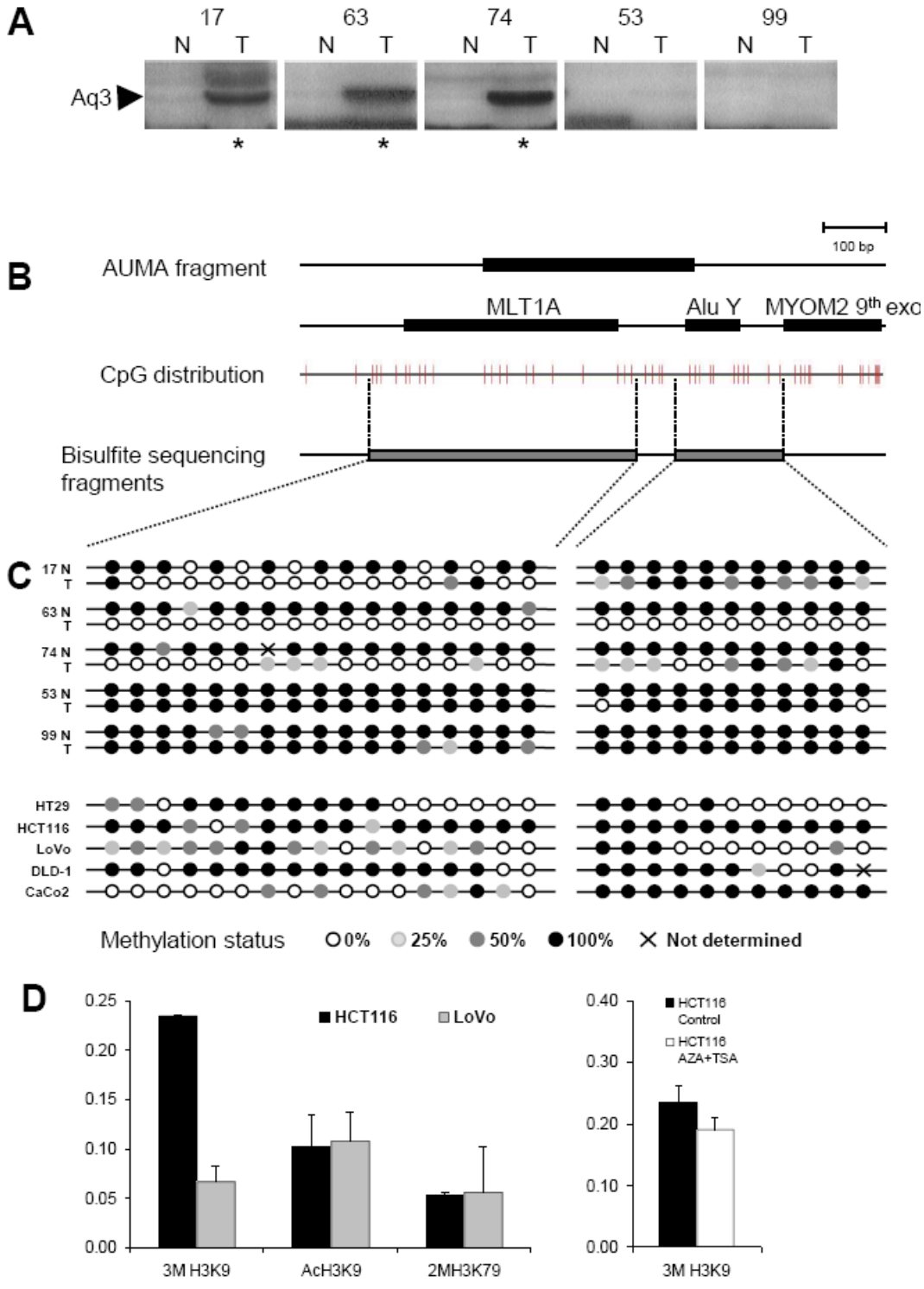

A
17
63
74
53
99
N T N T N T N T N T
Aq3
*
*
*
B
100 bp
AUMA fragment
MLT1A
Alu Y
MYOM2 9th exo
CpG distribution
Bisulfite sequencing fragments
C
17 N T
63 N T
74 N T
53 N T
99 N T
HT29
HCT116
LoVo
DLD-1
CaCo2
Methylation status
0%
25%
50%
100%
Not determined
D
HCT116
LoVo
0.25
0.20
0.15
0.10
0.05
0.00
3M H3K9
AcH3K9
2MH3K79
HCT116 Control
HCT116 AZA+TSA
0.40
0.30
0.20
0.10
0.00
3M H3K9

FIGURE 7

Supplemental Data

Supplemental METHODS

QUANTIFICATION OF UNMETHYLATED ALU (QUMA)

ALU consensus sequence around the SmaI site (up and down DNA strands)

unmethylated SmaI site
5′-...GGAGAATCGCTTGAA**CCC** ↓ **GGG**AGGCGGAGGTTGCAG...-3′
3′-...CCTCTTAGCGAACTT**GGG** ↑ **CCC**TCCGCCTCCAACGTC...-5′

Adaptor	Primer name
P-5′-TCAGAGCTTTGCGAAT-3′	5PMCF
3′-AGTCTCGAAACGCTTAAGCC-5′	Blue

Ligated fragment and upstream and downstream QUMA primers (green captions)

5′-CCGTCTCTACTAAAAATACA-3′ (Alu1)
5′-AACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCC..
3′-TTGGGGCAGAGATGATTTTTATGTTTTTAATCGGCCCGCACCACCGCGCGCGGACATTAGG..

..CAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAA**CCC**TCAGAGCTTTGCGAAT-3′
..GTCGATGAGCCCTCCGACTCCGTCCTCTTAGCGAACTT**GGG**AGTCTCGAAACGCTTAAGCC-5′
3′TTGGGAGTCTCGAAACGCTTA-5′(BauTT)

QUMA Normalization

Primers (5′-3′): CCGTCTCTACTAAAAATACA, ATTCTCCTGCCTCAGCCT

A consensus Alu and the control region amplified to normalize QUMA is shown bold. Primer sites are shown in green. SmaI site is shown in blue.

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACCTGA
GGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACC**CCGTCTCTACTAAAAATACAAAAAT**
TAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAA
TCGCTTGAA**CCCGGG**AGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCG
ACAGAGCGAGACTCCGTCTCAAAAAAAA

QUMA standard

Genomic fragment used to generate the QUMA standard. Fragment was generated by PCR from genomic DNA (locus chr17:18,206,278-18,206,720) using the primers:
GTTACCCAGAGAATAGACTG, AAAACACAATTACCCACTTCC.

Alu Sx sequence is shown in lower case, unique sequence is shown in uppercase:

GTTACCCAGAGAATAGACTGGAAACACCAGTTTAGGTTCTTGCAGAAATAACATTGTAggccgggcgcg
gtggctcatgcctgtaatcccagcattttgggaggccgaggtgggcggatcacttgaggttagtagttggagaccagcctggccaacat
ggtgaaaccccgtctgtactaaaaatacaaaaaaacctggctgcccatggtggcacacacctataatcccagctactcggaaggctga
ggcaggagaatcgcttgaa**cccggg**aggtggaggttgcagtgagcggagatcacgccactgcactccagcctgagtgacagagtg
agactctatctcaaaagaaagaaagagagagaaagaaaagaaAAAAAAGAAAAGAAATAACGTAAATTTATTACTTG
GGGGAACTTGGAAC**GGAAGTGGGTAATTGTGTTTT**

AMPLIFICATION OF UNMETHYLATED ALU (AUMA)

Consensus Alu and AUMA primer sequences

Accession numbers of consensus Alu sequences: Alu J (U14567); Alu Sb (U14568); Alu Sb1 (U14569); Alu Sc (U14571); Alu Sp (U14572); Alu Sq (U14573); Alu Sx (U14574)
Available from GeneBank database at http://ncbi.nih.gov/

Human consensus ALU interspersed repetitive sequence (SmaI site is highlighted) obtained with ClustalW software (European Bioinformatics Institute, EMBL-EBI at http://www.ebi.ac.uk/clustalw/)

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACCTGA
GGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAG
CCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAA**C**
CCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAG
ACTCCGTCTCAAAAAAAA

ALU consensus sequence around the SmaI site (up and down DNA strands)

unmethylated SmaI site
5′-...GGAGAATCGCTTGAA**CCC** ↓ **GGG**AGGCGGAGGTTGCAG...-3′
3′-...CCTCTTAGCGAACTT**GGG** ↑ **CCC**TCCGCCTCCAACGTC...-5′

Adaptor sequence (two views of the same adaptor are shown as ligated to the Alu tail or the head)

		Primer name
Tail-ligation	5′-CCGAATTCGCAAAGCTCTGA-3′	Blue
	3′-TAAGCGTTTCGACACT-5′-P	5PMCF
Head-ligation	P-5′-TCAGAGCTTTGCGAAT-3′	5P-MCF
	3′-AGTCTCGAAACGCTTAAGCC-5′	Blue

Ligated fragment and Alu downstream PCR primers (green captions)

5′-ATTCGCAAAGCTCTGAGGGAGGCGGA-3′ BAd-AGGCGGA
5′-ATTCGCAAAGCTCTGAGGGAGGC-3′ BAd-AGGC
5′-ATTCGCAAAGCTCTGAGGGAG-3′ BAd-AG
5′-CCGAATTCGCAAAGCTCTGA**GGG**AGGCGGAGGTTGCAG...-3′
3′-TAAGCGTTTCGACACT**CCC**TCCGCCTCCAACGTC...-5′

Ligated fragment and Alu upstream PCR primers (green captions)

5′-...GGAGAATCGCTTGAACCCTCAGAGCTTTGCGAAT-3′
3′-...CCTCTTAGCGAACTTGGGAGTCTCGAAACGCTTAAGCC-5′
3′-TTGGGAGTCTCGAAACGCTTA-5′ BAu-TT
3′-ACTTGGGAGTCTCGAAACGCTTA-5′ BAu-TTCA
3′-CGAACTTGGGAGTCTCGAAACGCTTA-5′ BAu-TTCAAGC

QUMA PCR conditions

Amplification was performed using 1 ng of the ligated product in a reaction volume of 10 µl. Mastermix (Roche) was prepared to a final concentration of 3.5 mM MgCl2 and 1 µM of each primer. PCR amplification consisted of 35 three-step cycles (10 s at 95 ºC, 20 s at 65ºC and 30 s at 72ºC), preceded by a denaturation step of 10 min at 95ºC.

AUMA PCR conditions

Amplification was performed using 3 µl of the ligated product (~12ng) in a reaction volume of 25 µl containing 0.8 µM of primer, 2U Taq DNA polymerase (Roche Diagnostics), 100 µM of each dNTP, and PCR buffer (10mM Tris-HCl pH 8.0, 50mM potassium chloride, 1.5mM $MgCl_2$). $MgCl_2$ was added to the PCR mix to a final concentration of 2.3mM. PCR conditions were identical for all primers and consisted of 35 three-step cycles (30 s at 95ºC, 30 s at 65ºC and 1 min at 72ºC). PCR cycles were preceded by denaturation for 1 min at 95ºC and ended with an extension step of 5 min at 72ºC. Reactions were performed in a Programmable Thermal Controller PTC100 (MJ Research Inc., Watertown, MA). The PCR products were diluted 1:4 in formamide dye buffer and denatured for 3 min at 95ºC. Three µl of denatured product were run on a 6% polyacrylamide 8M urea sequencing gel at 55W for 6 h. Initial experiments were visualized by silver staining of the gels. To improve the overall quality of fingerprints in normal-tumor comparisons, radioactive AUMA's were performed. One µCi α-^{32}P-dCTP (Amersham Biosciences) was added to the PCR mix. The gels were dried under vacuum at 85ºC and exposed to an X-ray film at room temperature without an intensifier screen for 3-6 days.

Supplemental Table 1. Sequences isolated from AUMA (part 1)

Band ID	*Size (bp)*	% GC	SmaI site in band ends	Chromosome map (Location[a])	Gene	Repetitive elements in band ends (5'/3')
AU-66	405	52	None / Sma I	1p36.12 (21835410-21835814)	RAP1GA1	MIRb / Alu Sp
AU-82	137	63	Sma I / Sma I	1q21.1 (147553872-147554008)	AK124550[b]	None / tRNA
AU-18	186	49	Sma I / Sma I	1q21.1 (147561130-147561315)	None	None / tRNA
AU-168	137	61	Sma I / Sma I	1q21.1 (147939418-147939554)	None[b]	None / tRNA
AU-47	186	46	Sma I / Sma I	1q21.1 (147946674-147946859)	None	None / tRNA
AU-136	603	47	None / Sma I	1q21.3 (151471158-151471495)	None	None / None
Ar7 c6	322	57	Sma I / Sma I	1q23.2 (156829373 – 156829688)	KCNJ10	None / None
AU-72	206	66	Sma I / Sma I	1q23.3 (159334967-159335172)	KARCA1[b]	None / None
AU-131	266	47	Sma I / None	1q23.3 (159767613-159767838)	None	None / None
AU-48	400	45	Sma I / Sma I	1q24.2 (165950975-165951374)	None[b]	None / tRNA
AU-45	792	33	Sma I / None	1q25.2 (174982256-174983047)	PAPPA2	Alu Y / None
Aj2 c1	464	49	Sma I / Sma I	1q32.2 (206389079 – 206389542)	MGC29875[b]	Alu Sq / None
AU-96	464	50	Sma I / Sma I	1q32.2 (208067307-208067770)	None[b]	Alu Sq / None
AU-64	438	51	Sma I / Sma I	1q32.3 (209975376-209975807)	None	None / None
AU-22	495	54	Sma I / Sma I	1q41 (219027406-219027900)	MOSC1[b]	None / Alu Sx
AU-57	210	56	Sma I / None	2p11.2 (88172667-88172870)	SMYD1	LTR33
AU-84	598	55	Sma I / Sma I	2p11.2 (85846392-85846989)	ATOH8	None / None
Ad3 c1	795	56	Sma I / Sma I	2p23.3 (27096094 – 27096888)	None	None / Alu Y
AU-23	195	59	Sma I / Sma I	2p24.1 (23712291-23712485)	BC069271 / AL834515	None / None
AU-113	989	52	None / Sma I	2p25.1 (10076680-10077668)	None	L1MA1 / None
AU-164	506	32	None / Sma I	2q13 (112505156-112505661)	None	None
Al1 c2	423	52	Sma I / Sma I	2q14.2 (121559187 – 121559609)	None	Alu Y / None
Ar3 c3	335	57	Sma I / Sma I	2q14.3 (127875175 – 127875509)	AF370412[b]	None / MIRb
At3 c3	288	52	Sma I / Sma I	2q21.1 (130817182 – 130817463)	IMP4[b]	None / Alu Sg
AU-8	388	54	Sma I / Sma I	2q21.1 (130817422-130817709)	IMP4[b]	None / Alu Sg
Ar3 c1	334	51	Sma I / Sma I	2q31.2 (180028122 – 180028455)	None	Alu Sx / None
AU-65	437	58	Sma I / Sma I	2q35 (219433280-219433716)	WNT6[b]	None / None
AU-114	322	53	Sma I / Sma I	2q37.3 (237777804-237778125)	None	Alu Sx / None
AU-40	263	56	Sma I / Sma I	3p22.1 (40631798-40632060)	None[b]	None / None
AU-33	393	53	Sma I / Sma I	3q26.2 (170347518-170347910)	EV11	None / None
Ao1 c2	373	60	Sma I / Sma I	4q31.3 (154527715 – 154528087)	TRIM2[b]	None/None
AU-105	570	44	Sma I / None	4q35.2 (190917251-190917820)	None	None / None
Ah1 c3	575	44	Sma I / Sma I	4q35.2 (191055406 – 191055980)	None	None / None
AU-86	320	43	None / Sma I	5p12 (45741040-45741359)	None	None / Alu Y
AU-135	380	46	None / Sma I	5p13.3 (33041752-33042131)	None	MER4B / Alu Y
AU-180	316	50	Sma I / None	5q33.1 (148845290-148845605)	None	Alu Jb / None
Ae2 c6	717	47	Sma I / Sma I	5q33.1 (151709944 – 151710660)	None	None / Alu Sx
Ar3 c7	335	56	Sma I / Sma I	5q35.1 (167846537 – 167846876)	RARS	None / Alu Y
AU-166	984	53	Sma I / Sma I	5q35.1 (171911500-171912483)	None	Alu Sc / Charlie 7
Ap1 c6	364	56	Sma I / Sma I	5q35.2 (175157319 – 175157677)	CPLX2[b]	None / MIR
AU-109	525	47	None / Sma I	6p12.3 (48557303-48557827)	None	L1MA3 / MER1B
AU-117	198	46	Sma I / None	6q13 (70235881-70236078)	None	HERV9
AU-79	164	60	Sma I / Sma I	6q21 (107887407-107887570)	PDSS2[b]	None / None
AU-125	307	41	Sma I / None	6q22.33 (130283187-130283493)	None	None / MLT1H1
AU-30	460	51	Sma I / Sma I	6q27 (165197147-165197606)	None	L1Mda / Alu Sq
Aq1 c1	354	53	Sma I / Sma I	6q27 (167075092 – 167075445)	RP6SKA2	None / None
Au1 c1	279	49	Sma I / Sma I	7p14.2 (35944501- 35944769)	None	LINE L2 / Alu Yd2
AU-50	268	52	Sma I / Sma I	7p15.2 (25857505-25857772)	None	None / None
Ah2 c10	563	48	Sma I / Sma I	7p22.2 (2893157 – 2893719)	None	LTR / Alu Sc
AU-167	234	59	Sma I / Sma I	8p23.1 (12508845-12509078)	None[b]	None / None
Av3 c1	245	49	Sma I / Sma I	8p23.2 (2994895 – 2995136)	None	None / None
Aq3 c6	345	58	Sma I / Sma I	8p23.3 (2007340 – 2007685)	MYOM2	LTR / Alu Y
Av2 c2	242	52	Sma I / Sma I	8q24.11 (117883300 – 117883541)	None	Alu Y / LTR
AU-140	423	49	None / Sma I	8q24.22 (132120338-132120760)	ADCY8	None / None
AU-41	444	57	Sma I / Sma I	8q24.3 (142607517-142607960)	None	None / None

Supplemental Table 1. Sequences isolated from AUMA (part 2)

Band ID	*Size (bp)*	% GC	Sma I targets in band ends (5'/3')	Chromosome map (Location[a])	Gene	Repetitive elements in band ends (5'/3')
AU-187	302	51	Sma I / None	9p11.1 (47044445-47044746)	None	Alu Sp / None
Ag2 c5	615	55	Sma I / Sma I	9p13.1 (38683335 – 38683950)	None	Alu Y / None
AU-16	603	58	Sma I / Sma I	9p24.2 (2231367-2231969)	None[b]	None / None
Ao1 c0	345	45	Sma I / Sma I	9q21.31 (79626081 – 79626425)	None	Alu Sx / Alu Jo
AU-68	310	55	Sma I / Sma I	9q22.1 (89779683-89779992)	None[b]	None / Alu Sg
AU-150	306	54	Sma I / Sma I	9q34.11 (129597457-129597762)	None	AluY
AU-61	361	46	None / Sma I	10q21.1 (53837803-53838163)	None	L2 / Alu Sc
AU-99	448	54	Sma I / Sma I	10q26.11 (120504982-120505429)	None[b]	None / Alu Y
Ar7 c5	324	55	Sma I / Sma I	10q26.13 (125843458 – 125843781)	None	None / Alu Sp
AU-1	448	59	Sma I / Sma I	11q12.1 (59089995-59090442)	None	None / None
AU-100	280	52	Sma I / Sma I	11q13.1 (64748856-64749135)	FLJ16331	Alu Sc / MIR
AU-108	327	51	Sma I / Sma I	11q13.2 (68668053-68668379)	None	Alu Sc / MER74A
Ao1 c1	333	45	Sma I / Sma I	11q13.4 (73980931 – 73981257)	POLD3[b]	MER44A / None
AU-155	219	53	Sma I / Sma I	11q21 (95763104-95763322)	None[b]	None / None
AU-17	668	54	Sma I / Sma I	11q23.3 (116164068-116164735)	None[b]	None / None
An1 c2	390	43	Sma I / Sma I	11q23.3 (116395630 – 116396019)	KIAA0999	Alu Sc / None
Ao2 c1	371	46	Sma I / Sma I	11q24.3 (129899206 – 129899576)	None	LINE L2 / Alu Sg/x
Ar3 c5	334	49	Sma I / Sma I	11q24.3 (130094098 – 130094431)	None	Alu Sx / None
AU-179	160	62	Sma I / Sma I	chr11 (not mapped)	None	None / None
AU-157	629	43	Sma I / None	12p12.1 (25467561-25468189)	None	Alu Sq / None
Aj1 c2	462	57	Sma I / Sma I	12p13.31 (6527527 – 6527992)	HOM-TES-103[b]	None / None
AU-10	673	43	None / Sma I	12p13.32 (3851608-3852280)	PARP11	MER39 / Alu Sg/x
Ar5 c2	327	59	Sma I / Sma I	12q13.11 (47397075 – 47397401)	CCNT1[b]	None / Alu Sg
Ai2 c5	511	39	Sma I / Sma I	12q21.31 (84722784 – 84723294)	PAMCI	Alu Y / None
Ai1 c1	520	51	Sma I / Sma I	12q24.11 (109686793 – 109687312)	None	None / Alu Sx
AU-13	249	53	None / Sma I	12q24.32 (124864438-124864686)	None	MERSA / Alu Sq
AU-127	328	38	Sma I / None	12q24.32 (126451330-126451657)	None	Alu Sx / L1ME1
AU-92	491	50	Sma I / Sma I	13q12.13 (26373474-26373964)	None	None / Alu Y
Ah2 c8	563	50	Sma I / Sma I	13q21.32 (64681523 – 64682085)	None	Alu Y / LTR
AU-174	617	46	Sma I / Sma I	14q32.2 (97438753-97439369)	None	None / Alu Y
AU-80	303	46	Sma I / Sma I	15q11.2 (18704809-18705111)	None	None / Alu Sq
Aj1 c9	468	52	Sma I / Sma I	15q12 (23575207 – 23575674)	ATP10A	Alu Sx / None
Aj1 c6	467	44	Sma I / Sma I	15q22.2 (58086012 – 58086478)	None[b]	None / None
AU-44	268	63	Sma I / Sma I	15q22.31 (63949311-63949578)	RAB11A[b]	None / None
Ao1 c6	374	53	Sma I / Sma I	15q23 (68573371 – 68573744)	None	None / MIR
AU-20	249	43	Sma I / Sma I	15q24.3 (74421623-74421871)	ISL2[b]	None / None
Al1 c6	428	49	Sma I / Sma I	15q26.3 (96971519 – 96971946)	None	Alu Sq / None
AU-60	460	49	Sma I / Sma I	16p11.2 (31100903-31101362)	FUS	Alu Sx / None
Ac4 c1	903	54	Sma I / Sma I	16p12.1 (27944929 – 27945831)	AY358206	None / Alu Sx
AU-34	434	52	Sma I / Sma I	16p13.13 (11742883-11743316)	TXNDC11	L2 / Alu Sx
AU-112	204	44	None / Sma I	16p13.2 (8523674-8523877)	None	None / None
As3 c6	312	63	Sma I / Sma I	16p13.3 (3160474 – 3160779)	None[c]	None / None
Au4 c1	274	57	Sma I / Sma I	16p13.3 (3162096- 3162369)	None	tRNA / None
AU-56	619	51	Sma I / Sma I	16p13.3 (3179681-3180299)	None[b]	tRNA / None
AU-126	315	54	Sma I / Sma I	16p13.3 (392266-392580)	DECR2[b]	Alu Sx / None
AU-104	504	58	Sma I / Sma I	16p13.3 (4466925-4467428)	HMOX2[b]	None / Alu Sx
Ah2 c6	563	52	Sma I / Sma I	16q21.1 (85551517 – 85552079)	None	None / None
Ai1 c3	515	55	Sma I / Sma I	17p11.2 (18206450 – 18206964)	SHMT1[b]	Alu Sx / MIR
AU-35	247	51	Sma I / None	17q21.31 (38831114-38831360)	None	Alu Sx / MIRb
AU-73	279	52	None / Sma I	17q22 (53031500-53031778)	MSI2	None / None
AU-55	480	49	Sma I / Sma I	17q24.3 (67555550-67556026)	None	None / Alu Sg
Ar7 c1	323	56	Sma I / Sma I	17q25.3 (73693981 – 73694303)	TK1	None / None
Ah1 c1	566	49	Sma I / Sma I	18p11.23 (7492459 – 7493024)	None	None / Alu Sg
AU-152	401	55	Sma I / Sma I	18p11.31 (3442451-3442851)	TGIF[b]	None / None
At1 c8	295	38	Sma I / Sma I	18q21.2 (51649154- 51649432)	None	None / None
AU-159	667	50	Sma I / Sma I	18q21.31 (53256789-53257455)	ONECUT2[b]	None / None
Al1 c4	429	58	Sma I / Sma I	18q22.3 (70314985 – 70315413)	CNDP2[b]	None / None

Supplemental Table 1. Sequences isolated from AUMA (part 3)

Band ID	*Size (bp)*	% GC	Sma I targets in band ends (5'/3')	Chromosome map (Location[a])	Gene	Repetitive elements in band ends (5'/3')
AU-186	419	59	Sma I / Sma I	19p13.11 (19600679-19601097)	None[b]	None / Alu Sc
AU-29	279	53	Sma I / Sma I	19p13.13 (13165698-13165976)	None	Alu Sx / Alu Sg/x
As3 c5	312	58	Sma I / Sma I	19p13.2 (9799858 - 9800169)	UBL5[b]	None / None
AU-149	549	53	Sma I / Sma I	19p13.3 (1970712-1971260)	None	None / Alu Y
At1 c3	295	59	Sma I / Sma I	19p13.3 (5273523 - 5273811)	None	None / Alu Sp
AU-24	250	48	Sma I / Sma I	19q13.2 (45624073-45624322)	None[b]	MIRb / MIR
AU-101	498	48	Sma I / Sma I	19q13.2 (46810414-46810911)	None	Alu Sx / None
Ao1 c4	371	50	Sma I / Sma I	19q13.32 (53550176 - 53550540)	AK001784	Alu Sx / MIRb
Aj1 c1	461	56	Sma I / Sma I	19q13.33 (54885550 - 54886010)	CPT1C[b]	None / None
AU-162	398	49	Sma I / Sma I	19q13.43 (61982532-61982929)	ZIM2	None / MER20
AU-2	251	58	Sma I / Sma I	20q11.23 (34926178-34926428)	None[b]	None / Alu Sg
Ak1 c7	434	53	Sma I / Sma I	20q13.12 (42690824 - 42691257)	ADA	None / None
AU-43	395	57	Sma I / None	20q13.13 (45894701-45895095)	None	MLT2B4 / None
AU-124	298	52	Sma I / Sma I	20q13.31 (55628389-55628686)	ZBP1	Alu Sx / None
AU-97	540	58	Sma I / Sma I	20q13.33 (61126250-61126789)	None	None / None
AU-133	398	56	Sma I / Sma I	21q22.13 (37004089-37004486)	SIM2	Alu Sx / None
AU-130	682	56	Sma I / Sma I	21q22.3 (43449720-43450401)	None	Alu Sg/x / None
AU-139	459	44	None / Sma I	22q11.1 (15411531-15411988)	None	ERVL-B4 / ERVL-B4
AU-106	432	53	Sma I / Sma I	22q12.3 (35742025-35742452)	TST	Alu Sg / None
Al1 c1	426	52	Sma I / Sma I	22q12.3 (35736578 - 35737003)	TST	Alu Sg / None
AU-14	290	50	Sma I / Sma I	Xp22.13 (17472645-17472934)	NHS	Alu Sx / None
AU-19	689	47	None / Sma I	Xq11.2 (65039112-65039800)	None	None / Alu Sg
AU-121	328	44	Sma I / Sma I	Xq13.3 (75267216-75267543)	None	LTR17 / L1PA11
AU-138	880	50	Sma I / Sma I	Yp11.2 (10493777-10494656)	None	LTR1 / None
AU-51	629	48	Sma I / Sma I	Yp11.31 (1736039-1736667)	None	Alu Sx / None

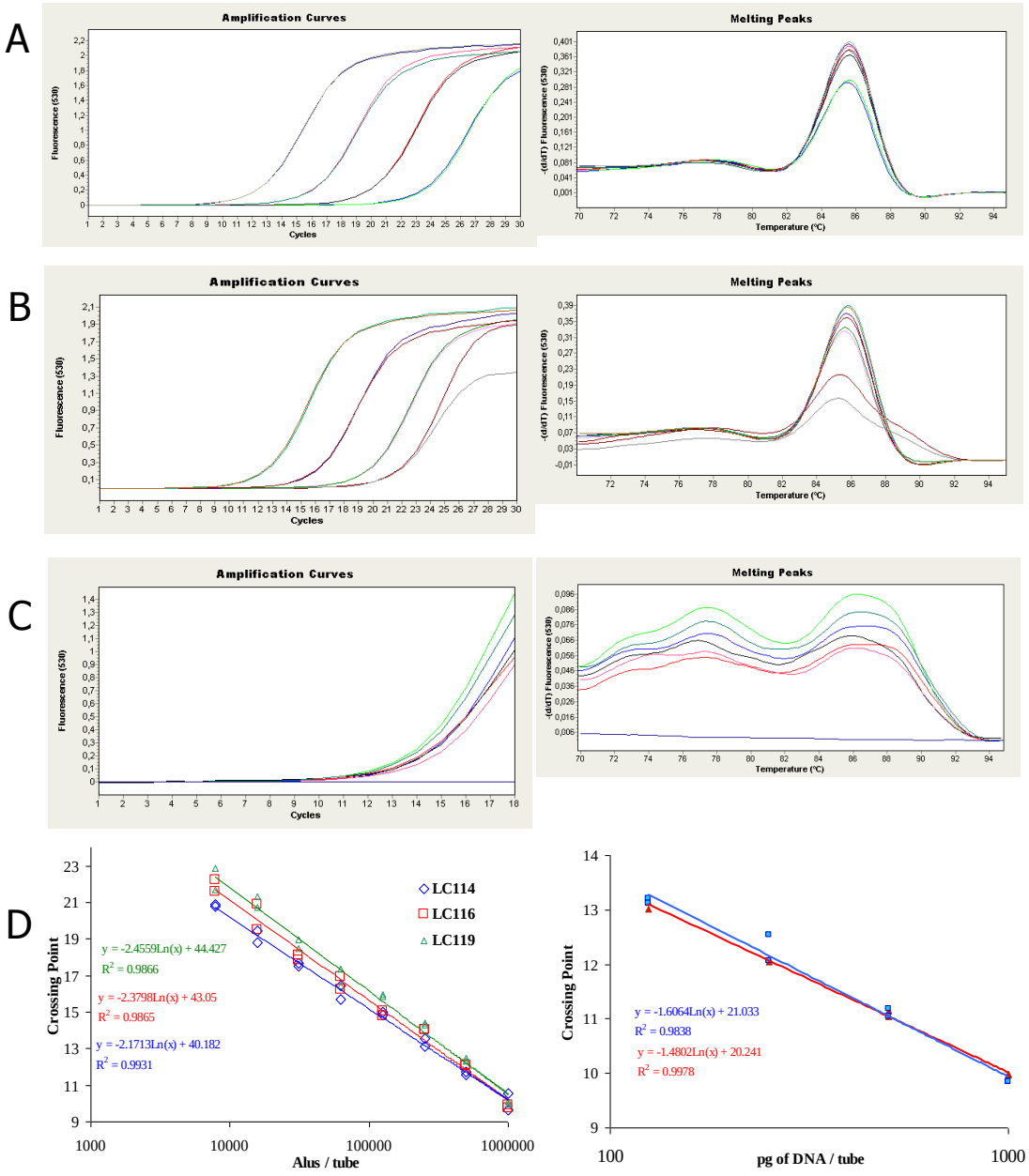
[a] Nucleotide position within the contig (strand +). NCBI Build 36.1 of the human genome.

[b] The whole sequence or a fragment of the sequence lays not further than 200 bp of a predicted CpG island.

Supplemental Table 2. Bisulfite sequencing and gene expression primer sequences

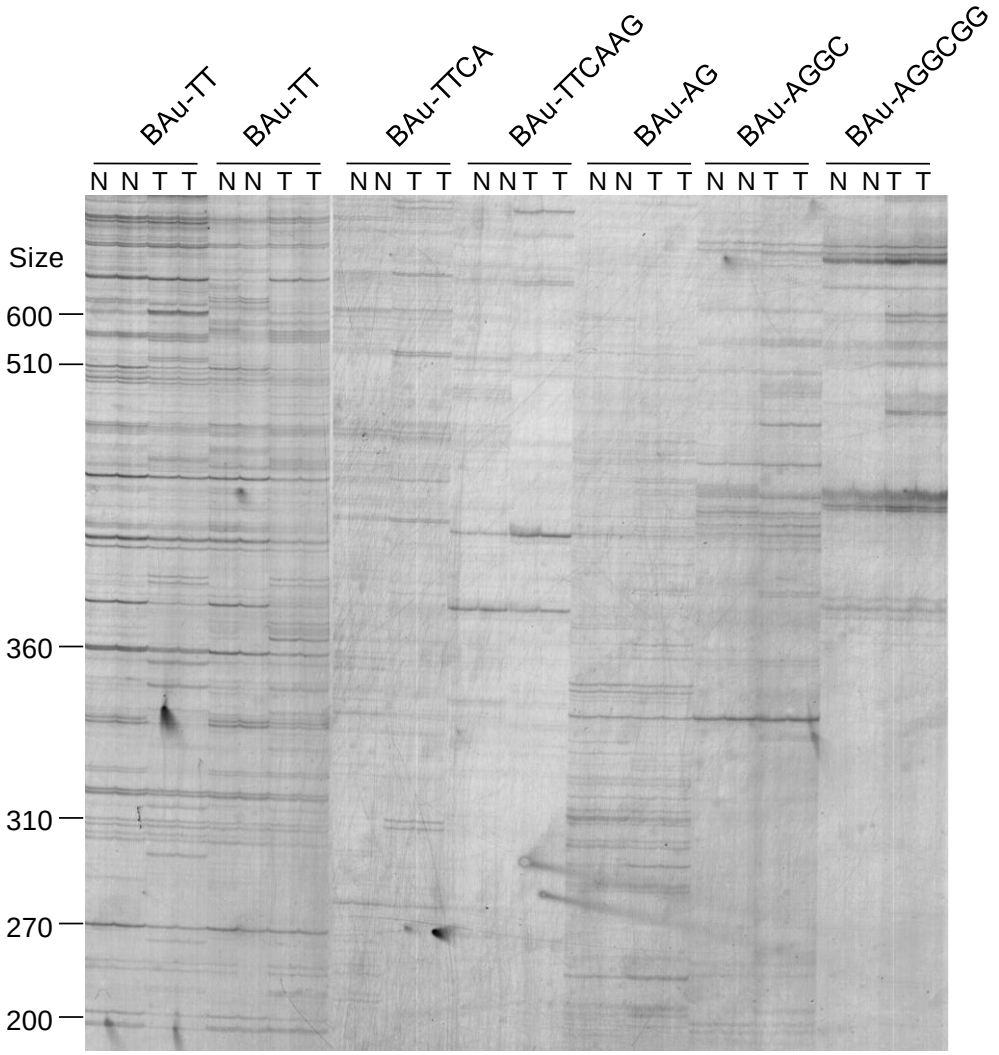
Purpose	Primer Sequence (5'-3')
Aq3 band (MLT1A element), bisulfite sequencing, first PCR	GTTGGATTTAAATTTTTGGTT TTAACAACAAACACACTCAA
Aq3 band (MLT1A element), bisulfite sequencing, nested PCR	GATTAGGTTTTTAGTTTTGT CCCACAACATAAAACACC
Aq3 band (AluY element), bisulfite sequencing, first PCR	GTGTTTTATGTTGTGGGG CCCAAACTTCTCCAAAAACTA
Aq3 band (AluY element), bisulfite sequencing, nested PCR	ATTTGAGTGTGTTTGTTGTT ATAACAAAAACACTACACAA
Aq3 band (AluY element), ChIP analysis	GTGGGGCGGTGACAAAGACG AGGGGCACTGCACAGAAACGAT

Supplemental Figures



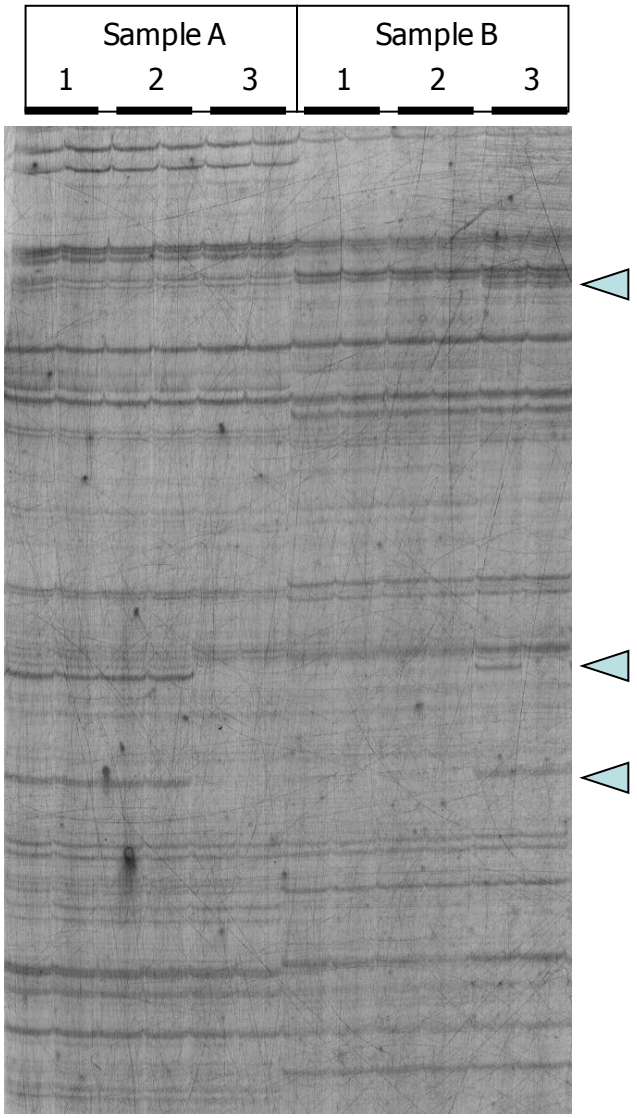
Supplemental Figure 1

Real time analysis of QUMA amplification and melting curves. (A) QUMA analysis of the AluSx sequence used as standard in concentrations ranging from 10^4 to 10^7 molecules per reaction tube. Due to the presence of a unique Alu sequence, melting curves show a single peak. (B) QUMA analysis of the AluSx sequence used as standard (equivalent concentrations ranging from 10^4 to 10^7 molecules per ng of DNA were added) spiked on a genomic DNA digested with XmaI. Sensitivity was in the range of 10^4 molecules per ng of DNA. (C) QUMA analysis of normal and tumor samples. Melting curves show multiple peaks due to the presence of different Alu elements. (D) Three QUMA dilution curves in three independent experiments (left) and QUMA dilution curves of two sample DNAs (right). In all cases analysis was performed in duplicate. The linearity of the response is observed in standard and sample dilutions.



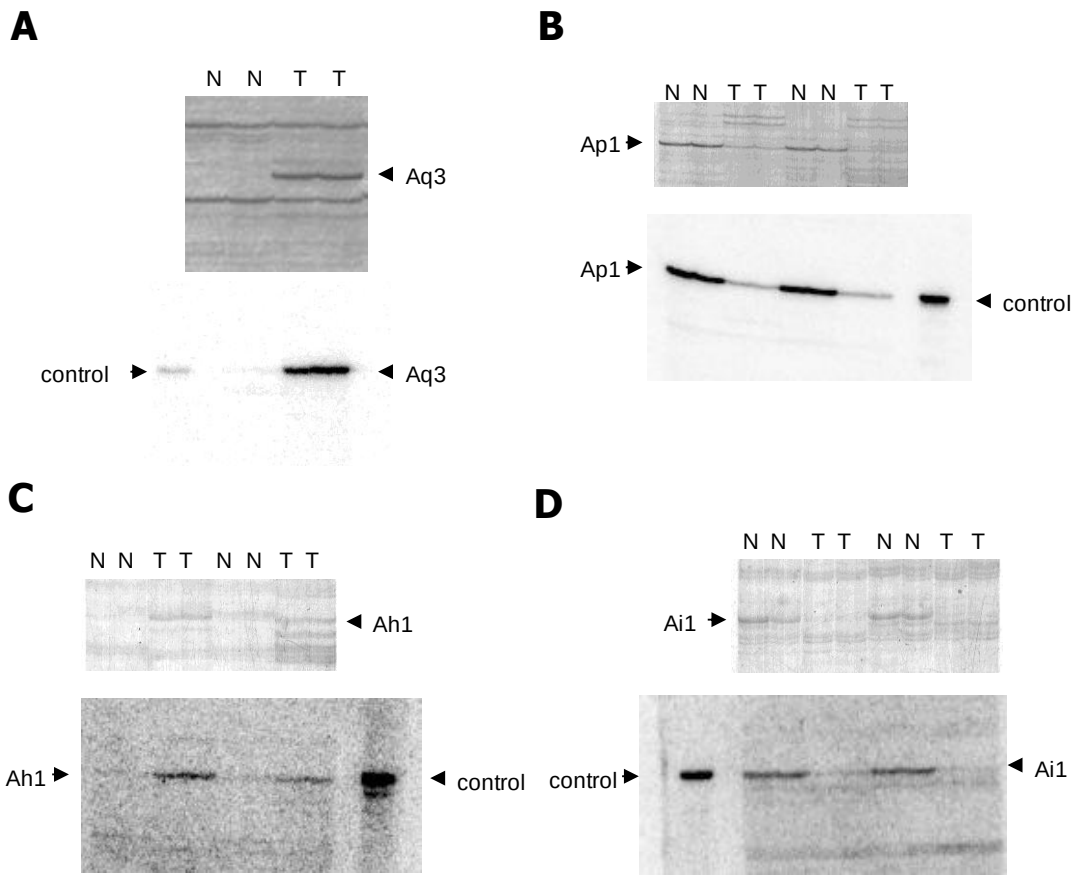
Supplemental Figure 2

Representative fingerprints of AUMA products obtained with primers amplifying towards Alu promoter (BAu-TT, BAu-TTCA and BAu-TTCAAG) and towards Alu poly-A tail (BAd-AG, BAd-AGGC, BAd-AGGCGG). Duplicate analysis of normal-tumor pairs was performed. One case is shown for each primer except for Bau-TT, for which two cases are illustrated. Approximate band size (bp) is shown at the left. Gels were silver stained for band visualization.



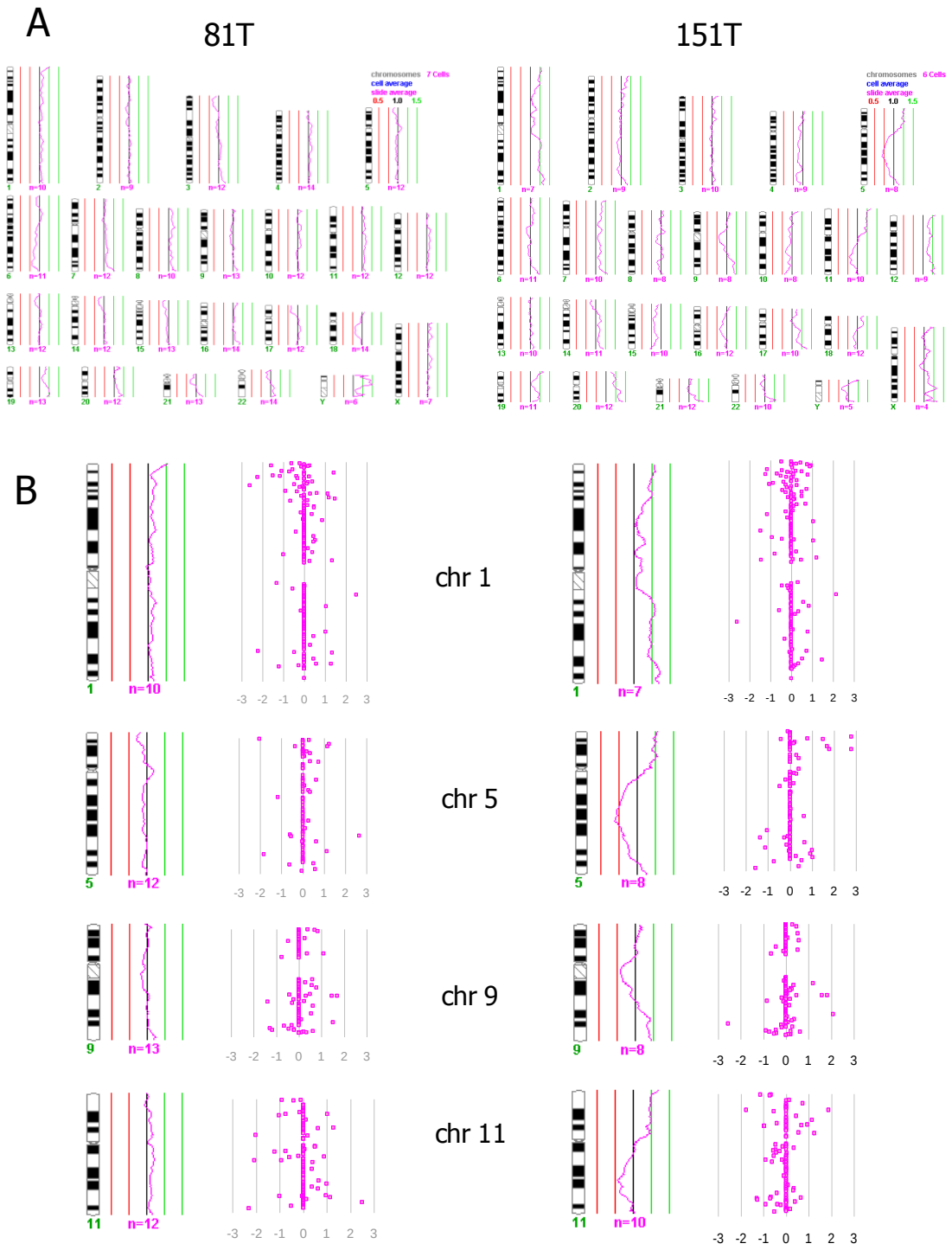
Supplemental Figure 3

Interassay reproducibility of AUMA. Two sample DNAs (A and B) were analyzed in three different experiments (1 to 3). PCR was performed in duplicate for each replica, run in denaturing polyacrylamide gel and silver stained. While most bands maintained the display among replicas, a few bands (marked with and arrowhead) showed inconsistent display in both intraassay and interassay replicates. Bands with irreproducible display were not considered for further analysis and were not included in the accounting of normal-tumor differences.



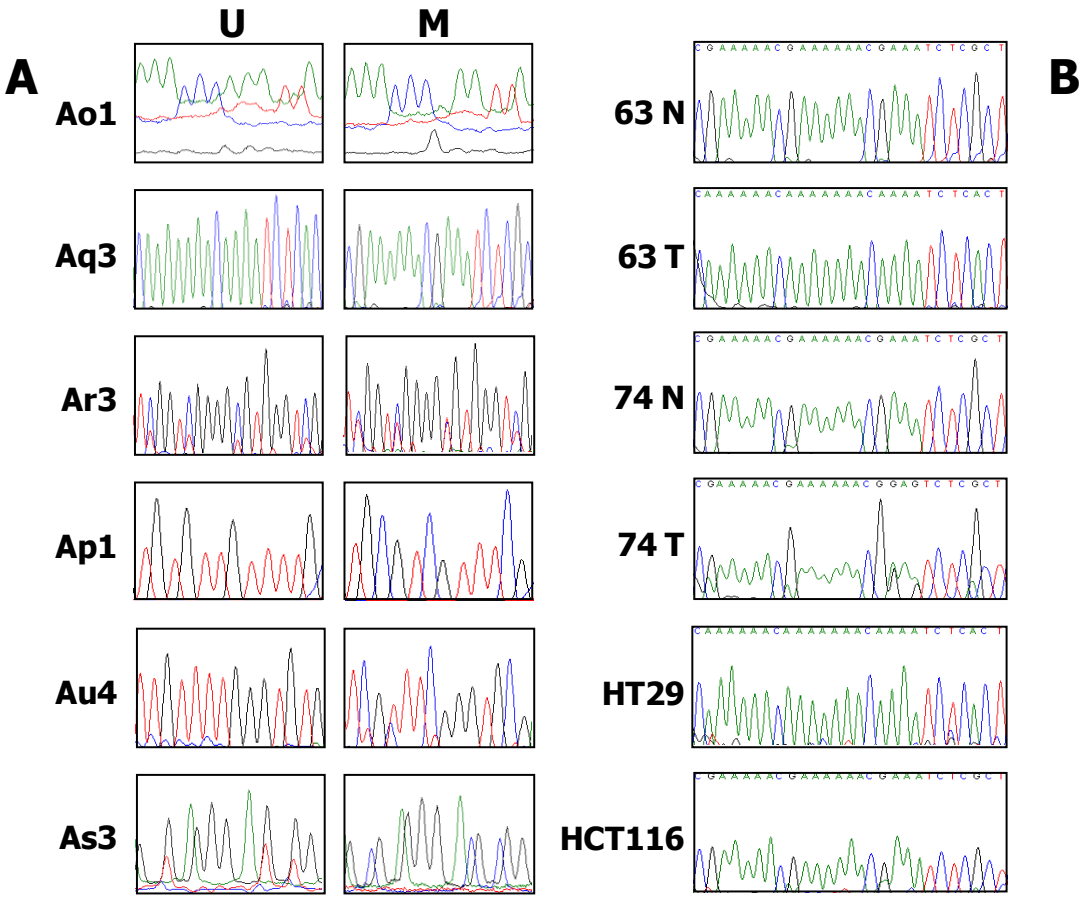
Supplemental Figure 4

Southern Blot confirmation of AUMA isolated bands. Two hypomethylated (A and C) and two hypermethylated AUMA bands (B and D) were labeled by PCR amplification in the presence of [α-^{32}P]dCTP. Top of each panel corresponds to the silver stained AUMA fingerprint that harbors the band to be tested while bottom of each panel corresponds to the AUMA gel blotted to a nitrocellulose membrane hybridized with the band of interest. For each hybridization a positive control consisting in the isolated band that had to be hybridized was included.



Supplemental figure 5

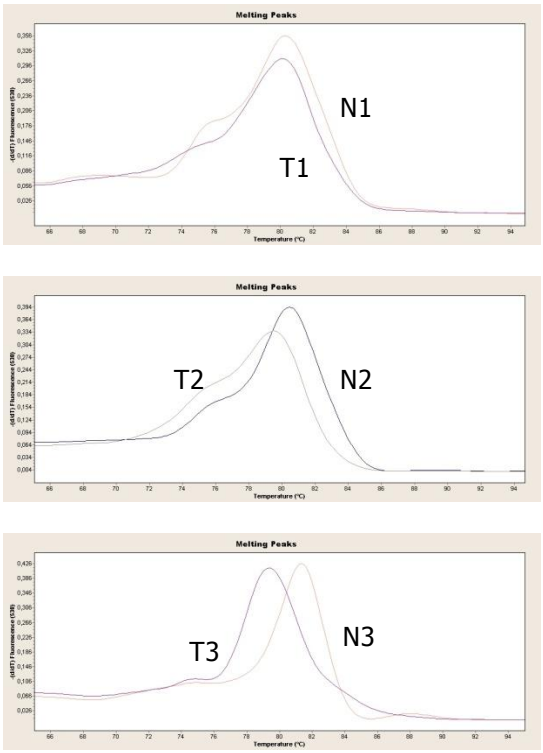
(A) CGH ideograms of two tumors are shown. (B) Comparison of the genomic profile of selected chromosomes 1, 5, 9 and 11 and the corresponding differential methylation profiles determined by competitive hybridization of AUMA products from normal and tumor tissue to BAC arrays are shown. Positive values (to the right) indicate hypomethylations, negative values (to the left) indicate hypermethylations.



Supplemental Figure 6

Direct sequencing of Bisulfite treated DNAs. (A) Confirmation of the methylation change seen in AUMA fingerprints of 6 different bands. All bands except Ao1 and Aq3 have been sequenced with forward primers. Raw data electropherograms for bands Ao1 and As3 are shown. (B) Sequences corresponding to Aq3 hypomethylated band. A completely demethylated tumor (63T), a partially demethylated tumor (74T) and a demethylated cell line (HT29) are shown. All normal tissues and HCT116 cell line are heavily methylated at all CpGs. The sequence shown corresponds to a fragment of the Alu Y element. All sequences were obtained with reverse primer.

A



Supplemental Figure 7

Dissociation curves of bisulfite treated DNAs amplified by real time PCR. Results corresponding to three normal (N)-tumor (T) pairs are illustrated. Methylated sequences exhibit a higher melting temperature than unmethylated sequences due to a higher CG content. (A) Analysis of the AluSx sequence represented by Aq3 AUMA band. In all normal cases the sequence is methylated. Tumor 1 is methylated, tumor 2 is partially methylated, and tumor 3 is fully unmehylated.

CAPÍTOL 3

Bivalent domains enforce transcriptional memory of DNA methylated genes in colorectal cancer.

Jairo Rodríguez, Mar Muñoz, Laura Vives, Gus Frangou, Mark Groudine, Miguel A. Peinado.

Article en preparació.

Bivalent domains enforce transcriptional memory of DNA methylated genes in colorectal cancer.

El DNA es troba empaquetat en el nucli eucariòtic en forma d'una nucleoproteïna anomenada cromatina, la qual és el substrat de la regulació epigenètica. Així doncs, de l'estructura de la cromatina en depenen processos tant importants com la replicació, recombinació i transcripció, a més a més d'afectar el grau d'empaquetament dels cromosomes. L'alteració dels mecanismes de regulació de la cromatina *in vivo* comporten l'aparició de processos com el càncer.

La regulació epigenètica es dóna a dos nivells: la metilació del DNA i les modificacions de les histones, i totes dues tenen efectes sobre l'estructura de la cromatina, a través del posicionament dels nucleosomes, les unitats bàsiques de la cromatina. L'estudi dels patrons de metilació en càncer indiquen que la metilació del DNA en regions promotores és un mecanisme molt freqüent d'inactivació gènica en tumors, la qual va associada a patrons de modificacions d'histones de cromatina inactiva.

Els darrers estudis globals sobre la distribució genòmica de modificacions d'histones en càncer indiquen que els gens silenciats amb promotors hipermetilats presenten uns patrons de cromatina específics sobre els seus promotors molt similars als descrits en cèl·lules mare embrionàries (ESCs), els dominis bivalents, en els quals conviuen dues marques tradicionalment relacionades amb estats transcripcionals oposats, la trimetilació de la histona H3 en la lisina 4 (H3K4me3) i 27 (H3K27me3). Aquests resultats aporten noves evidències moleculars a un hipotètic origen *stem* del càncer.

En aquest treball hem aprofundit en l'estudi dels patrons epigenètics aberrants en càncer colorectal (CCR), amb una especial atenció sobre el silenciament transcripcional associat a la hipermetilació de promotors. Així hem pogut identificar un nou conjunt de gens veïns hipermetilats i silenciats en CCR en la regió genòmica 5q35.2, els quals posseeixen promotors amb dominis bivalents. Aquest fet evidencia l'existència d'una relació entre les maquinàries de metilació del DNA i de regulació de la memòria cel·lular en ESCs i cèl·lules diferenciades a través dels complexes que regulen els nivells de H3K4me3 i H3K27me3, les proteïnes Trithorax i Polycomb, respectivament.

De forma inesperada, els resultats obtinguts demostren que la reexpressió per mitjà de drogues desmetilants del DNA i inhibidors de HDACs dels gens silenciats no aconsegueix esborrar els dominis bivalents, fet que posa de manifest l'existència de mecanismes de memòria cel·lular que mantenen l'estat transcripcional en les cèl·lules tumorals i molt probablement en les cèl·lules normals. Contràriament, gens que s'expressaven a un alt nivell en el teixit normal passen a infraexpressar-se després del tractament amb les drogues, adquirint modificacions d'histones pròpies d'un estat inactiu.

Bivalent domains enforce transcriptional memory of DNA methylated genes in colorectal cancer.

Jairo Rodriguez[1,2], Mar Muñoz[2], Laura Vives[2], Gus Frangou[3], Mark Groudine[3], Miguel A. Peinado[1,2]

[1] Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet, Barcelona, Spain
[2] Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC), Badalona, Barcelona, Spain
[3] Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

Running title: Bivalent domains enforce transcriptional memory

Keywords: DNA methylation, transcriptional silence, bivalent domains, demethylating agents, cancer.

Correspondence: Miguel A. Peinado
Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC)
Crta Can Ruti, Camí de les Escoles, Badalona, 08916 Barcelona, Spain
email: map@imppc.org
Tel 34-93 4978694
Fax 34-93 4978697

Abstract

There is little doubt that epigenetic abnormalities lay at the foundation of carcinogenesis, providing an alternative to genetic disruption for almost all defined pathways. Although the list of silenced genes with promoter DNA methylation is long and still growing, no answer is given as to why some genes become DNA-methylated during carcinogenesis. Recent works reporting on genome-wide maps of histone modifications have provided some clues, indicating that stem-cell chromatin signatures are found over the promoters of genes that will become methylated in adult cancers. Here we report a novel set of neighboring DNA methylated genes in a region spanning 1.2 Mb in colorectal carcinoma cells, where promoter DNA methylation occurs only at genes that are kept in a low transcriptional state by the presence of bivalent domains. Upon drug-induced demethylation, silenced genes become upregulated but bivalent domains do not resolve into an active chromatin conformation. Furthermore, active genes in the region become downregulated after drug treatment, accompanied by a *de novo* formation of bivalent domains. Our results demonstrate that bivalent domains mark the promoters of genes that will become DNA methylated in tumor cells, enforcing transcriptional silence and thus connecting stem cell biology and tumor development.

Introduction

During decades, scientists have focused on the understanding of the origins of cancer, trying to dissect the sequence of events that deregulate the complex networks governing homeostasis in multicellular organisms. Cancer is a disease that ultimately alters gene expression. In this context, even though we have more precise maps of the human genome[1], this has proved insufficient to understand how genetic programs are read and translated into transcriptional patterns in normal and pathological cells. A large body of data indicates that information other than the encoded within the DNA sequence is required. This "other" information, termed epigenetic (which literally means "over genetic"), is defined as the heritable changes that affect gene expression without altering the DNA sequence[2]. To date there is little doubt that epigenetic events, in an intimate cooperation with genetic events, are involved in every step of tumorigenesis. Moreover, for all the key cellular pathways disrupted in human cancers, epigenetic alterations provide an alternative mechanism to genetic inactivation of tumor suppressor genes (Reviewed in[3-5]) and, for a growing number of genes, epigenetic inactivation represents the only inactivating mechanism[3,4].

Epigenetic information is encoded as a heritable combination of chemical modifications to both DNA and its packaging histones (Reviewed in[6,7]). Methylation of the cytosine base within the CpG dinucleotide is the main epigenetic modification of the DNA (Reviewed in[2,8]). Much of the human genome is CpG depleted, but this dinucleotide can be found at close to its expected frequency in small genomic regions (200bp to a few kb), known as CpG islands[9,10]. These areas are usually "protected" from methylation and are located in the proximal promoter regions of 75% of human genes[2,8,11]. Methylated CpG islands are strongly and hereditably repressed[8] and therefore, DNA methylation has been considered as a mark of long-term inactivation[8,12,13].

Abnormal DNA methylation patterns are a common hallmark of all human cancers (Reviewed in[14]). Even though widespread genomic hypomethylation was the first DNA methylation abnormality detected in tumors (Reviewed in[15,16]), most investigations have focused on the study of abnormal *de novo* methylation of CpG islands and its association with the transcriptional silencing of many cancer-related genes[3,4,14,15,17-19]. To date, little is known about the mechanisms that bring *de novo* methylation to the promoter regions of genes that are usually methylation-free, during carcinogenesis. In the above context, chromatin, and more importantly, histones and their modifications, are revealed as a key component of the silencing machinery that may have a profound impact on the DNA methylation patterns. Silenced genes, including those with promoter DNA methylation, have been found to contain a specific code of histone modifications across the promoter region and within the body of the gene, that are thought to be a signature of the sequence of silencing events that lead to transcriptional inactivation (Reviewed in[20]). Additionally, an extra layer of complexity is added by the fact that the DNA methylation mark itself can be read by specific proteins that are able to ultimately recruit activities that alter chromatin composition (Reviewed in[8]). This highlights the notion that a crosstalk, but not a unidirectional flow of information, exists between DNA methylation and histone modifications to orchestrate transcriptional silencing.

In recent years, a very exciting concept has emerged which hypothesizes the existence of cancer stem cells, which would be responsible for perpetuating the tumor population (Reviewed in[5]). A growing number of evidences suggest that disruption in cancer cells of the two main cell memory systems involved in the maintenance of a stem cell state, Trithorax (Trx) and Polycomb Group (PcG) proteins may be involved in tumor-associated aberrant gene silencing and promoter DNA methylation. In this context, epigenetic stem cell-like signatures have been found in a large set of DNA-methylated genes in different types of tumors: trimethylated lysine 27 has been found to pre-mark genes in colorectal and prostate cancer cells[21]; methylated lysine 27 and components of the Polycomb Repressor Complex 2 (PRC2) responsible for the deposition of this mark, SUZ12 and EED, have been found over the promoters of a different subset of genes in primary colorectal tumors[22]; and the activation mark dimethyl lysine 4 in histone H3 together with methylated lysine 27 have been found to coexist over the promoter regions of DNA methylated genes in embryonic carcinoma (EC) cells[23], in an epigenetic landscape that mimics the bivalent domains described by Bernstein and colleagues for a subset of key developmentally regulated genes in embryonic stem (ES) cells[24].

The above scenario suggests that the aberrant *de novo* DNA methylation that so commonly affects cancer-related genes, could be a direct consequence of an aberrant chromatin environment, driven by the abnormal presence of the polycomb-mediated mark methylated lysine 27. Further evidence supporting this hypothesis comes from experiments conducted by Vire and collaborators[25], revealing that the histone methyltransferase responsible for the deposition of the methyl mark over lysine 27, Enhancer of Zeste Homolog 2 (EZH2), which forms part of the Polycomb Repressor Complex 2 (PRC2)(Reviewed in[26]), physically interacts with DNMT activities to bring *de novo* DNA methylation over the promoters of polycomb regulated genes. However, recent experiments by McGarvey and colleagues[27] show that only lightly DNA-methylated cancer-related genes become upregulated upon EZH2 knockdown, while densely DNA-methylated genes remain unaffected in terms of DNA methylation and transcriptional activity. Consequently, further experiments must be performed to ascertain the possible causes of aberrant *de novo* DNA methylation in cancer.

Global approaches that couple the use of antibodies against methylated cytosine to microarray analysis have helped to better understand how abnormal methylation patterns are established during tumor development, bringing to light some unveiled features. It has been recently shown that de novo methylated genes in the colorectal cancer cell line CaCo2, which fall in a few functional categories, tend to cluster together in the genome, suggesting the existence of an instructive *de novo* methylation mechanism directing DNA methylation to specific set of closely related genes[28]. In agreement with these observations, two genomic regions where DNA methylation encompasses large portions of the cancer cell genome, affecting multiple neighboring CpG island-containing genes, have been described in a cancer system[29,30]. Similarly, other regions have been described where silencing of multiple neighboring genes in large genomic regions occur in the absence of DNA methylation[31]. Taken together, these experiments highlight the necessity of exploring the cancer cell epigenome to search for clusters of DNA-methylated genes in order to gain insights into how aberrant epigenetic landscapes are set during tumorigenesis.

Here, we report the characterization of the region around a novel DNA-methylated gene, the CPLX2 locus, showing a novel set of neighboring genes that undergo frequent promoter DNA hypermethylation in human primary colorectal tumors and cell lines. The genes, which include hormone receptors, cell-to-cell adhesion and signaling molecules and tissue specific genes, are found in a genomic region spanning over 1.2 Mb mapping in 5q35.2. Chromatin analysis has revealed that DNA methylated genes are kept in a silent state by the presence of a stem cell-like chromatin pattern defined by the simultaneous presence of the trimethylated forms of lysines 4 and 27 in histone H3 and hypoacetylated forms of histones H3 and H4. We further demonstrate the direct implication of the polycomb repressor complex 4 in the silencing of this region.

Reactivation of the silenced genes with the demethylating agent 5azaC and the HDAC inhibitor TSA results in a striking pattern of histone modifications across the entire region, indicating that drug treatment is not able to erase chromatin patterns at DNA-methylated promoters, but rather forces the cell to retain the bivalent domains by adjusting the levels of both lysine 4 and 27 trimethylation.

Our results show for the first time the concurrent DNA methylation of neighboring tumor-associated genes across a large genomic region, which are embedded in a stem cell-like chromatin pattern where DNA methylation locks in a silent state. These experiments, in concordance with recent reports, directly point to a direct involvement of the two main cell-memory systems, polycomb and trithorax, in the aberrant promoter DNA methylation that so commonly affects silenced genes in cancer.

Materials and methods

Tissues and cell lines

A series of 121 colorectal tumors (113 carcinomas and 8 adenomas) and their paired areas of normal colonic mucosa was used in the analysis. Samples were collected simultaneously as fresh specimens and snap-frozen within 2 h of removal and then stored at -80ºC. All samples were obtained from the Ciutat Sanitària i Universitària de Bellvitge (Barcelona, Spain). Transformed cell content was higher than 75% in most tumor specimens as assessed by histological examination. The study protocol was approved by the Ethics Committee. Human colon cancer cell lines (HT29, SW480, HCT116, LoVo, DLD-1, CaCo-2 and LS174T) were obtained from the American Type Culture Collection (ATCC; Manassas, VA). The KM12C and KM12SM cell lines were generously provided by A. Fabra. DNA from tumor-normal pairs was obtained by conventional organic extraction and ethanol precipitation. DNA purity and quality was checked in 0.8% agarose gel electrophoresis. RNA from cell lines was obtained by phenol-chloroform extraction and ethanol precipitation following standard procedures.

Amplification of Unmethylated Alu (AUMA)

The genomic region methylated in colorectal tumors and cell lines described here was identified as a single band undergoing recurrent hipermethylation in DNA methylation fingerprints (Figure

1A), obtained by the application of a novel DNA methylation screening method named Amplification of Unmethylated Alu (AUMA)[32]. Briefly, one microgram of DNA was digested with 20 U of the methylation sensitive restriction endonuclease SmaI (Roche Diagnostics GmbH, Mannheim, Germany) for 16h at 30ºC, leaving cleaved fragments with blunt ends (CCC/GGG). Adaptors were prepared incubating the oligonucleotides Blue (CCGAATTCGCAAAGCTCTGA) and the 5' phosphorilated MCF oligonucleotide (TCAGAGCTTTGCGAAT) at 65ºC for 2 min, and then cooling to room temperature for 30-60 min. One microgram of the digested DNA was ligated to 2nmol of adaptor using T4 DNA ligase (New England Biolabs, Beverly, MA). The products were purified using the GFX Kit (Amersham Biosciences, Buckinghamshire, UK) and eluted in 250 µl of sterile water.

Subsequent PCR amplification using a single primer named BAuTT (ATTCGCAAAGCTCTGAGGGTT) allowed the generation of fingerprints that were resolved on denaturing sequencing gels. Radioactive AUMA's were performed for normal-tumor comparisons, yielding a highly reproducible and clear band patterning that allowed a feasible identification of specific bands undergoing tumor-associated hypermethylation (bands with a decreased intensity in the tumor compared to the normal tissue) and tumor-associated hypomethylation (bands *de novo* appearing or with an increased intensity in the tumor, compared to the normal tissue).

Bisulfite genomic sequencing

The bisulfite reaction was carried out on 2µg of mechanically sheared DNA for 16h at 55˚ under conditions previously described[29]. Prior to sequencing, DNA was amplified using a nested or semi-nested PCR approach, as appropriate. A minimum of two independent PCRs were carried out and pooled together to ensure a representative sequencing. Alternatively, DNA methylation was assessed by real-time PCR melting curve analysis. Positions and sequence composition of the analyzed CpG islands are listed in Supplementary Table 1 and bisulfite genomic sequencing PCR primers are listed in Supplementary Table 2.

Gene expression analysis

Whole cDNA was obtained by retrotranscription of 500 ng of whole RNA with M-MLV retrotanscriptase (Invitrogen, Carlsbad, CA) using random hexamers (Amersham Biosciences) at 37ºC for 1h. cDNA levels were quantified using the LightCycler 2.0 real time PCR system with Fast Start Master SYBR Green I kit (Roche Diagnostics). For a 10 µl PCR reaction volume, 1 µl of cDNA and 9 µl of mastermix were added to each capillary. Mastermix was prepared to a final concentration of 3.5 mM MgCl2 and 1 µM of each primer. Primers used for expression analysis are listed in Supplementary Table 3.

Chromatin Immunoprecipitation (ChIP) assay

Briefly, $6x10^6$ cells were washed twice with PBS and cross-linked on the culture plate for 15 min at room temperature in the presence of 0.5% formaldehyde. Cross linking reaction was stopped by adding 0.125M glycine. All subsequent steps were carried out at 4˚C. All buffers were pre-chilled and contained protease inhibitors (Complete Mini, Roche). Cells were washed twice with PBS and then scraped. Collected pellets were dissolved in 1ml lysis buffer (1% SDS, 5mM EDTA,

50mM Tris pH8) and were sonicated in a cold ethanol bath for 10 cycles at 100% amplitude using a UP50H sonicator (Hielscher, Teltow, Germany). Chromatin fragmentation was visualized in 1% agarose gel. Obtained fragments were in the 200 to 500 bp range. Soluble chromatin was obtained by centrifuging the sonicated samples at 14.000g for 10 min at 4°C. The soluble fraction was diluted 1/10 in dilution buffer (1% Triton X-100, 2mM EDTA, 20mM Tris pH8, 150mM NaCl) then aliquoted and stored at -80°C until use.

Immunoprecipitation was carried out at 4°C by adding 5 to 10 µg of the desired antibody to 1ml of chromatin. Soluble chromatin was immunoprecipitated with specific antibodies against BmiI, acetylated (Ac) H3K9, Ac H3K14, Ac H4K12, Ac H4K16, dimethylated (me2) H3K9me2, trimethylated (me3) H3K9me3, H3K27me3 (Upstate, Millipore, Billerica, MA), H3K4me3 (Abcam, Cambridge, UK), EZH2 (Santa Cruz, Santa Cruz, CA), SirT1 (Delta Biolabs, Gilroy, CA) and IgG negative control (Jackson Immunoresearch, West Grove, PA). Chromatin-antibody complexes were pulled down using a 50% slurry protein A/G (Upstate) and subsequently washed and eluted according to the manufacturer's instructions. Eluted samples were cleaned using the Jet quick PCR product purification spin kit (Genomed, Löhne, Germany).

Enrichment for a given chromatin modification was quantified as a fold enrichment over the input using quantitative real time PCR platform (Light Cycler 2.0, Roche). The input value was obtained for each amplification reaction and sample as the mean value of the 1/10 and 1/50 dilutions. Amplification efficiency and linearity were assessed by using serially-diluted samples that allowed the generation of a standard curve for every PCR. All quantifications were performed in duplicate. Primers used for Real Time PCR ChIP analysis are listed in Supplementary Table 4.

Drug treatments

Briefly, HCT116 cell line was seeded at low density 24h prior to treatment. 5-aza-2'-deoxicytidine (5azaC) was added to media for 48h, after which drug was removed and cells were allowed to recover in fresh media for further 24h before harvesting DNA and RNA. For 5azaC and Trichostatin A (TSA) co-treatments, cells were grown in the presence of 5azaC for 48h after which TSA was added for further 18h. Before harvesting RNA and DNA, fresh media was added for further 24h.

Results

Discovery of a new DNA-methylated gene in colorectal tumors

By applying AUMA to a series of colorectal tumors and their paired normal mucosas[32] (unpublished results), we identified a specific band, termed Ap1 (Figure 1A), which was found extensively methylated in 36 out of 50 (72%) tumor samples. Gel isolation, cloning and sequencing, allowed us to map the Ap1 band on the genomic region 5q35.2. As expected, the Ap1 band contained the two SmaI sites on both ends. The SmaI site on the 3'end of the band was inside of a MIR interspersed repeat (Figure 1B). Direct sequencing of bisulfite treated DNAs showed that this MIR repeat is devoid of methylation in normal samples but partially to

completely methylated in tumor samples and colorectal cancer cell lines (Figure 1C). The 5' end of the Ap1 band mapped inside of the first intron of the Complexin 2 (CPLX2) gene (Figure 1B). Since this SmaI site is only 31 bp away from the CPLX2 CpG island (CpG125), we assessed its methylation status by bisulfite sequencing, showing that this normally unmethylated CpG island can be found partially to heavily methylated in colorectal tumors and cell lines (Figure 1C).

Methylation around the CPLX2 locus encompasses multiple CpG islands

As depicted in figure 2A, the genomic region containing the CPLX2 gene is moderately rich in genes and CpG islands. Given that previous works had reported that large genomic regions may undergo concurrent methylation of multiple CpG islands in tumors[29], we further extended the methylation analysis to genes upstream and downstream the CPLX2 locus. In a first setting, CPLX2 locus neighboring CpG islands were analyzed for DNA methylation by bisulfite sequencing in five normal-tumor pairs and 8 colorectal cell lines (figure 2B). The first island upstream of the CPLX2 gene (CpG72) is located 137.8 Kb away and is associated to the Histamine Receptor H2 (HRH2) gene, while the first island downstream (CpG125) is situated 73.8 Kb away and is associated to a transcriptional variant of the CPLX2 gene named AK124.837 (Figure 2A), which has been reported to be transcriptionally controlled from an alternative promoter[33].

In a similar way to CPLX2 CpG island (CpG118), the HRH2 CpG island (CpG78) was found to be completely devoid of methylation in 4 normal tissues (we detected low methylation in one normal tissue), while moderate methylation levels were detected for the same samples in the CpG125 island. Both islands were found heavily methylated in all tumors and cell lines (CpG72) and in 6 of 7 cell lines and 4 of 5 tumors (CpG125). The next two islands to be tested were CpG93, associated to the Sideroflexin 1 (SFXN1) and CpG52 associated to the Tho Complex 3 (THOC3), laying 317.5 Kb upstream and 170.2 Kb downstream of the CPLX2 CpG island respectively. Both were the first islands to be found completely methylation-free in all normal-tumor pairs and cell lines. The last two islands to show methylation around the CPLX2 gene were CpG145 associated to the Dopamine Receptor D1 (DRD1), laying 351.2 Kb upstream, and the CpG78, located inside of a segmental duplication, mapping 262.8 Kb downstream. In the case of the DRD1 CpG island, it was found methylation-free in all normal tissues, while it was methylated in all cell lines and it was only found heavily methylated in one tumor sample and partially methylated in two other tumors. On the contrary, CpG78 contained low levels of DNA methylation in all normal tissues, turning to medium methylation in one tumor sample and heavy methylation in the other four tumors and all cell lines. The region from the DRD1 gene to the CpG78 (the two most distal ends of the methylated set of genes around CPLX2) spanned so far 617.6 Kb.

More genes were included in the analysis downstream of the methylated CpG78 (see figure 2A). A second DNA-methylated region was found affecting CpG91, CpG114 and CpG79. CpG91 is a CpG island associated to the 3' ends of the Protocadherin Lung, Kidney and Colon gene (PcLKC) and the G protein-regulated inducer of neurite outgrowth (GPRIN1). It was found to be non-methylated in all normal-tumor pairs except for one tumor tissue were low methylation levels were detected. Four of the 7 cell lines displayed heavy methylation in the same island. PcLKC

has no CpG island associated to its proximal promoter region (Figure 2A). CpG114 is associated to the GPRIN1 promoter and was found partially methylated in 4 of 5 normal tissues, showing increased methylation levels in 3 tumors. Cell lines exhibited again higher methylation levels, with 5 cell lines showing heavy methylation, one showing medium methylation and only one cell line showed no methylation. Finally, CpG79 associated to the β-synuclein gene was methylation free in 4 of 5 normal tissues (one normal sample exhibited low methylation levels) while 3 tumors displayed medium methylation and 2 tumors showed heavy methylation. Three of the cell lines displayed heavy methylation, two displayed medium methylation levels and two cell lines were methylation-free. The rest of the CpG islands analyzed showed to be methylation free in all normal-tumor pairs and all cell lines. These include CpG47, CpG100 associated to the Clathrin B (CLTB) gene, CpG126 associated to the Ring Finger 44 (RNF44) gene and CpG47 associated to the Tetraspanin 17 (TSPAN17) gene. Summarizing all the above mentioned, we have discovered a novel set of DNA-methylated, tumor-associated neighboring genes, that lay in two different subregions, for which the most distal ends, DRD1 and SNCB, map 1203.8 Kb away.

Methylation of multiple genes in 5q35.2 is a common event in colorectal carcinogenesis

We have determined the methylation status of the most heavily methylated genes HRH2, CPLX2, AK124.837 and SNCB genes over a series of 121 colorectal normal-tumor pairs, which included both adenomas and carcinomas. Real-time PCR melting curve analysis was the method used here for a high throughput analysis (see materials and methods) (Figure 3A). In those cases were methylation status was not clear, direct bisulfite sequencing was performed. In concordance with direct bisulfite sequencing data, the four analyzed genes were found extensively methylated in all tumor samples, including adenomas, being HRH2 the most frequently methylated (72.7%), followed by CPLX2 (70.8%), SNCB (68.6%) and AK124.837 (63%)(Figure 3C).

Methylation of any of the four genes affected virtually all tumors (119 out of 121) (Figures 3B and 3C). Among carcinomas, few cases exhibited methylation of only 1 gene (8.1%), while this figure was 26.3% for two methylated genes, 36.3% for three methylated genes and 28.1% for four methylated genes (Figure 3B). Adenomas showed a slightly different distribution, with most samples displaying 1 (25%) or 2 (27%) methylated genes and a lower percentage (12.5%) showing methylation of three or four genes (Figure 3B). Even though the genes described in this region tend to be methylated concurrently, a methylation pattern could not be discerned, therefore indicating that there is not a unique order in the methylation of the genes in this region (Figure 3C).

Transcriptional silence across genes in 5q35.2

Expression profiles of a total of 12 genes (including those genes with a methylated CpG island and a subset of those displaying no-methylation) were investigated for a subset of 16 normal tumor pairs. As illustrated in Figure 4A, DNA-methylated genes are almost invariably downregulated, with the exception of CPLX2, which is methylated and moderately upregulated in tumor sample 10655, HRH2 and SNCB which are methylated and upregulated in tumor

sample 10659 and HRH2, which is methylated and upregulated in tumor sample 10886 (Figure 4A). Interestingly, transcriptional silencing also affects genes that are non-methylated. Among these genes, there are those that have been found methylated in tumors and cell lines (DRD1, HRH2, CPLX2, AK124.837, GPRIN1 and SNCB), genes in which DNA methylation has never been found (SFXN1, CLTB, RNF44 and TSPAN17) and genes without CpG island (PcLKC). An extreme example of this situation is tumor sample 10845, which in spite of displaying no methylation at any promoter CpG island, shows downregulation of 10 of the 12 genes analyzed (Figure 4A).

Mean expression ratios (Figure 4B) and mean absolute expression values (Figure 4D) for the 16 normal-tumor pairs clearly indicate that most of the genes in 5q35.2 undergo transcriptional downregulation in tumors, with the exception of THOC3 and RNF44 genes, which tend to be moderately upregulated. Nevertheless, the different genes in the region do not become downregulated to the same extent, being AK124.837, PcLKC and SNCB the most downregulated ones. With some exceptions, colon cancer cell lines also exhibited methylation and silencing of the same genes affected in tumors. SFXN1, which is upregulated in cell lines and downregulated in primary tumors, was the main difference (Figure 4C-E). A perfect association between promoter DNA methylation and transcriptional silencing was observed in HCT116 cells (Figure 4C) and the rest of cell lines (data not shown). SW480 cell line displays a characteristic profile of the region. Specifically, CPLX2 and AK124.837 escape transcriptional silencing in accordance with the unmethylated status of their CpG islands, higher expression levels for unmethylated genes such as THOC3, RNF44 and SNCB and very high levels for the unmethylated genes CLTB, GPRIN1 and TSPAN17 (Figure 4E).

As a whole, expression profiles of this region in normal and tumor cells suggests the existence of interspersed regions containing highly transcribed genes (SFXN1, genes from THOC3 to RNF44 and TSPAN17), with other regions containing low-expression genes (DRD1, genes from HRH2 to AK124.837 and genes from PcLKC to SNCB).

Chromatin profiles identify different chromatin domains from DRD1 to TSPAN17

In order to gain insights into the mechanisms that lead to transcriptional silencing in tumors we profiled active and inactive histone marks along the region from DRD1 to TSPAN17 (Figure 5). Chromatin profiles were performed using the chromosomically stable HCT116 cell line. The close similarity between the expression and methylation profiles displayed by HCT116 cell line and primary tumors makes this cell line a good model into which perform this kind of experiments.

A panel of 17 primer sets were used in the analysis (Supplementary table 4), covering promoter regions (all genes) and CpG islands and intronic regions of the three most DNA methylated and downregulated genes (HRH2, CPLX2 and SNCB). As summarized in figure 5B, the "active" mark acetylated lysine 9 on histone H3 (Ac H3K9) is highly enriched over the promoter regions of the most highly transcribed genes (Figure 5A), defining three peaks of acetylation in SFXN1, RNF44 and TSPAN17. Hypoacetylation pockets included the DRD1 gene, the region spanning from HRH2 to AK124.837, and the region including PcLKC and SNCB. Noteworthy, all these genes appeared hypermethylated in cancer cells. Interestingly, acetylation of the Lysine 14 in histone

H3 (Ac H3K14) was already loaded in all regions tested, not showing any correlation with the active or silent genes (Supplementary figure 1). Nevertheless, the different genes showed slight differences in the content of this mark, with the highest amounts detected in the HRH2 promoter, CPLX2 intronic region, PcLKC promoter and SNCB promoter and intronic region, suggesting a weak correlation between this mark and inactive genes. Acetylated lysine 16 in histone H4 (Ac H4K16) was globally depleted in the region, except for the highly expressed RNF44 and the CpG island-free PcLKC gene. Very low levels of acetylated lysine 12 of the histone H3 was detected (see supplementary figure 1). The last "active" mark profiled was the trimethylated form of the lysine 4 in histone H3 (H3K4me3). This mark was found to be present in moderate amounts in all regions tested, with the highest enrichment over the actively transcribed, methylation-free SFXN1 and RNF44 genes and in the promoter CpG islands of the DNA methylated genes HRH2 and CPLX2 (Figure 5B).

With regards to histone modifications associated to inactive states, we surveyed for the dimethylated and trimethylated forms of the lysine 9 in the histone H3 (H3K9me2 and H3K9me3, respectively) and the trimethylated lysine 27 in the histone H3 (H3K27me3). Surprisingly, while no significant levels of H3K9me2/me3 could be found at any of the regions tested, including intronic regions of silenced genes HRH2, CPLX2 and SNCB (see supplementary figure 1) high levels of H3K27me3 were found mainly over the CpG islands of the DNA methylated genes DRD1, HRH2, CPLX2, AK124.837 and the DNA unmethylated but downregulated SNCB gene (Figure 5B). These are the same genes that showed depletion for Ac H3K9 and Ac H4K16 and moderate levels of Ac H3K14 and H3K4me3. Conversely, low to undetectable levels of the H3K27me3 mark were found over the promoters of the DNA unmethylated, active genes SFXN1, THOC3, RNF44 and TSPAN17. Even though CLTB and PcLKC were found downregulated in the HCT116 cell line, they showed very low levels of H3K27me3 over their promoter regions.

Demethylating agents derepress silent genes and affect the expression of active genes across 5q35.2

Re-expression patterns were assessed using HCT116 cell line for genes spanning from DRD1 to TSPAN17 using a pharmacological approach with the demethylating agent 5azaC and the class I and II HDAC inhibitor TSA, both alone and in combination. Genes could be grouped into four different groups according to their behavior after drug treatment (Figure 6). The first group of genes, corresponding to the two DNA-methylated genes HRH2, CPLX2, the unmethylated SNCB and the CpG island-free PcLKC were upregulated to different levels upon 5azaC treatment alone. Only the CPLX2 gene showed a synergistic re-expression after 5azaC and TSA co-treatments, while the rest of the genes in this group showed weak (HRH2 and SNCB) or no (PcLKC) upregulation of the co-treatment over the 5azaC treatment alone. A second group, including the DNA-methylated genes DRD1 and AK124.837, were only moderately upregulated after drug co-treatment, remaining unaffected after 5azaC treatment alone. The third group of genes, CLTB and TSPAN17 (both DNA-methylation free) were unaffected by drug treatment, either 5azaC and TSA alone or in combination. Finally, a fourth group included the highly-expressed, DNA metylation-free genes SFXN1, THOC3 and RNF44. Contrary to the rest of the genes in the region, these genes became moderately downregulated after drug treatment and

co-treatment. 5azaC and 5azaC/TSA co-treatments dropped SFXN1 expression levels to 57% and 63% respectively, while these figures were 71% and 46% for THOC3 and 64% and 17% for RNF44 (Figure 6). TSA treatment alone was unable to upregulate genes with methylated promoters, while downregulated genes displaying unmethylated promoters were weakly upregulated upon TSA treatment alone (see supplementary figure 3).

Chromatin modifications of drug-induced, re-expressed genes.

We have described how drug treatment differentially affects gene expression across 5q35.2, depending on gene basal expression level and DNA methylation status. Thus, we wondered how the histone modification landscape was affected by the drug treatment. Chromatin state in 17 regions was characterized in the untreated HCT116 cell line and compared with cells treated with 5azaC and 5azaC/TSA (Figure 7). A two to four fold enrichment for the Ac H3K9 mark was seen in all genes tested after 5aza/TSA co-treatment, being the promoters and CpG island regions of silenced genes (DRD1, HRH2, CPLX2 and SNCB) the areas showing the highest enrichment, plus the THOC3 gene, which also displayed a moderate enrichment for this modification. Interestingly, the enrichment was more clearly seen over CpG islands than over the promoter and intronic regions of DNA-methylated genes, as is the case of HRH2, CPLX2 and SNCB (Figure 7). On the other hand, only DNA-methylated genes displayed a low enrichment for this active mark after 5azaC treatment alone (promoter region and CpG island of HRH2 and SNCB genes and CpG island of CPLX2 gene), while the rest of the genes displayed a mild decrease for the same modification (Figure 7).

An unexpected result was obtained regarding Ac H4K16 modification. Treatment with 5azaC alone induced a low to moderate erasure of this mark at virtually all regions analyzed. On the contrary, drug co-treatment had the opposite effect, highly enriching this modification over HRH2, CPLX2 and SNCB, having a moderate enrichment over DRD1 and CLTB and a low effect over PcLKC and TSPAN17 (Figure 7). Highly expressing genes showed no enrichment for this modification. This is of special relevance given the fact that the HDAC is supposed to deacetylate this aminoacid residue, the NAD+ dependent class III HDAC SirT1, has been previously involved in the silencing of other DNA methylated tumor suppressor genes[34]. Interestingly, the present data suggests an active role of SirT1 in the maintenance of the repressed state, independently of promoter DNA methylation.

Only silenced genes, DNA methylated or not, were enriched for H3K4me3 after both 5azaC and combined drug treatment, while highly expressing genes had their H3K4me3 levels depleted after the treatments (Figure 7). Globally, H3K4me3 is the only mark that showed a clear enrichment over the promoters of silenced genes after both 5azaC treatment alone and co-treatment. The highest enrichments were detected once again around the promoter regions of the DNA-methylated genes DRD1, HRH2, CPLX2, AK124.837 and SNCB (Figure 7). Moderate enrichments were seen over CLTB and PcLKC. On the other hand, H3K4me3 levels were reduced after 5azaC treatment alone at the SFXN1 promoter, while depletion for this mark was seen only after drug co-treatment at the THOC3 and TSPAN17 promoters. Only RNF44 showed reduced levels for this mark after both 5azaC treatment and co-treatment with TSA.

Finally, and most surprisingly, while H3K27me3 levels were weakly affected by 5azaC treatment alone, co-treatment with TSA clearly induced a global gain of this histone modification across the entire region (Figure 7). Enrichment for this mark affected all genes, including active genes, which had been previously characterized to have very low or undetectable levels of this mark. These results clearly indicate that this classic inactive histone modification not only remains after 5azaC treatment but increases with transcriptional upregulation upon drug co-treatment.

Components of polycomb repressor complexes are found at the inactive genes.

The expression profiles and histone modification patterns before and after drug treatment clearly suggest that a specific histone code exists around the promoters of silenced genes, especially over those with promoter DNA methylation. The specific presence of H3K27me3 indicates that components of the polycomb group of proteins may be mediating, at least in part, the silencing that affects most of the genes across this region. We therefore attempted to identify the polycomb proteins EZH2 and BmiI, over the promoter regions of the silenced genes. Additionally, specific enrichment of Ac H4K16 over the promoters of silenced genes predicted the presence of the HDAC SirT1, which was also included in the analysis. As expected, the histone methyltransferase responsible for the deposition of the methylation mark over lysine 27, EZH2, was found enriched over the promoters of the silenced genes where we had previously detected enrichment for its methylation mark, H3K27me3 (Figure 8). This was consistent with the presence, over the same regions of the polycomb member BmiI, which forms part of the polycomb repressor complex 1 (PRC1) that mediates recognition of the H3K27me3 mark, and the HDAC SirT1.

Drug treatment resulted in an increase of the EZH2 protein over the promoter regions of all genes, which provides a plausible explanation for the global increase of the H3K27me3 mark upon reactivation. BmiI was consistently depleted in all genes after reactivation. Finally, while 5azaC treatment alone increased the levels of SirT1 at some promoters (HRH2, CPLX2, AK124.837 and CLTB), drug co-treatment tended to reduce the levels of this HDAC at all promoters tested (Figure 8).

Discussion

Concurrent methylation of multiple CpG islands across 5q35.2 is a common event in colorectal carcinogenesis

Aberrant *de novo* DNA methylation is the most well characterized epigenetic alteration in carcinogenesis and is associated with the transcriptional silencing of a large list of genes within multiple pathways relevant to tumor biology[5,14]. Even though multiple hypermethylation events can be found affecting a single tumor cell, hypermethylation has been always considered to be restricted to individual, discrete CpG island-containing genes. This paradigm has changed since the discovery that promoter DNA methylation and chromatin remodeling can also encompass several megabases, as has been recently described in colorectal[29,30] and bladder cancers[31]. By applying a DNA methylation screening methodology that specifically targets unmethylated

repeats, we have unmasked the presence of an unmethylated MIR repetitive element besides the Complexin 2 (CPLX2) CpG island, both of which undergo extensive DNA methylation in tumors and cell lines (Figure 1C). Further characterization of the genomic region around the CPLX2 gene, has unveiled a novel set of genes undergoing promoter DNA methylation (Figures 2B and 3C). Even though the region affected does not encompass a complete chromosomal band, as reported by Frigola and collaborators for 2q14.2, the number of genes affected by the presence of promoter DNA methylation is significant and further experiments directed to ascertain the possible roles of these genes in carcinogenesis are warranted. These findings, together with unpublished results from our laboratory and previous reports[29,30] suggest that promoter DNA methylation affecting multiple neighboring genes is not an uncommon event in colorectal carcinogenesis and surely deserves further investigation.

Results clearly indicate that hypermethylation of HRH2, CPLX2, AK124.837 and SNCB is a highly frequent event in colorectal carcinogenesis (Figure 3C), affecting well over 60% of the tumors tested. Moreover, hypermethylation was also detected in a significant fraction of the adenomas, thus suggesting an early onset of the methylation in this region during tumorigenesis. These results are extremely relevant, as early detection of tumors is one of the most critical parameters affecting patient survival (reviewed in[35]). Our results show that only a very small fraction of all the tumors analyzed do not contain any methylated gene (less than 2% of the samples). Future experiments will elucidate the usefulness of this region as a diagnostic and/or prognostic marker, and will assess the feasibility of its detection in different samples (i.e., the analysis of stools or circulating DNA).

Additionally, even though the highest methylation levels have been found restricted to tumors, low to moderate DNA methylation levels have also been detected in normal tissue from different patients (Figure 2B). Normally methylated genes include the alternative transcript AK124.837 (CpG125), BC042064 (CpG78) and GPRIN1 (CpG114). Low level DNA methylation was found over CpG79, associated to SNCB, in only one normal sample. Interestingly, BC042064 gene is located within a segmental duplication that has evolved from a sequence located further downstream chromosome 5 in 5q36.1. Methylation of this region in the normal tissue suggests that it may have some specific feature that renders it prone to become methylated, hypothesis that surely deserves further investigations.

DNA-methylation and histone signatures define active and inactive chromatin domains in 5q35.2

The genomic region presented here represents a unique model into which further investigate the complex regulatory mechanisms that lead to transcriptional inactivation associated to DNA methylation. As we have mentioned above, the presence of highly expressing genes among the DNA-methylated, silenced genes, provides a collection of positive and negative controls that can be directly compared, therefore allowing the identification of common features that define the different transcriptional states. In the above scenario, we provide evidence for the presence of isolated expression domains characterized by specific patterns of histone modifications and DNA methylation. The active compartments, where SFXN1, THOC3, RNF44 and TSPAN17 are embedded, are characterized by many active-associated chromatin features: high transcription

rates are associated with unmethylated promoter CpG islands (Figures 5A and 5C), and moderate to high levels in the promoter regions of the active histone modifications Ac H3K9, Ac H4K16 and H3K4me3, and low or inexistent levels of the inactive histone modification H3K27me3 (Figure 5B). On the other hand, the inactive compartment is defined by low transcriptional activity associated with methylated promoter CpG islands (Figure 5A and 5B), low levels of the active histone modifications Ac H3K9, Ac H4K16 and H3K4me3, and moderate to high levels of the inactive H3K27me3 histone modification mainly over the DNA methylated CpG island (Figure 5B).

The fact that transcriptionally active and inactive regions are arrayed one after another (Figure 5A) lead us to think that some insulator mechanism must operate in the region. Only one factor with insulator activity has been identified in mammals so far, the CTCF factor, which is essential in the formation of differentially methylated imprinted domains (Reviewed in[36]). The use of global approaches to map CTCF targets genome-wide in the human genome have identified many non-imprinting related locations at which CTCF may regulate distinct epigenetic states, some of them across the region described here[37]. Some of these CTCF sites in 5q35.2 are found at intergenic regions between genes with opposite transcriptional and chromatin states (Supplementary Table 5), in agreement with a CTCF role in marking boundaries of different histone methylation domains[37]. Future experiments should uncover a possible role for this insulator protein in the maintenance of the transcriptional silencing in this region.

As shown in figure 4B, there is a global trend towards downregulation affecting most of the genes in the region, with the exception of THOC3 and RNF44, which appear upregulated in the tumors. The study of a set of differentially expressed genes indicates that only low expressing or silent genes in the normal tissue become DNA methylated in both tumors and cell lines, in agreement with previous reports on DNA methylated genes in cancer[28]. These results, together with the fact that transcriptional downregulation affects many unmethylated genes in the region, are consistent with the hypothesis that DNA methylation is not the primary event leading to transcriptional silencing in carcinogenesis, but rather locks in an already existing silent state. Nevertheless, DNA methylation plays a dominant role in the silencing mechanism, as the silent genes cannot be reactivated unless first demethylated. Thus, the use of drugs that inhibit HDAC activity such the one used here has no effects on the transcriptional activity of DNA-methylated genes but rather synergize with 5azaC once the DNA has been demethylated (Figure 6 and Supplementary Figures 2 and 3). Moreover, the inhibitory effects of DNA methylation extend over non-CpG island containing genes such as PcLKC and unmethylated genes such as SNCB, which can be reactivated after 5azaC treatment alone, as has been previously shown for other genes[29].

The fact that some methylated genes appear upregulated in a limited number of tumors could be explained due to contamination of the tumor sample with normal cells. Alternatively, we could also hypothesize that only a small fraction of the cells in the tumor present a hypermethylated promoter, thus masking the effects of methylation on the expression levels. Both situations could be responsible for the results obtained.

A stem cell-like chromatin landscape defines silencing across 5q35.2

To date, the origins of human cancer remain obscure. A very exciting and promising possibility comes from recent studies of the epigenetic landscape in murine embryonic stem (ES) cells, where a striking balance of the two opposite histone modifications, the "active mark" H3K4me3 and the "inactive mark" H3K27me3, exist over the promoter regions of key developmentally-regulated genes, in what the authors call bivalent domains[24]. The presence of these two marks over the subset of genes reported by Bernstein and colleagues correlates with low levels of transcription of the same genes, which suggests that bivalent domains keep key developmentally-regulated genes in a silent state, but ready for activation, which reflects the pluripotency of these cells at the chromatin level[24].

A possible origin of cancer from stem or early progenitor cells has been previously postulated[38-41]. Supporting evidence to this hypothesis has come from the discovery that promoter regions of genes frequently hypermethylated in colorectal cancer tend to be occupied by polycomb proteins in human embryonic stem cells, even though the authors do not provide evidence for polycomb protein occupancy in cancer cells[22]. Furthermore, the discovery that bivalent domains embody the promoter regions of a subset of silenced genes in embryonic carcinoma (EC) cells, the malignant counterparts of ES cells[23], indeed suggests that cancer cells have a distinctive stem cell-like chromatin pattern. In this context, DNA methylation has been suggested to be the factor locking in these chromatin patterns in adult cancers, which in contrast to ES and EC cells, display high levels of promoter DNA methylation[23].

Consistent with the above mentioned, in the HCT116 cell line both H3K4me3 and H3K27me3 have been found over the promoters of all DNA-methylated genes and the DNA unmethylated, but severely downregulated SNCB gene, across 5q35.2 (Figure 5B). Noteworthy, SNCB is not methylated in the HCT116 cell line, but it is heavily methylated in multiple cell lines and tumor tissues (Figures 2B, 3C and 4A). The bivalent domains reported here embody areas of low expression, discontinued by active domains enriched for the active mark H3K4me3 and depleted for the H3K27me3 inactive mark. Unlike embryonic EC, where bivalent domains keep key genes ready for activation with DNA-methylation free promoters[23], HCT116 cancer cells have dense promoter DNA methylation, which prevents any further transcription activation. Concurrent methylation of multiple neighboring CpG islands has been previously reported in colorectal cancer[29,30], and genome-wide maps of DNA methylation in the colorectal cancer cell line CaCo2 showed that DNA methylated genes tend to cluster together[28]. Our results not only provide another example of concurrent DNA methylation of multiple neighboring CpG islands in cancer, but also indicate that concurrent DNA methylation occurs at genomic regions where genes under the control of bivalent domains cluster together.

Ooi and colleagues have recently reported on a previously unnoticed association between DNA methylation and histone H3 tail[42]. Their results indicate that unmethylated Histone H3 at lysine 4 can lead to DNA methylation mediated by the DNA methyltransferase DNMT3L, therefore linking the absence of a key active histone modification to DNA methylation. Nevertheless, the DNA-methylated genes reported here contain moderate levels of methylated H3K4 and

therefore the results obtained by Ooi et al. do not provide a convincing explanation as to why these genes become DNA-methylated.

According to the presence over the promoters of silenced gens of H3K27me3, further chromatin immunoprecipitation experiments have revealed the presence of the PRC2 member histone methyltransferase EZH2, responsible for the deposition of this histone modification (Reviewed in[26]) and the member of the PRC1 BmiI. Interestingly, together with EZH2 and BmiI we have also detected the presence of the histone deacetylase SirT1, which is in agreement with the specific enrichment of acetylated H4K16, the lysine residue deacetylated by SirT1, seen over the promoters of silenced genes upon drug treatment (Figure 7). This class III HDAC has been shown to be directly involved in the repression of DNA-methylated genes in different cancer models[34] and is associated to transcriptional repression mediated by polycomb in *Drosophila*[43]. Interestingly, experiments conducted by Kuzmichev and colleagues[44] have allowed the identification of a previously undescribed polycomb complex, named PRC4, which in addition to the basal components EZH2 and SUZ12 contains the SirT1 HDAC and the Eed isoform Eed2. As noted by Kuzmichev et al., the fact that PRC4 shows preference for methylating H1K26, a target for deacetylation by SirT1 in vivo, implies that PRC4 can also methylate H3K27 or alternatively, that another polycomb repressor complex methylates this position[44]. The fact that Eed2 is a specific isoform expressed exclusively in cancer cells and undifferentiated embryonic stem (ES) cells, further strengths the hypothesis of a cancer-stem cell origin of cancer. Based upon this observations, we conclude that the PRC4 complex is associated with the deposition of H3K27 mark and the specific depletion in the acetylated form of lysine 16 in histone H4 over the silenced genes, which together with a specific combination of other histone marks defines transcriptional silencing across 5q35.2

Bivalent domains persist after drug-induced gene reactivation

Comparison of DNA methylation between the normal tissue and the tumor tissue and cell lines has revealed the existence of tumor-specific DNA methylation (Figure 2B and 4C), which chromatin immunoprecipitation experiments over the HCT116 cell line have linked to the presence of bivalent domains and histone hypoacetylation (Figure 5B and 5C). Nevertheless, we can not confirm that the trimethylated mark on lysine 27 is cancer specific, mainly due to the absence of a cellular model that resembles the normal colorectal epithelium in which to perform such experiments. In this context, recent works focused on the genomic distribution of different histone methylation signatures in murine ESCs have revealed the presence of bivalent domains exclusively over the same genes described here[45], with some major differences in the case of human ESCs[46,47] (Supplementary Table 5). Similar experiments performed over human embryonic stem cells[48] failed to detect the polycomb proteins SUZ12 and Eed and the methylated H3K27 over the same promoter regions, probably indicating different onset times for the chromatin patterns over these genes in different model organisms and different cell types. Together, these results indicate that H3K27me3, together with low levels of H3K4me3, are already present at the genes that will become DNA-methylated during carcinogenesis.

As shown before for a subset of genes, bivalent domains in murine ESCs tend to become selectively enriched for any of the two opposite marks upon cell differentiation. Therefore, it

could be predicted that the bivalent domains described here would also become enriched for the active mark H3K4me3 and depleted for the inactive mark H3K27me3 upon drug-induced reactivation. Unexpectedly, both H3K4me3 and H3K27me3 were found enriched upon reactivation, therefore suggesting that cellular memory systems prevent these bivalent domains from being perturbed. In agreement with these results we have found that EZH2, the enzyme responsible for H3K27me3 deposition, tends to be enriched in most of the promoter regions analyzed after drug treatment in comparison to the untreated cell line (Figure 7), thus providing an explanation for the H3K27me3 increase. These results are consistent with previous reports on drug reactivation of DNA methylated genes, which also showed that both H3K4me3 and H3K27me3 become enriched after gene reactivation, although the authors show that EZH2 levels decrease at the same promoters[49].

Spreading of H3K27me3 to adjacent active domains upon drug treatment

We have shown how the EZH2 protein increases at promoters of most genes across 5q35.2, regardless of their basal level in the untreated cells (Figure 8), in agreement with the increase in the H3K27me3 levels at silenced promoters (Figure 7). But unexpectedly, H3K27me3 mark lost its specific dispersed distribution after drug co-treatment, showing a dramatic increase over the promoter regions of active genes (Figure 7). This enrichment provides a full explanation for the downregulation that this subset of active genes suffer after 5azaC and TSA co-treatments, together with a depletion of the H3K4me3 mark over the same promoters. Therefore, by using 5azaC and TSA we have broken the balance of H3K4me3 and H3K27me3 exclusively over the promoters of active genes, while silenced genes retain their bivalent domains, which further provides evidence for the existence of isolated chromatin and expression domains.

We provide further evidence for a dominant role of H3K4me3 and H3K27me3 over the rest of histone modifications tested here. In this context, even though we have seen a global increase in the lysine 9 acetylation levels at all sites tested across 5q35.2, breaking the balance between the polycomb and trithorax marks at the active promoters has been sufficient to provoke their transcriptional downregulation.

Summarizing, our data clearly support the hypothesis that genes marked with bivalent domains become preferentially DNA methylated during carcinogenesis, although at present we can't give a reasonable explanation for this phenomenon. The bivalent domains over the promoters of DNA methylated genes described here exist in undifferentiated ESC from both human and mouse, with some major differences, therefore leading to the conclusion that DNA methylation locks in silence at genes that are already kept in a low transcriptional state by the presence of these domains. Drug reactivation not only fails to erase bivalent domains, but also induces the spreading of the silencing mark, most probably by breaking the insulator mechanisms that keep neighboring genes in distinct and opposite transcriptional states.

References

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
2. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
3. Jones, P.A. & Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-28 (2002).
4. Herman, J.G. & Baylin, S.B. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* **349**, 2042-54 (2003).
5. Jones, P.A. & Baylin, S.B. The epigenomics of cancer. *Cell* **128**, 683-92 (2007).
6. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669-81 (2007).
7. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
8. Goll, M.G. & Bestor, T.H. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**, 481-514 (2005).
9. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-82 (1987).
10. Takai, D. & Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740-5 (2002).
11. Fazzari, M.J. & Greally, J.M. Epigenomics: beyond CpG islands. *Nat Rev Genet* **5**, 446-55 (2004).
12. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54 (2003).
13. Yoder, J.A., Walsh, C.P. & Bestor, T.H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**, 335-40 (1997).
14. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* **8**, 286-98 (2007).
15. Feinberg, A.P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143-53 (2004).
16. Ehrlich, M. DNA methylation in cancer: too much, but also too little. *Oncogene* **21**, 5400-13 (2002).
17. Plass, C. Cancer epigenomics. *Hum Mol Genet* **11**, 2479-88 (2002).
18. Lund, A.H. & van Lohuizen, M. Epigenetics and cancer. *Genes Dev* **18**, 2315-35 (2004).
19. Laird, P.W. Cancer epigenetics. *Hum Mol Genet* **14 Spec No 1**, R65-76 (2005).
20. Li, B., Carey, M. & Workman, J.L. The role of chromatin during transcription. *Cell* **128**, 707-19 (2007).
21. Schlesinger, Y. et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**, 232-6 (2007).
22. Widschwendter, M. et al. Epigenetic stem cell signature in cancer. *Nat Genet* **39**, 157-8 (2007).
23. Ohm, J.E. et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39**, 237-42 (2007).
24. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).

25. Vire, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-4 (2006).
26. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**, 413-43 (2004).
27. McGarvey, K.M., Greene, E., Fahrner, J.A., Jenuwein, T. & Baylin, S.B. DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. *Cancer Res* **67**, 5097-102 (2007).
28. Keshet, I. et al. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* **38**, 149-53 (2006).
29. Frigola, J. et al. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* **38**, 540-9 (2006).
30. Hitchins, M.P. et al. Epigenetic Inactivation of a Cluster of Genes Flanking MLH1 in Microsatellite-Unstable Colorectal Cancer. *Cancer Res* **67**, 9107-16 (2007).
31. Stransky, N. et al. Regional copy number-independent deregulation of transcription in cancer. *Nat Genet* **38**, 1386-96 (2006).
32. Rodriguez, J. et al. Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Submitted* (2007).
33. Raevskaya, N.M. et al. Structural organization of the human complexin 2 gene (CPLX2) and aspects of its functional activity. *Gene* **359**, 127-37 (2005).
34. Pruitt, K. et al. Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. *PLoS Genet* **2**, e40 (2006).
35. Etzioni, R. et al. The case for early detection. *Nat Rev Cancer* **3**, 243-52 (2003).
36. Delaval, K. & Feil, R. Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev* **14**, 188-95 (2004).
37. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37 (2007).
38. Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J. & Clarke, M.F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-8 (2003).
39. Clarke, M.F. & Fuller, M. Stem cells and cancer: two faces of eve. *Cell* **124**, 1111-5 (2006).
40. Pardal, R., Clarke, M.F. & Morrison, S.J. Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer* **3**, 895-902 (2003).
41. Jordan, C.T. Searching for leukemia stem cells--not yet the end of the road? *Cancer Cell* **10**, 253-4 (2006).
42. Ooi, S.K. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-7 (2007).
43. Furuyama, T., Banerjee, R., Breen, T.R. & Harte, P.J. SIR2 is required for polycomb silencing and is associated with an E(Z) histone methyltransferase complex. *Curr Biol* **14**, 1812-21 (2004).
44. Kuzmichev, A. et al. Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc Natl Acad Sci U S A* **102**, 1859-64 (2005).

45. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
46. Pan, G. et al. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell stem cell* **1**, 299-312 (2007).
47. Zhao, X.D. et al. Whole-genome mapping of histone H3 lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell stem cell* **1**, 286-298 (2007).
48. Lee, T.I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-13 (2006).
49. McGarvey, K.M. et al. Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. *Cancer Res* **66**, 3541-9 (2006).

Figure Legends

Figure 1. Panel A contains a fragment of a polyacrilamyde denaturing gel displaying the Ap1 band (arrowhead). The presence of the band in all normal tissues indicates the unmethylated state of the SmaI sites at both band ends in the normal colon. Loss of band intensity (marked by an asterisk) in tumor samples 53, 81 and 99 is indicative of a tumor-associated methylation of one or both of the SmaI target sites. Therefore these three normal-tumor pairs are used as Ap1 positive methylation controls for subsequent analysis. Tumor samples 72 and 74 do not display any band intensity loss, and are considered as negative methylation controls. Genetic structure of the region containing the amplified band is depicted in B. DNA is represented by solid lines and the amplified Ap1 fragment, the CPLX2 CpG island, and the MIR and MIRb repetitive elements are represented by black boxes. CpG dinucleotides are represented as short vertical lines across a horizontal line (DNA). Taking into account CpG distribution and Ap1 band location, two bisulfite sequencing fragments were designed (gray boxes), one of them located over the CPLX2 CpG island and a second fragment located over the MIR repetitive element (named CPLX2 MIR), covering each fragment one of the two SmaI sites. Methylation data from direct sequencing of bisulfite-treated DNAs (C) was obtained for the 5 normal-tumor pairs in A and for a collection of 8 colorectal cancer cell lines. Each CpG dinucleotide is represented by circles. Sequencing data clearly corroborates the unmethylated state of all normal tissues and the tumor-specific methylation of both CPLX2 CpG island and MIR element in methylation positive controls and cell lines, with the exception of CaCo2 and SW480 cell lines. In the case of SW480 cell line, the unmethylated status of the CPLX2 CpG island correlated with detectable transcription levels of CPLX2 (D).

Figure 2. The genomic region containing the CPLX2 locus is located near the 5q telomere at 5q35.2 (A). The region covered in this study spans over 1.2 Mb and includes 18 genes, from most centromere-proximal DRD1 to most telomere-proximal TSPAN17. CpG islands (CpGi) across 5q35.2 are named according to the number of CpG residues they contain, except for the CPLX2 MIR fragment, which is not a CpG island itself. Direct sequencing methylation data for seven of the eight colorectal cancer cell lines and the 5 normal-tumor pairs is summarized in B. Each box represents a CpG island, which can be found unmethylated (white), partially methylated (gray) or heavily methylated (black). ND, not determined.

Figure 3. Real Time melting curve analysis was used to determine methylation status of 4 of the *de novo* methylated genes in a series of 121 normal-tumor pairs that included both adenomas and carcinomas. Panel A contains melting curve analysis for a negative methylation control (Case 861) and a positive methylation control (case 659). Only positive control tumor 861 displays an increase in the melting temperature of the amplified DNA fragment with respect to the melting temperature of the normal tissue. No melting temperature increase is seen for negative control case 861. Even though methylation affected both adenomas and carcinomas, the latest exhibited, on average, a higher number of methylated genes (B). Panel C contains methylation data summary for the 4 analyzed genes (Y axis) in 118 of the 121 normal-tumor pairs (X axis; 3 carcinomas failed to amplify 1 or more genes and therefore are not shown). Each white box corresponds to an unmethylated status and a black box to a methylated status.

Carcinomas are clustered together (samples to the left) and represented in order of their methylation frequency. Adenomas (Ad) are clustered at the right side of the panel and are also represented in order of their methylation frequency. The percentage of tumors with each methylated gene is given.

Figure 4. Expression ratios and promoter DNA methylation status were determined for 12 genes across 5q35.2 in a series of 16 normal-tumor pairs and HCT116 cell line (A). Expression values are represented as the log2 of the tumor-normal ratio. Therefore, genes that are upregulated in the tumor have values above 0 and genes downregulated in the tumor have values below 0. Those expression bars colored in red correspond to a methylated promoter and those colored in green correspond to unmethylated promoters. PcLKC is the only gene in this study that doesn't have a promoter CpG island, and therefore its methylation status is referred here as unmethylated (green). Mean values of the expression ratios of all 16 normal-tumor pairs are plotted in B. Expression ratios and methylation data for the HCT116 cell line are given in C. In this case, values for each gene were obtained as the ratio of the HCT116 cell line expression and the mean expression values of the 16 normal tissues. Panel D represents mean absolute expression values across the same 12 genes across 5q35.2 for the 16 normal tissues (green), the 16 tumor tissues (red) and HCT116 cell line (black). The absolute expression values of the HCT116 cell line were compared to the absolute expression levels of a panel of other colorectal cancer cell lines used in this study (E). ND, not determined.

Figure 5. Absolute expression data for genes across 5q35.2 (A) were compared to the levels of different histone modifications (B) in the cell line HCT116. IgG values represent the negative control. The presence of the three activation marks Ac H3K9, Ac H4K16 and H3K4me3 and the inactive mark H3K27me3 was ascertained. Regions analyzed included promoter regions of all genes, and for three additional genes (HRH2, CPLX2 and SNCB) the CpG island (C) and a distal intronic region (I) were analyzed besides the promoter region (P). In those cases where the amplified region falls inside or nearby a CpG island, the CpG island methylation status is noted as methylated (gray boxes) or unmethylated (white boxes). Chromatin immunoprecipitation values are shown as the enrichment fraction over input. ND, not determined.

Figure 6. Absolute expression data for genes across 5q35.2 in the cell line HCT116 after drug treatment. White bars correspond to the untreated control cell line, while light gray corresponds to the 5azaC treatment and dark gray corresponds to the double 5azaC and TSA treatments.

Figure 7. Chromatin remodeling across 5q35.2 after drug treatment in the HCT116 cell line. The same histone modifications previously characterized are shown in relative levels to the untreated control cell line. Light gray bars correspond to the 5azaC treatment and black bars correspond to the double 5azaC and TSA co-treatment. Values above 0 correspond to a modification enrichment while values below 0 correspond to modification depletions in the treated cells. Dashed boxes group values for different primer sets in the same gene, corresponding to the promoter (P), CpG island (C) and intron (I) regions. ND, not determined.

Figure 8. Mapping of polycomb components EZH2 and BmiI and the SirT1 HDAC across gene promoters in 5q35.2. Absolute enrichment levels over the input are shown for the untreated control HCT116 cell line (white bars), 5azaC treated (light gray) and 5azaC and TSA co-treated cells (dark gray). Chromatin immunoprecipitation values are shown as the enrichment fraction over input.

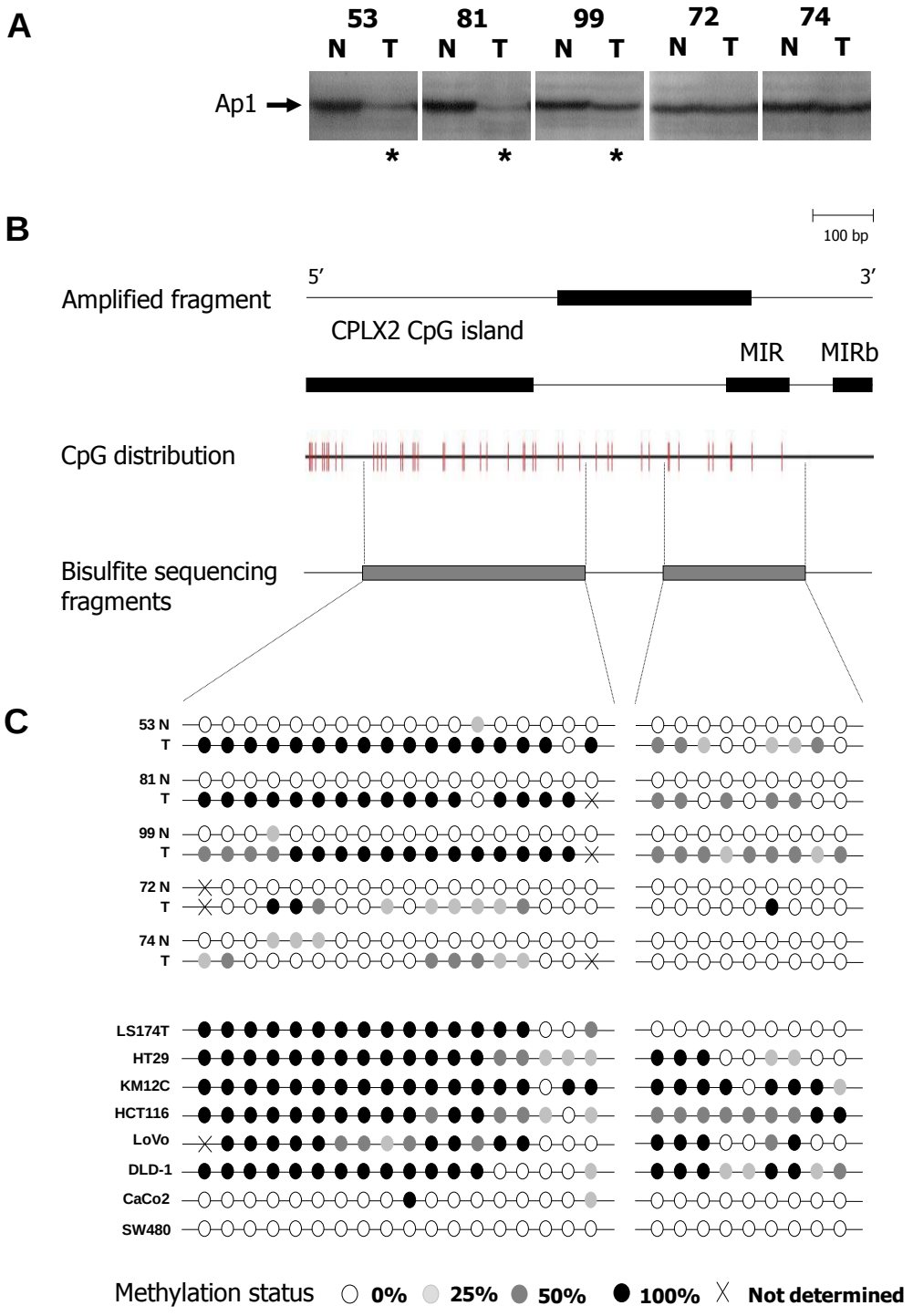

A
53
81
99
72
74
N T N T N T N T N T
Ap1
*
*
*
B
100 bp
Amplified fragment
5′
3′
CPLX2 CpG island
MIR
MIRb
CpG distribution
Bisulfite sequencing fragments
C
53 N
T
81 N
T
99 N
T
72 N
T
74 N
T
LS174T
HT29
KM12C
HCT116
LoVo
DLD-1
CaCo2
SW480
Methylation status 0% 25% 50% 100% Not determined
D
SW480
HCT116
LoVo
KM12C
CaCo2
DLD-1
200 pb
CPLX2
500 pb
18S rRNA

FIGURE 1

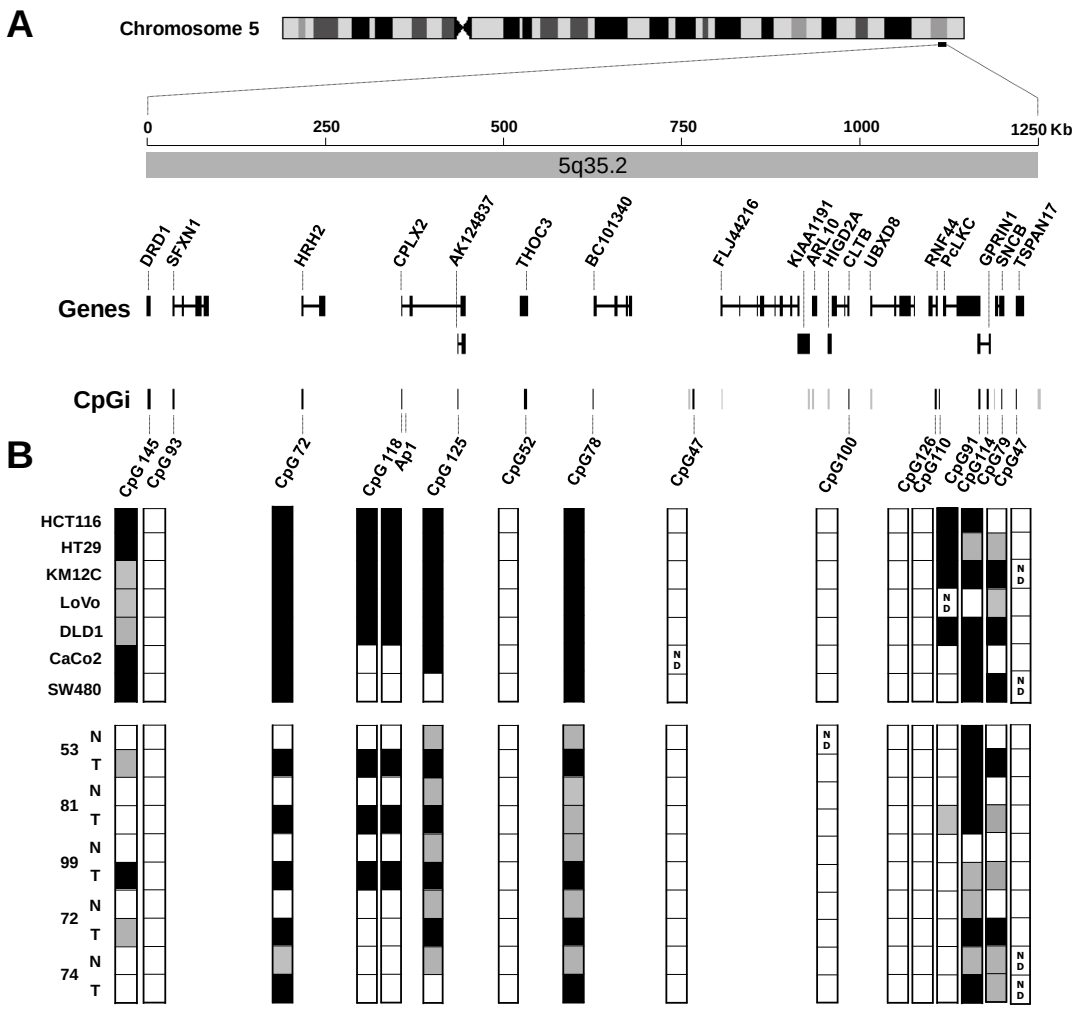

A
Chromosome 5
0
250
500
750
1000
1250 Kb
5q35.2
DRD1
SFXN1
HRH2
CPLX2
AK124837
THOC3
BC101340
FLJ44216
KIAA1191
ARL10
HIGD2A
CLTB
UBXD8
RNF44
PcLKC
GPRIN1
SNCB
TSPAN17
Genes
CpGi
B
CpG145
CpG93
CpG72
CpG118
Ap1
CpG125
CpG52
CpG78
CpG47
CpG100
CpG126
CpG110
CpG91
CpG114
CpG79
CpG47
HCT116
HT29
KM12C
LoVo
DLD1
CaCo2
SW480
53
81
99
72
74
N
T
ND

FIGURE 2

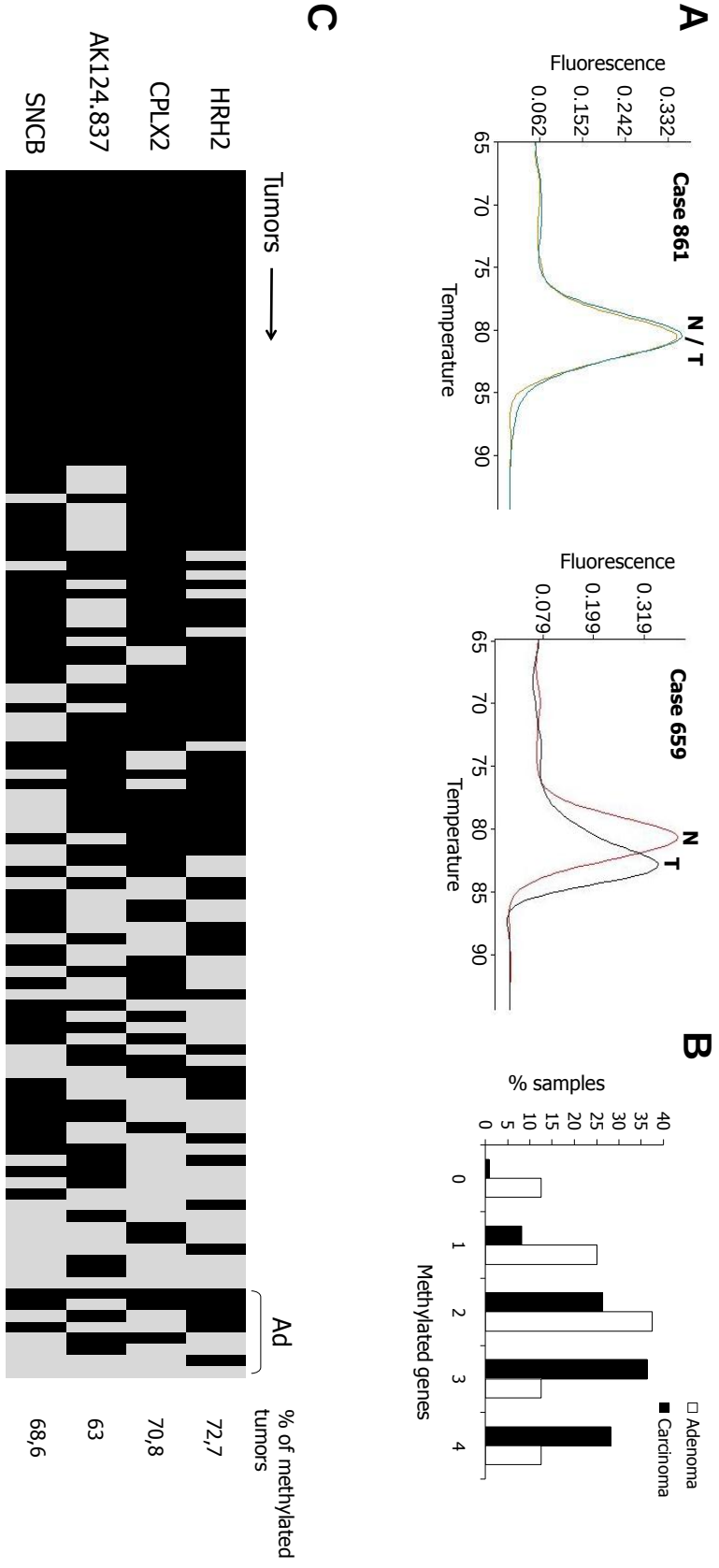

A
Fluorescence
0.332
0.242
0.152
0.062
Case 861
N / T
65 70 75 80 85 90
Temperature
Fluorescence
0.319
0.199
0.079
Case 659
N
T
65 70 75 80 85 90
Temperature
B
% samples
40 35 30 25 20 15 10 5 0
0 1 2 3 4
Methylated genes
Adenoma
Carcinoma
C
Tumors
HRH2
CPLX2
AK124.837
SNCB
Ad
% of methylated tumors
72,7
70,8
63
68,6

FIGURE 3

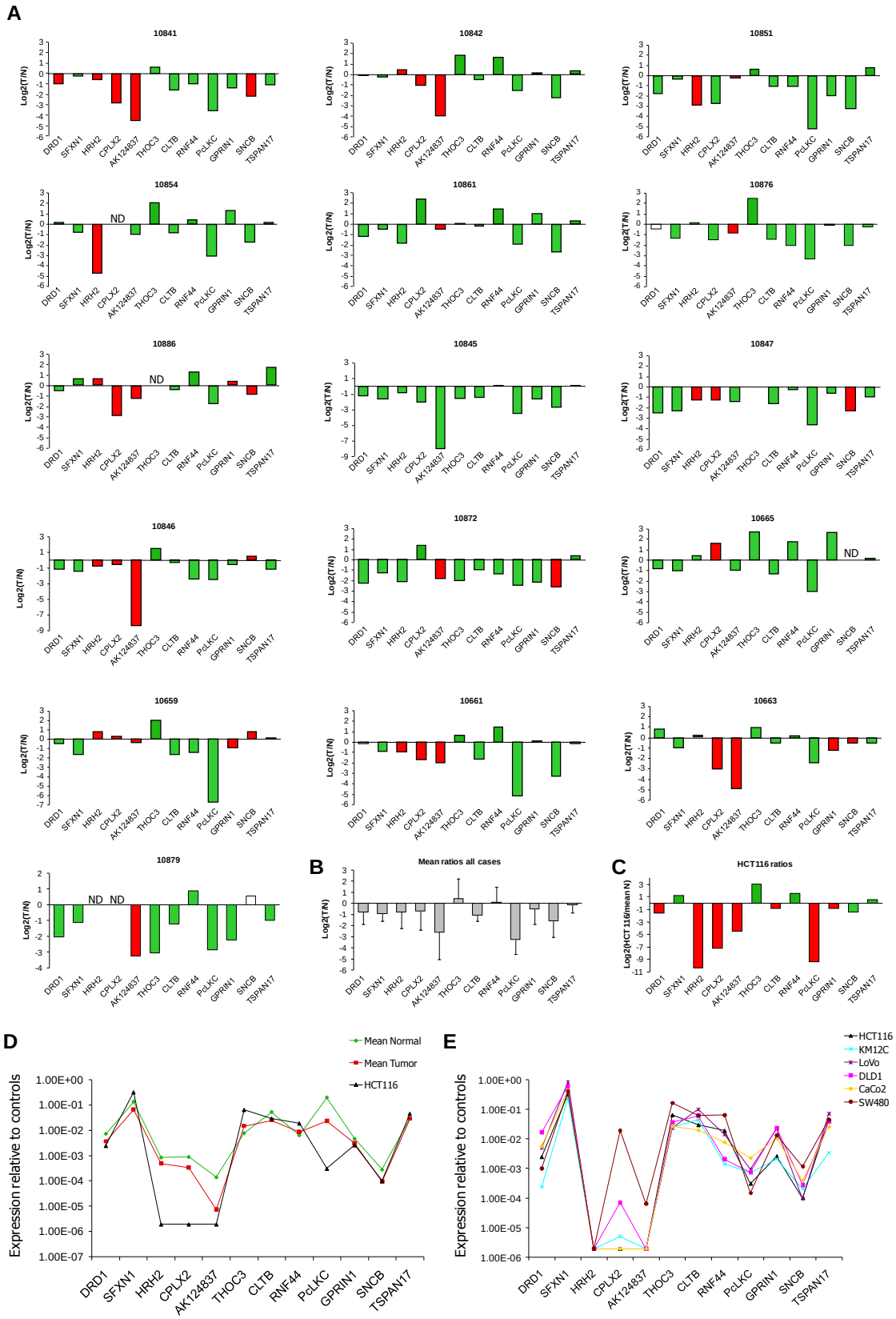

A
10841
10842
10851
10854
ND
10861
10876
10886
ND
10845
10847
10846
10872
10665
ND
10659
10661
10663
10879
ND ND
B
Mean ratios all cases
C
HCT116 ratios
Log2(T/N)
Log2(HCT 116/mean N)
DRD1 SFXN1 HRH2 CPLX2 AK124837 THOC3 CLTB RNF44 PcLKC GPRIN1 SNCB TSPAN17
D
Mean Normal
Mean Tumor
HCT116
Expression relative to controls
E
HCT116
KM12C
LoVo
DLD1
CaCo2
SW480

FIGURE 4

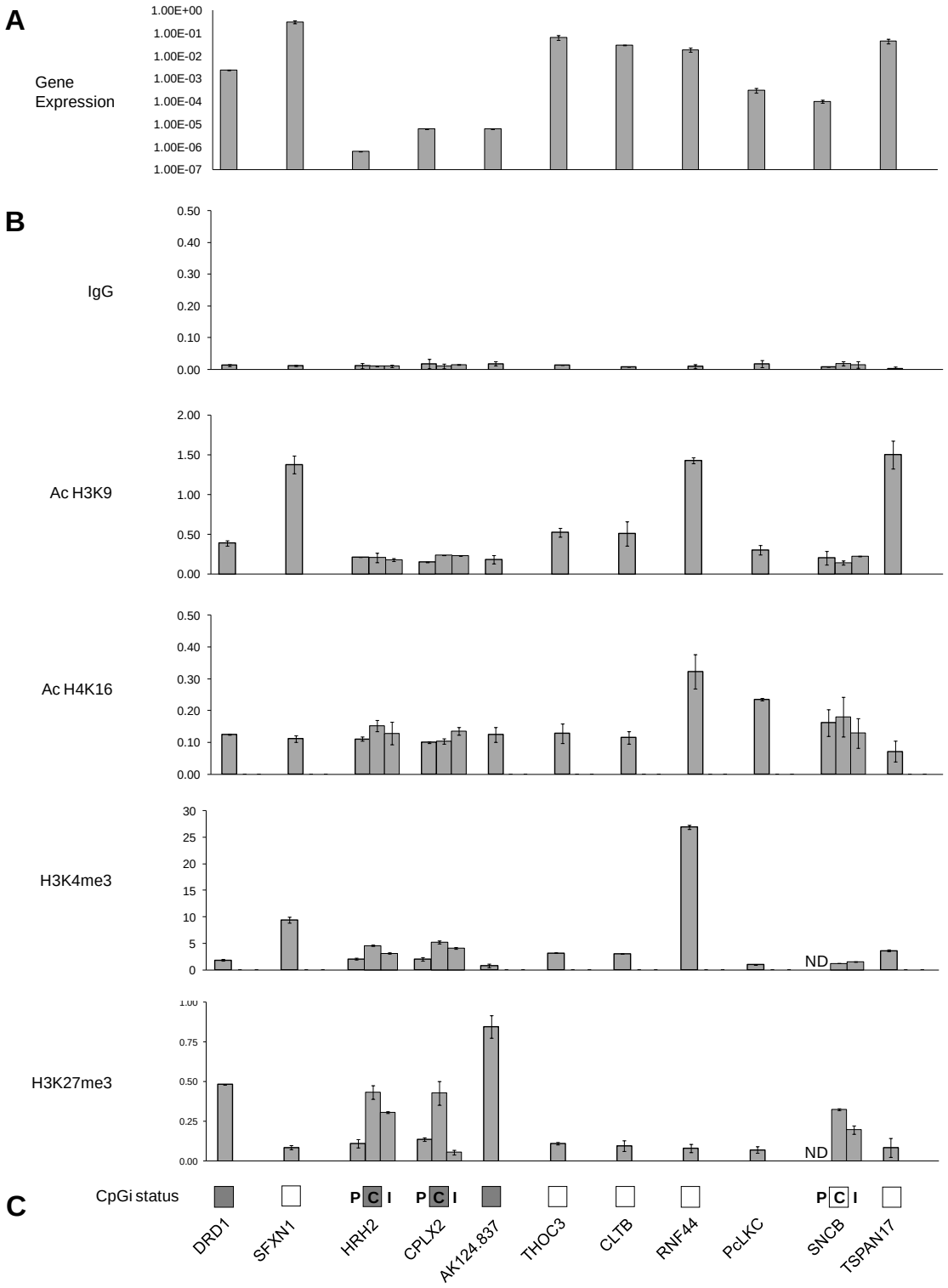

A
Gene Expression
1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06
1.00E-07
B
IgG
Ac H3K9
Ac H4K16
H3K4me3
H3K27me3
ND
ND
C
CpGi status
P C I
P C I
P C I
DRD1
SFXN1
HRH2
CPLX2
AK124.837
THOC3
CLTB
RNF44
PcLKC
SNCB
TSPAN17

FIGURE 5

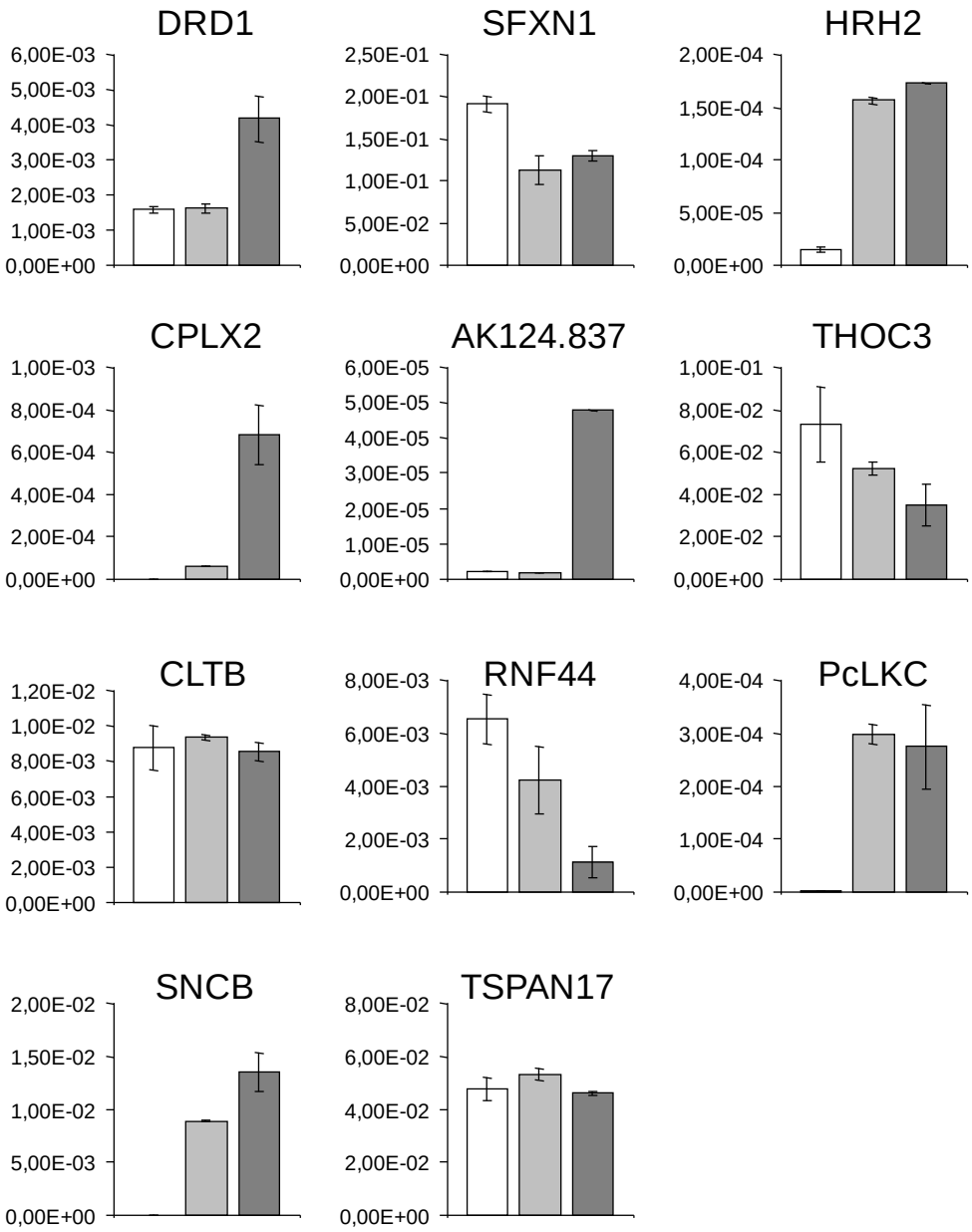

DRD1
6,00E-03
5,00E-03
4,00E-03
3,00E-03
2,00E-03
1,00E-03
0,00E+00
SFXN1
2,50E-01
2,00E-01
1,50E-01
1,00E-01
5,00E-02
0,00E+00
HRH2
2,00E-04
1,50E-04
1,00E-04
5,00E-05
0,00E+00
CPLX2
1,00E-03
8,00E-04
6,00E-04
4,00E-04
2,00E-04
0,00E+00
AK124.837
6,00E-05
5,00E-05
4,00E-05
3,00E-05
2,00E-05
1,00E-05
0,00E+00
THOC3
1,00E-01
8,00E-02
6,00E-02
4,00E-02
2,00E-02
0,00E+00
CLTB
1,20E-02
1,00E-02
8,00E-03
6,00E-03
4,00E-03
2,00E-03
0,00E+00
RNF44
8,00E-03
6,00E-03
4,00E-03
2,00E-03
0,00E+00
PcLKC
4,00E-04
3,00E-04
2,00E-04
1,00E-04
0,00E+00
SNCB
2,00E-02
1,50E-02
1,00E-02
5,00E-03
0,00E+00
TSPAN17
8,00E-02
6,00E-02
4,00E-02
2,00E-02
0,00E+00

FIGURE 6

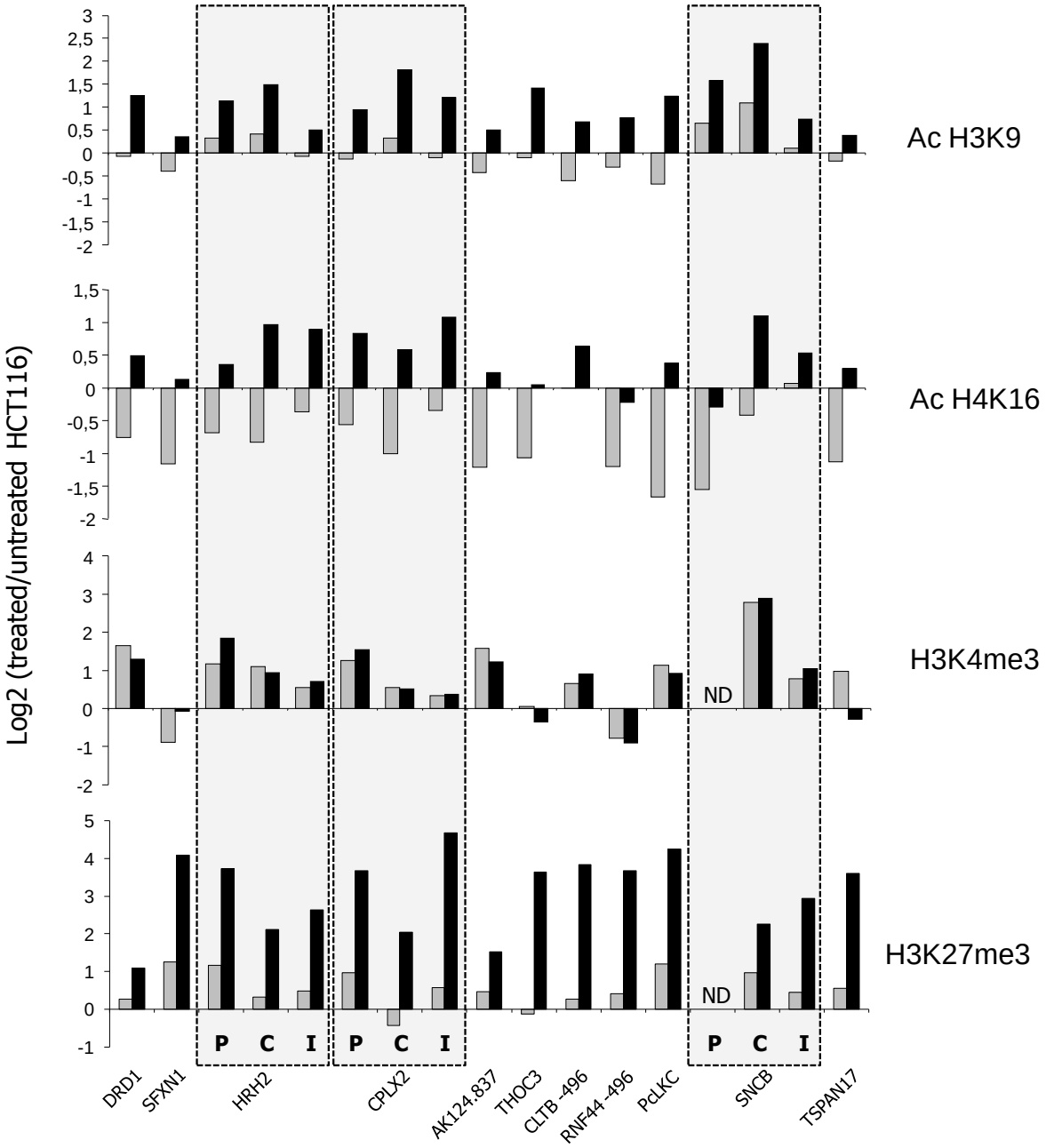

Log2 (treated/untreated HCT116)
Ac H3K9
Ac H4K16
H3K4me3
H3K27me3
ND
ND
P C I
P C I
P C I
DRD1
SFXN1
HRH2
CPLX2
AK124.837
THOC3
CLTB -496
RNF44 -496
PcLKC
SNCB
TSPAN17

FIGURE 7

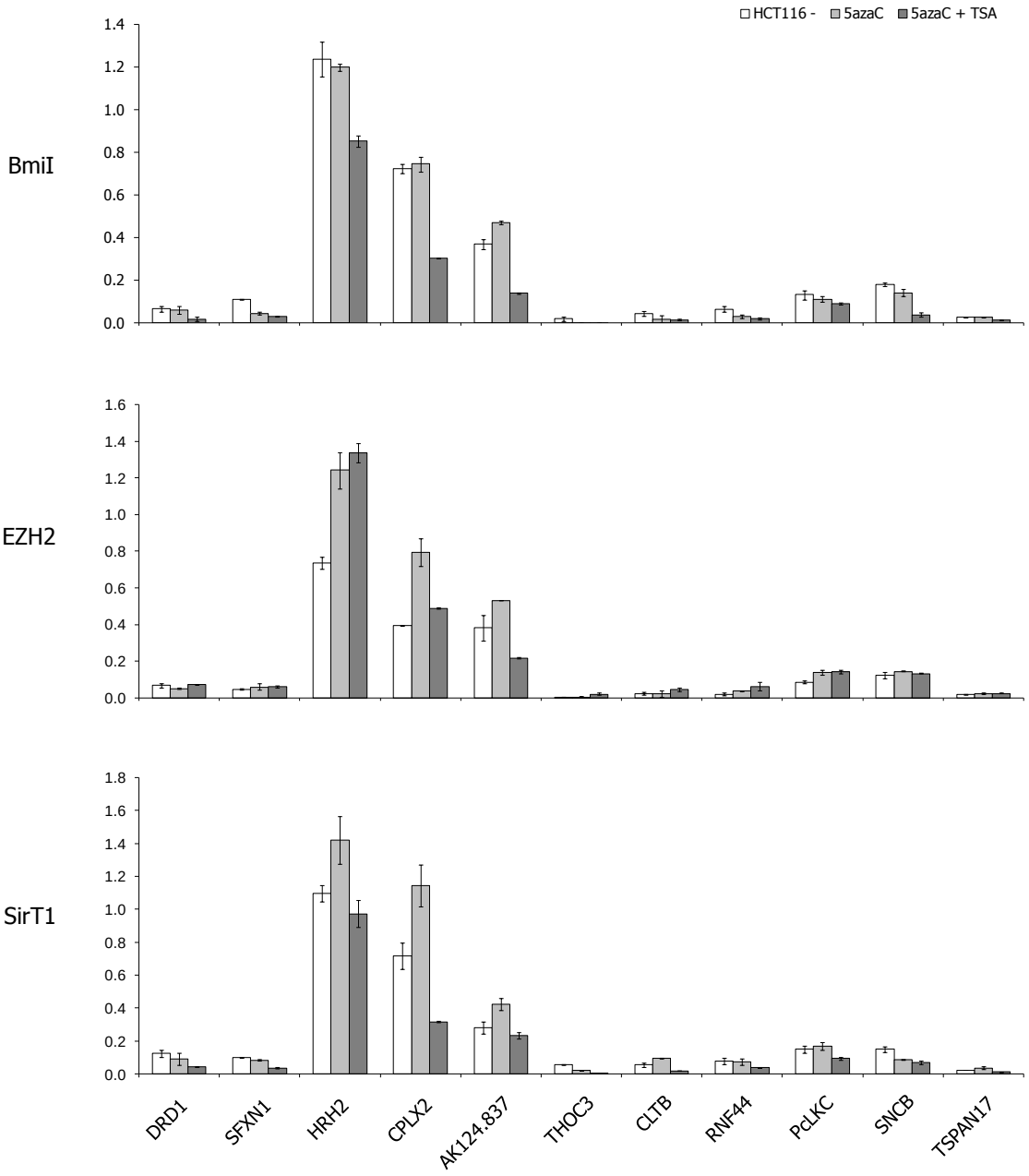

HCT116 -
5azaC
5azaC + TSA
BmiI
EZH2
SirT1
DRD1
SFXN1
HRH2
CPLX2
AK124.837
THOC3
CLTB
RNF44
PcLKC
SNCB
TSPAN17

FIGURE 8

Supplementary Table 1. Characteristics and positions of CpG islands across 5q35.2

Gene	CpG island	CpG island position[a]	Length bp	% G+C	% CpG[b]	Ratio O/E[c]	Relative position to CpG118[d]
DRD1	145	174803360-174804951	1592	65.2	18.2	0.86	-351265
SFXN1	93	174837949-174838671	723	74.3	25.7	0.94	-317545
HRH2	72	175017611-175018362	752	57.8	19.1	1.16	-137854
CPLX2	118	175156216-175157285	1070	71.8	22.1	0.86	0
AK124.837	125	175231122-175232672	1551	67.2	16.1	0.72	73837
THOC3	52	175327533-175327989	457	70.5	22.8	0.93	170248
BC042064	78	175420174-175421036	863	68.7	18.1	0.77	262889
BC034407	47	175558573-175559091	519	70.7	18.1	0.72	401288
CLTB	100	175775725-175776450	726	76.9	27.5	0.93	618440
RNF44	126	175896105-175897722	1618	61.6	15.6	0.82	738820
5' PcLKC	110	175902356-175903659	1304	52.8	16.9	1.21	745071
3' PcLKC	91	175956384-175957257	874	62	20.8	1.08	799099
GPRIN1	114	175969069-175970163	1099	63.1	20.7	1.04	811784
SNCB	79	175989127-175990100	974	55.4	16.2	1.09	831842
TSPAN17	47	176006745-176007229	485	71.1	19.4	0.8	849460

a. Base positions based on March 2006 build of the UCSC genome browser
b. Ratio of the observed CpG dinucleotides and the CpG island length
c. Ratio of the observed (O) and expected (E) CpG dinucleotides in each CpG island
d. Total distance between most distal ends CpG145 (DRD1) and CpG47 (TSPAN17): 1,203,869 bp. Base positions count from 5′ end

Supplementary Table 2. PCR primers used for bisulfite genomic sequencing.

CpG island	Gene	Product position[a]	Forward primer outer	Reverse primer outer	Forward primer inner	Reverse primer inner	Product length (bp)
CpG145	DRD1	174804213 - 174804583	TTTTAATTAGGGTTGGAGATGT	TTTCAAATTCTAACAATCATAAAC	TTCCCTATTTTCAAATTCTAACAA	None	371
CpG93	SFXN1	174837971 - 174838285	AAATGGAATATAGTGGTTTTAGT	TTCCCCACCCACCTAACTA	TGAGGATGGGTTTGGAGTTT	CCCACCCACCTAACTAAAC	315
CpG72	HRH2	175018104 - 175018391	GGGTGGATTTGGAAAGTGT	TTACTCTACTCATCCCACAA	None	TCCAAATATCCCCAACAAAA	288
CpG118	CPLX2	175157004 - 175157373	GGTTTAGTTAAGAGTGGGT	TACAACCTCCCAAATCTAAA	GGTAGAGGGGAAATAGGT	CCCCCCTACCTTCTCTA	370
	CPLX2 MIR	175157473 - 175157757	AGGTTTTTAAATAGGGAAGT	TTACCCAAAATCATATACTAA	TTTAAATAGGGAAGTTAGTTT	CACTCTACTATTCTCCTTA	285
CpG125	AK124.837	175231657 - 175231864	GGAGATTTAAGTTTGAATTTAG	CCCCCATCCCCTATCAATA	TTAAGTTTGAATTTAGAGAATAAT	TCTCCTCCCCTCCACCTA	208
CpG52	THOC3	175327833 - 175328014	GAGGGTTTTATGGTTGTAG	AACCAAATATTACTAAATAAACAA	TTTTAATTTATAATATTGTAGTAGT	None	182
CpG78	BC042064	175420128 - 175420425	GATGTATTTGATTTTGAGGATATT	TTACAATTACAACTCAAACACTA	None	ATACACAACCTACCTAATCTAAA	298
CpG47	BC034407	175558616 - 175558855	AGTTGTTTTGTTTGTTTTAGGAT	CCAAAACCACTATCTCTCCAA	GTTAGGTTTTGGGTATTAGTT	TACTCACAAAAATAAATCTTACTA	240
CpG100	CLTB	175775945 - 175776359	GAGAAGAAGTTAAAGTTATTAGT	CAAAATAAATCAAAAACCAAATCTA	None	ACAATACCTAACCCAAATATCC	415
CpG126	RNF44	175896049 - 175896291	AGAAGAAATTAGGTGGTTATAT	TCCCCTCTCCTACAACTAAA	GTGGGGTTTTGGGAAGTTTAT	CACTTCCTTTATTATTACTTTTA	243
CpG110	None	175903332 - 175903548	TAATAGTTGTTTGGGGTTAGT	ACAAAAACTACCACCACCTA	ATAGTTGTTTGGGGTTAGTAG	CCCTAATCCAAATTAACTCTAA	217
CpG91	GPRIN1	175956397 - 175956634	GTTTGTGATTAAGAAGGGAGT	AATACTAAACATAACCATCCAA	GTATAAAATGGGGGTAGATTAT	AACCATCCAAAAACATCTAAAA	238
CpG114	GPRIN1	175969000 - 175969337	GTATTTTGGGTGATAGGTGGT	AAATATCTATTTCCAAACTCCTC	TGTGTATGAGTTTAATGGAATGAT	AACCAAACTAACCTAACTATAAAT	338
CpG79	SNCB	175989268 175989476	TTTTTGGTTATGGATAGGTTT	CCCCAAATTTCAAAAACACTAA	GGTTTGGGGAGGGTGATAT	None	209
CpG47	TSPAN17	176006967 - 176007253	GGATTGAGTTGATTTATGGG	ATTCCTAAAAATACTAATACTTAC	GATTTTAAAGGGATTGAGGGT	None	287

a. Base positions based on March 2006 build of the UCSC genome browser
Primers are shown in the 5' to 3' orientation

Supplementary Table 3. PCR primers for expression analysis.

Gene	Product position[a]	Forward primer	Reverse primer	Product length
DRD1	174801483 - 174801614	CAATGGGGCCGCGATGTTTT	ATGCCAGCTGCCTCCTCCTTTTT	132
SFXN1	174869765 - 174871765	ATGCCGTCGTCAATTACACCAACAG	CTACGGCAGCAAAGGGAACAAAAC	174
HRH2	175017711 - 175042475	CGCGGACCGAGGCGAACC	GAATCCGAGGCACTGTCTGG	211
CPLX2	175238651 - 175239628	GAGGCGGAGCGGGAGAAGGTC	GCCCGGGCAGGTATTTGAGCA	214
AK124837	175231117 - 175239969	GGGAGCGGTAGAACGTCAGGGTAT	AGTGGGGAGCAGGGAGGAATGTAT	159
THOC3	175326847 - 175327758	GATGGCCCCGTGGTGCTCAG	CCCATGTCCCCGATAATTGTTTTCT	238
CLTB	175752420 - 175757578	GCAAGTGGCGAGAGGAGCAGAGGA	CGAGCGCAGGCGGGACACAT	287
RNF44	175888572 - 175889154	ACAGCCATCAGTCGGAGCAGA	GGCCCCACCCACAAGTTTC	251
PcLKC	175950826 - 175952388	GAAGCTTCAAGCTATGAAGG	CCTTGATTTCCTGACTGTTC	245
SNCB	175980367 - 175986061	AACATCGCAGCAGCCACAGGAC	GGGCAGGGACAGGGACAGAAT	231
TSPAN17	176011408 - 176012474	CCCCGTGTGGCTGTTTGTGGTAGT	CCGGTAGGCCTTGACGTTGTTGTT	226

a. Base positions based on March 2006 build of the UCSC genome browser
Primers are shown in the 5′ - 3′ orientation

Supplementary Table 4. Primer sets for Real Time PCR ChIP analysis.

Gene	Region[a]	Product position[b]	Forward primer	Reverse primer	Product length
DRD1	Promoter+CpGi	174803937 - 174804181	GCCGGCCGTTCTAGGAGTTGGT	GAGGACGCCCCGGTGAGTGC	245
SFXN1	Promoter+CpGi	174837851 - 174837971	GTGAACGGATGCTGGCTACG	ACGGTGGTGTGGGGCTCTG	121
HRH2	Promoter	175017179 - 175017347	TATTTCCCCCTTCTAGCCTGACTC	AGATTGTGAGCCCGCCCTTGA	169
	CpGi	175018145 - 175018361	TTTTGTCTCCGGGTCTGGGTGGTG	GCTGAGCGCTGCCTGTTGTGG	217
	First intron	175037215 - 175037372	ATAACCCCTCCCAGCCCATTTTG	GCACCGCCCAGTTCCTCATTCT	158
CPLX2	Promoter	175155821 - 175156051	TGAAGGATCTCGTCAAAAAGTTAG	GCAGTTCTGCCCTGAAGTTC	165
	CpGi	175157105 - 175157277	CCTCCAGACCCCCACCCCATCC	AGGCTTCCCCGCGGCTTCTCAG	173
	Second intron	175186220 - 175186399	TGGATGTGGGAAGGAGACGAAGAA	CAGTTTTGAGGGAGTGGGTGTTGG	180
	CPLX2 distal promoter	175230775 - 175230954	ACCCCTTGCCCCCACCACT	CGTCTCCTTCTCCCTCCATTCC	180
THOC3	THOC3 promoter	175328169 - 175328287	TTCTTCTGGGGTTGTTCCGTAATC	CACACCCGCTAGCCCTTTTCAT	119
CLTB	CLTB promoter	175776609 - 175776822	AGGCCGTGTTGTTTTGTGTATTTT	CAGGCAGGGAGGAGGCATTCAGA	214
RNF44	Promoter+CpGi	175897523 – 175897694	GCCGCCCCACTTCTCCTTGTT	GGGCGTGGCGGTTAAATGTCTC	172
PcLKC	Promoter	175908639 - 175908817	ACGGGGGCTCTGGCATCA	GGACAGGAGGAAGGGACC	179
SNCB	Promoter+CpGi	175990064 - 175990302	CAGCCACCGCTCGTTCCT	CCTCAACTCCCAACAGACTAC	239
	CpGi	175989488 - 175989660	GCCACTGCCTCGGTTATCC	AGCGTCTGCGTGTCCTGTG	173
	Fourth exon-intron	175985907 - 175986046	GGGGCGGGGATGGTGACA	ACAGGACTGGTGAAGAGGGAGGAAT	140
TSPAN17	Promoter+CpGi	176006849 - 176007062	TGCGAGGAGGTCCGAATCAG	GACCGCCCTCAGCATCCAG	214

a. CpGi, CpG island

b. Base positions Base positions based on March 2006 build of the UCSC genome browser

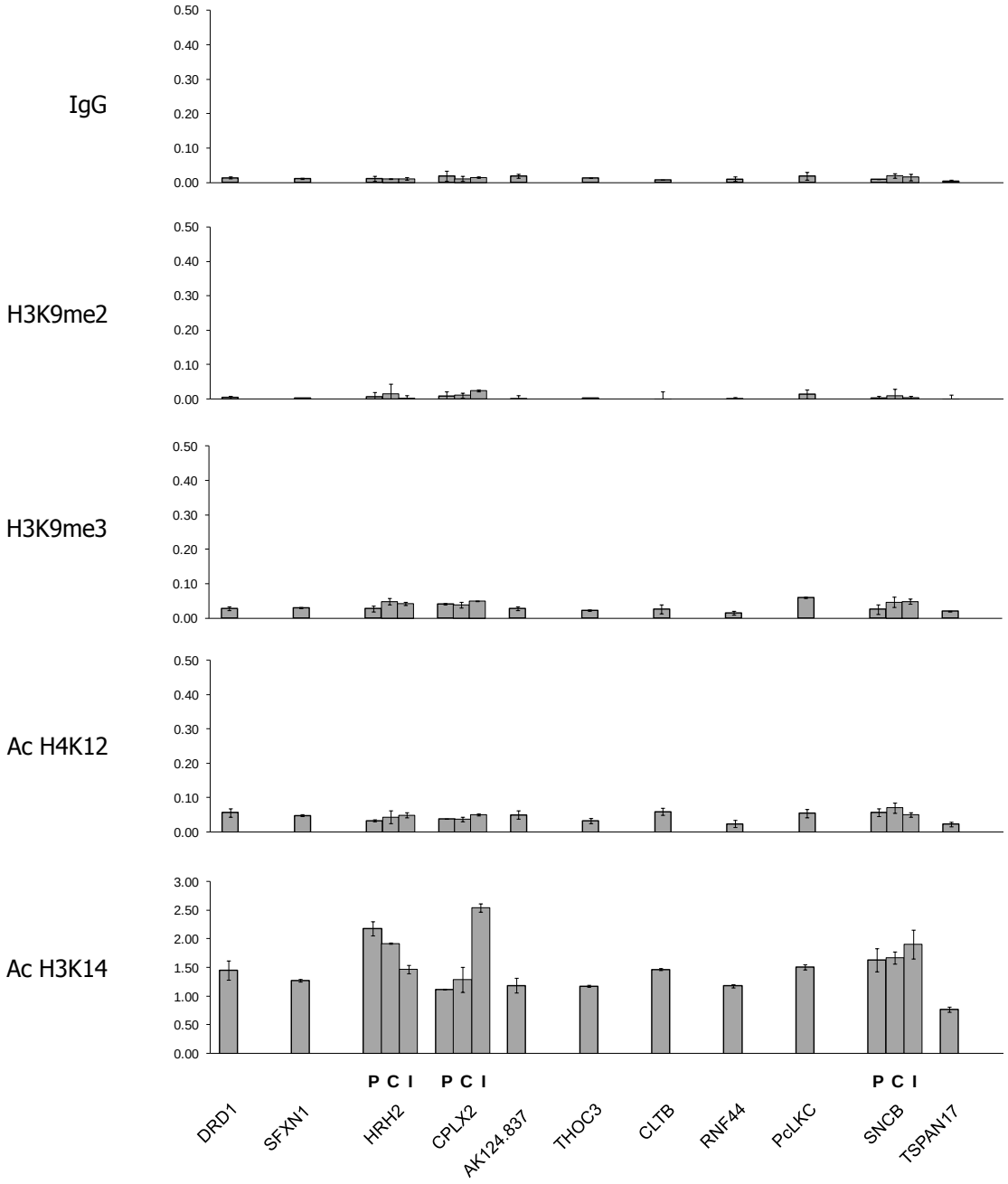
Primers are shown in the 5′ – 3′ orientation

Supplementary Table 5. Bivalent domains and CTCF binding across 5q35.2 in different cell types

Gene	HCT116 cells[a]	Murine ESC[b]	Human ESC[c]	Differentiated Human cells[d]	CTCF sites[d]
DRD1	Bivalent	Bivalent	Bivalent	H3K27me3	
SFXN1	H3K4me3	H3K4me3	H3K4me3	H3K4me3	SFXN1 and HRH2
HRH2	Bivalent	Bivalent	None	H3K4me3	HRH2 and CPLX2
CPLX2	Bivalent	Bivalent	H3K4me3	H3K27me3	
AK124.837	Bivalent	Bivalent	H3K4me3	H3K27me3	
THOC3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	THOC3 and BC042064
CLTB	H3K4me3	H3K4me3	Bivalent	H3K4me3	CLTB CpG island
RNF44	H3K4me3	H3K4me3	H3K4me3	H3K4me3	
PcLKC	Bivalent	H3K4me3	None	None	Intragenic PcLKC
SNCB	Bivalent	Bivalent	Bivalent	H3K27me3	
TSPAN17	H3K4me3	H3K4me3	H3K4me3	H3K4me3	Downstream TSPAN17

a. Data summary from Figure 5B
b. Data summarized from [Mikkelsen, 2007 #270]
c. Data summarized from [Pan, 2007 #296] and [Zhao, 2007 #297]
d. Data summarized from [Barski, 2007 #272]. Where two genes are shown, CTCF site maps at intergenic space between them
Primers are shown in the 5′ – 3′ orientation

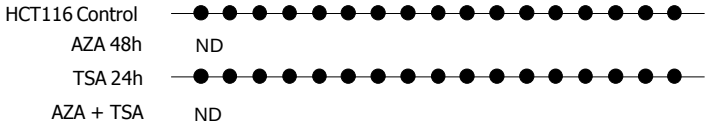
Supplementary Figures



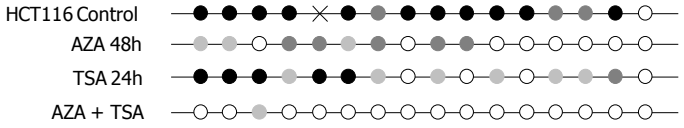
Supplementary Figure 1.

Levels of different chromatin modifications compared to the IgG negative control. Both H3K9me2 and H3K9me3 marks, associated to inactive chromatin, are found at very low levels across the region. Acetylation of the lysine 12 in histone H4 is also detected at low levels while acetylation of the lysine 14 in histone H3 is found at high levels across the whole region. All values are expressed as absolute levels over the input fraction.

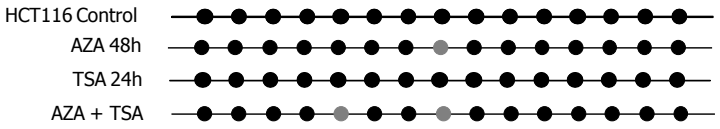
#72 HRH2 CpG island



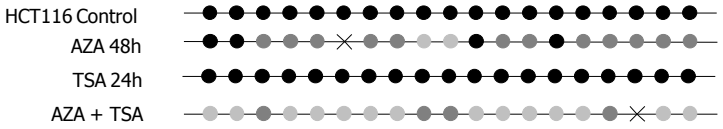
#118 CPLX2 CpG island



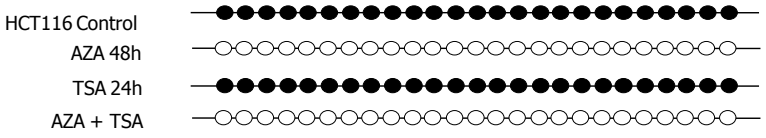
#125 AK124.837 CpG island



#78 BC042064 CpG island

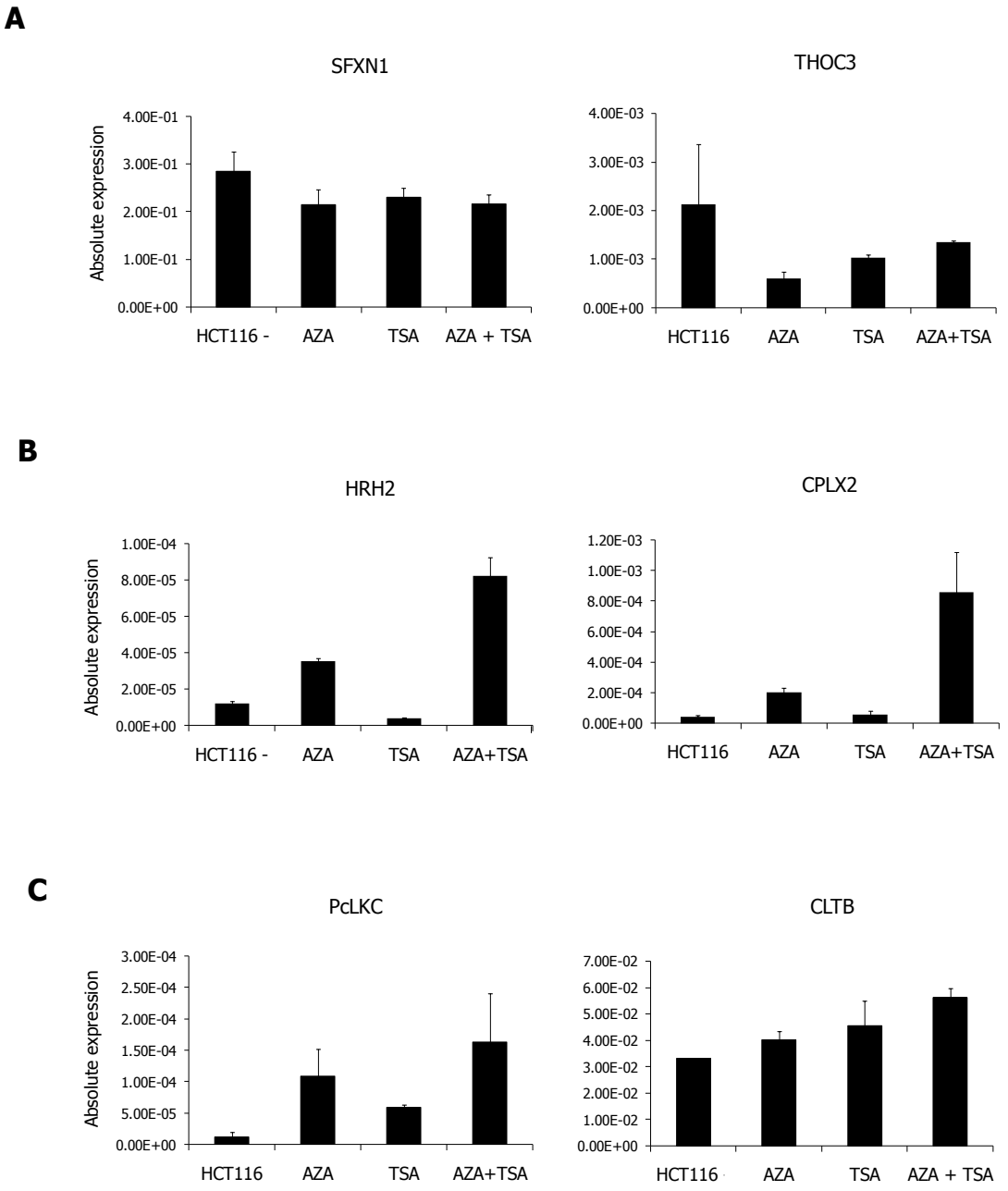


#91 GPRIN1 CpG island



Supplementary Figure 2.

DNA methylation data of methylated genes after 5azaC, TSA and co-treated cells. Individual CpG residues are represented by white circles (unmethylated), light gray circles (25% methylated), dark gray circles (50% methylated) and black circles (over 75% methylated). Data has been obtained by direct sequencing of sodium bisulfite treated DNAs. HRH2 CpG island recurrently failed to amplify the 5azaC treated DNAs. ND, not determined.



Supplementary Figure 3.

TSA treatment has different effects on the expression levels of genes across 5q35.2. Highly expressing genes become weakly downregulated after TSA treatment, similarly to the downregulation seen after 5azaC and drug co-treatment (A). DNA methylated, deeply silenced genes do not become upregulated after TSA treatment (B), while the downregulated PcLKC gene, which lacks promoter CpG Island, and the unmethylated and weakly downregulated CLTB gene, are the only genes in the region that weakly upregulate upon TSA treatment (C).