Supporting Information

Rodriguez *et al.* 10.1073/pnas.0810133105

SI 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 −80°C. 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, and CaCo-2 were obtained from the American Type Culture Collection (ATCC). The KM12C cell line was generously provided by A. Fabra (Institut d'Investigació Biomèdica de Bellvitge, L'Hospitalet, Spain). DNA from tumor-normal pairs was obtained by conventional phenol-chloroform extraction and ethanol precipitation. DNA purity and quality was checked by 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). One microgram of DNA was digested with 20 U of the methylation-sensitive restriction endonuclease SmaI (Roche Diagnostics) for 16 h at 30°C. This enzyme recognizes and cleaves the sequence CCCGGG leaving blunt ends (CCC/GGG). Adaptors were prepared by incubating the oligonucleotides Blue (CCGAATTCGCAAAGCTCTGA) and the 5′ phosphorylated MCF oligonucleotide (TCAGAGCTTTGCGAAT) at 65°C for 2 min, and then cooling to room temperature for 30–60 min. One microgram of the digested DNA was ligated to 2 nmol of adaptor using T4 DNA ligase (New England Biolabs). The products were purified using the GFX Kit (Amersham Biosciences) 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 AUMAs were performed for normal-tumor comparisons, yielding a clear and highly reproducible 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). For a detailed description of the method see reference (1).

Bisulfite Genomic Sequencing. The bisulfite reaction was carried out on 2 μg of mechanically sheared DNA for 16 h at 55°C under conditions previously described (2). Before sequencing, DNA was amplified using a nested or a semi-nested PCR approach, as appropriate. A minimum of two independent PCRs were carried out and pooled together to ensure a representative sequencing. Primers sequence may be requested form the authors.

Gene Expression Analysis. Whole cDNA was obtained by retrotranscription of 500 ng of whole RNA with M-MLV retrotranscriptase (Invitrogen) using random hexamers (Amersham Biosciences) at 37°C for 1 h. cDNA levels were quantified by using the LightCycler 2.0 real time PCR system with Fast Start Master SYBR Green I kit (Roche Diagnostics). For a 10-μl PCR 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. Sequence of PCR primers may be requested from the authors. Gene expression values were normalized by using cyclophilin A (PPIA), beta-2-microgluobulin (B2M), and 18S.

Chromatin Immunoprecipitation (ChIP) Assay. Briefly, 6×10^6 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 1 ml lysis buffer (1% SDS, 5 mM EDTA, 50 mM Tris, pH8) and were sonicated in a cold ethanol bath for 10 cycles at 100% amplitude using a UP50H sonicator (Hielscher). Chromatin fragmentation was visualized in 1% agarose gel, with the sonicated fragments in the 200–500 bp range. Soluble chromatin was obtained by centrifuging the sonicated samples at 14,000 $\times$ *g* for 10 min at 4°C. The soluble fraction was diluted 1/10 in dilution buffer (1% Triton X-100, 2 mM EDTA, 20 mM Tris, pH8, 150 mM NaCl), 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 1 ml of chromatin. Soluble chromatin was immunoprecipitated with specific antibodies against BMI1, CTCF, acetylated (Ac) H3K9, Ac H3K14, Ac H4K12, Ac H4K16, dimethylated (me2) H3K9, trimethylated (me3) H3K9, H3K27me3 (Upstate, Millipore), H3K4me3 (Abcam), EZH2 (Santa Cruz), SIRT1 (Delta Biolabs) and IgG (Jackson ImmunoResearch). Chromatin-antibody complexes were pulled down using a 50% protein A/G slurry (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).

Enrichment for a given chromatin modification was quantified as a fold enrichment over the input using a 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 sequence may be requested from the authors.

Drug Treatments. HCT116 cell line was seeded at low density 24 h before treatment. The drug 5-aza-2′-deoxycytidine (5-AzaC) was added to media (0.5 μM) for 48 h, after which it was removed and cells were allowed to recover in fresh media for further 24 h before harvesting DNA and RNA. For 5-AzaC and Trichostatin A (TSA) cotreatments, cells were grown in the presence of 5-AzaC for 48 h after which TSA (0.1 μM) was added for further 18 h. Before RNA and DNA harvesting, cells were allowed to recover in fresh media for 24 h.

1. Rodriguez J, *et al.* (2008) Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Nucleic Acids Res* 36:770–784.
2. Frigola J, Song J, Stirzaker C, Hinshelwood RA, Peinado MA, Clark S (2006) Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* 38:540–549.

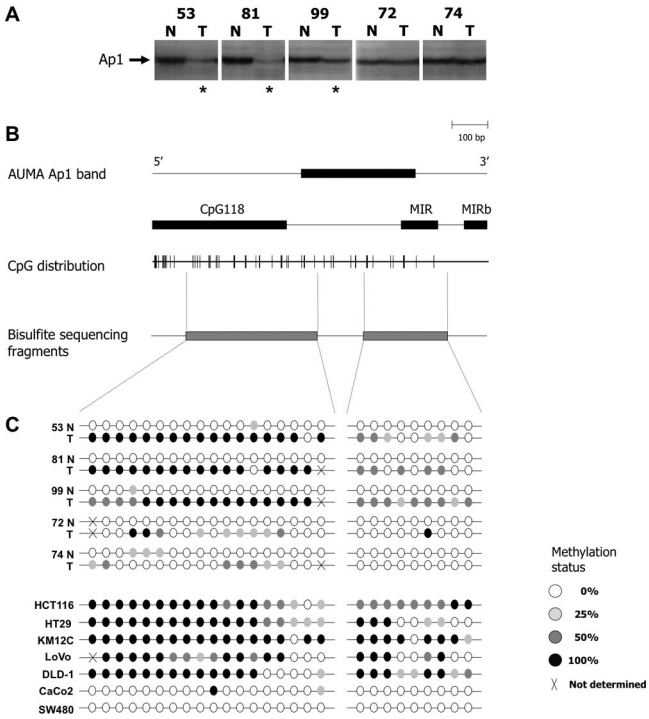

Fig. S1. Identification and characterization of recurrent DNA hypermethylation in colorectal cancer. AUMA Ap1 band was found to be methylated (missing) in 36 out of 50 (72%) tumor samples. (*A*) Fragment of a polyacrylamide denaturing gel displaying the AUMA Ap1 band in five normal-tumor pairs. The presence of the band in all normal tissues indicates the unmethylated state of the two SmaI sites flanking the AUMA targeted sequence. Decreased band intensity (as marked by an asterisk in three of the tumors) is indicative of methylation of one or both of the SmaI target sites in the tumor sample. (*B*) Gel isolation, cloning and sequencing, allowed us to map the Ap1 band on the genomic region 5q35.2, with its 3′ end is inside a MIR element. The amplified sequence was flanked by CPLX2 CpG island (CpG118) and a MIR element. CpG dinucleotides are represented as short vertical lines. Two fragments were analyzed by bisulfite sequencing (gray boxes), one of them encompassing the 3′ end of CpG118 and a second fragment encompassing the MIR repetitive element, therefore covering the two SmaI sites. (*C*) Methylation data were obtained by bisulfite sequencing for five normal-tumor pairs and seven colon cancer cell lines. Each CpG dinucleotide is represented by a dot. Three tumors exhibited a high degree of methylation, in agreement with AUMA data. The 5′ end of the Ap1 band mapped inside of the first intron of the CPLX2 gene, only 31 bp away from the CPLX2 CpG island (CpG118). CpG118 and the MIR element are normally unmethylated, but may be found partially to heavily methylated in colorectal tumors and cell lines.

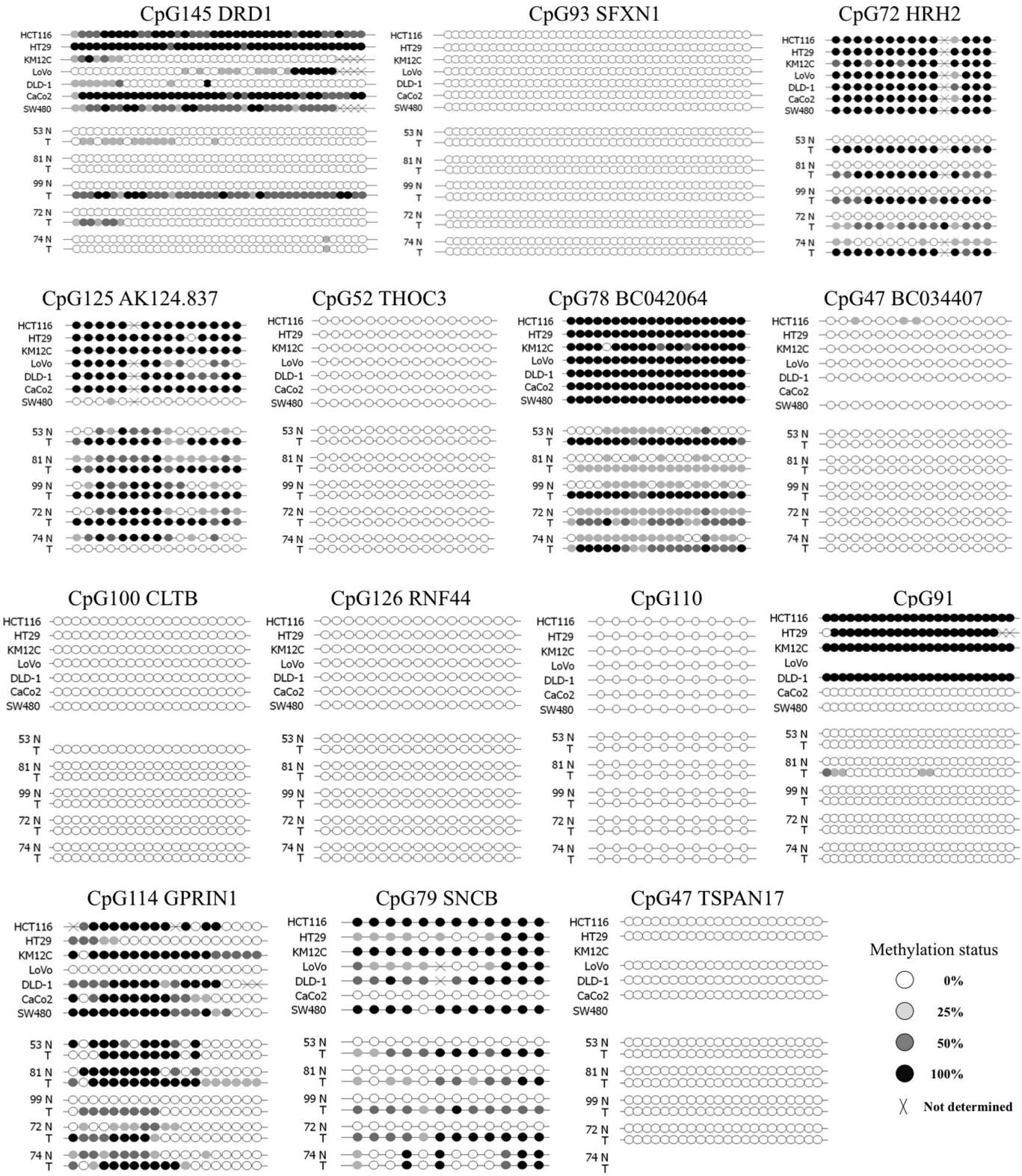

Fig. S2. Direct sequencing data of sodium bisulfite-treated DNAs of seven cell lines and five normal tumor pairs for CpG islands across 5q35.2. Data summarized in Fig. 1*A*.

Fig. S3. Expression and promoter DNA methylation status for 12 genes across 5q35.2 in a series of 16 normal-tumor pairs. Expression values are represented as the $\log_2$ of the tumor-normal ratio. Therefore, genes that are up-regulated in the tumor have values above 0 and down-regulated genes in the tumor have values below 0. The tumor DNA methylation status is noted as green and red, for unmethylated and methylated promoters, respectively. In normal tissue, all CpG islands were fully unmethylated (data not shown). PcLKC is the only gene in this study that does not have a promoter CpG island, and therefore its methylation status is depicted as a gray bar. ND, not determined.

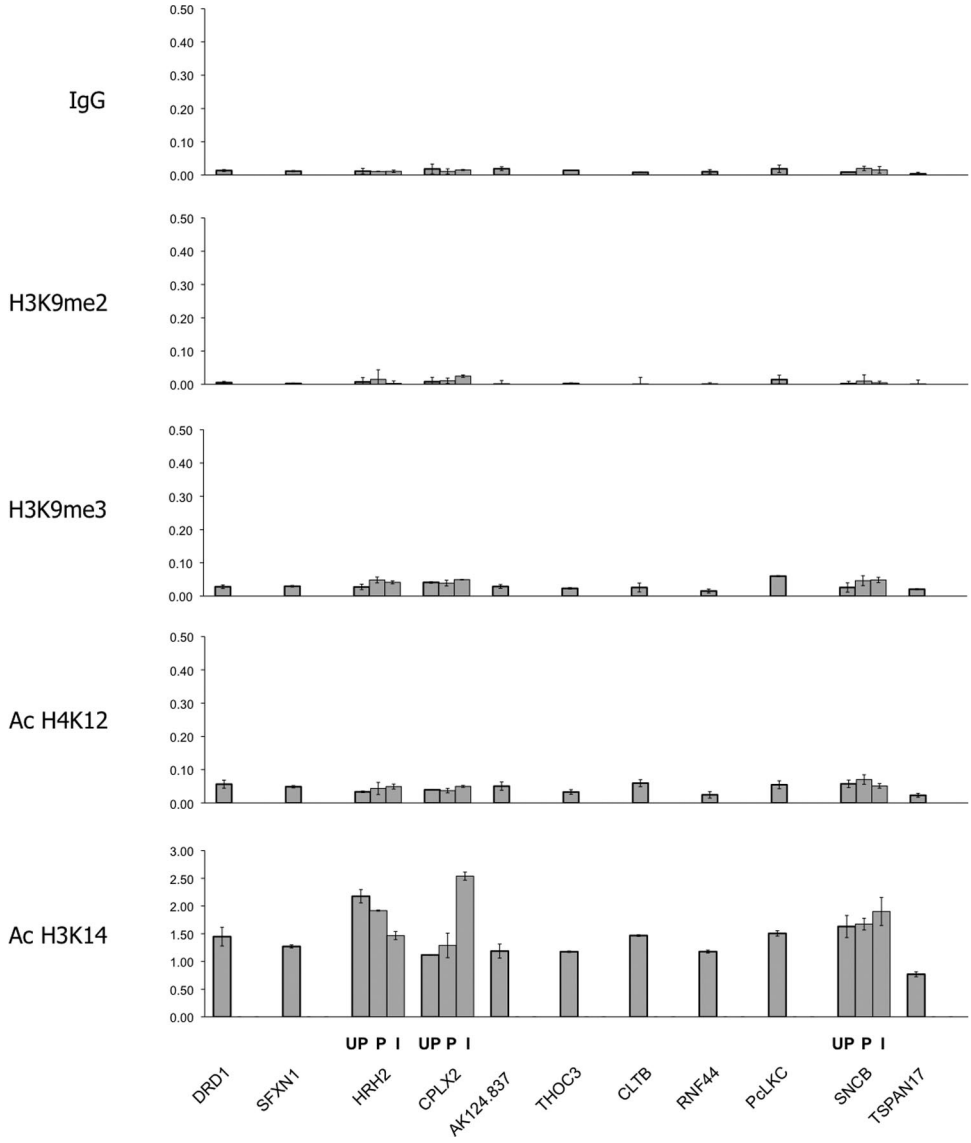

Fig. S4. Levels of different chromatin modifications compared to the IgG negative control across genes in 5q35.2. For three of the genes (HRH2, CPLX2, and SNCB) two additional regions were amplified besides the promoter (P) region, corresponding to an upstream promoter (UP) region and a distal intronic (I) region. Both H3K9me2 and H3K9me3 marks, associated to inactive chromatin, are found at very low levels at all sites tested, including introns. 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.

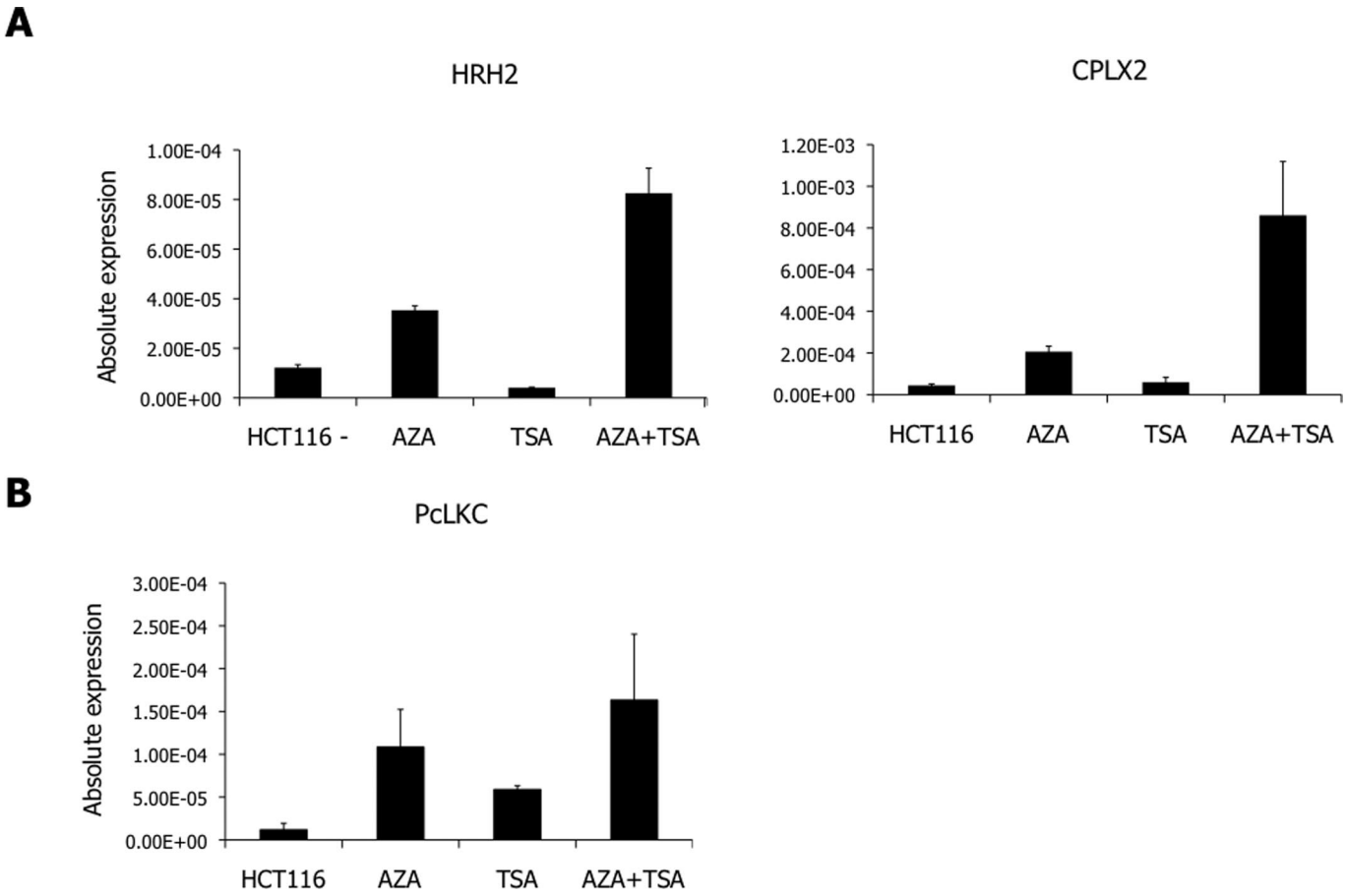

Fig. S5. Genes with promoter DNA methylation do not become re-expressed after TSA treatment (*A*) while genes such as the down-regulated PcLKC gene, which lacks promoter CpG island, become weakly up-regulated after TSA treatment (*B*).

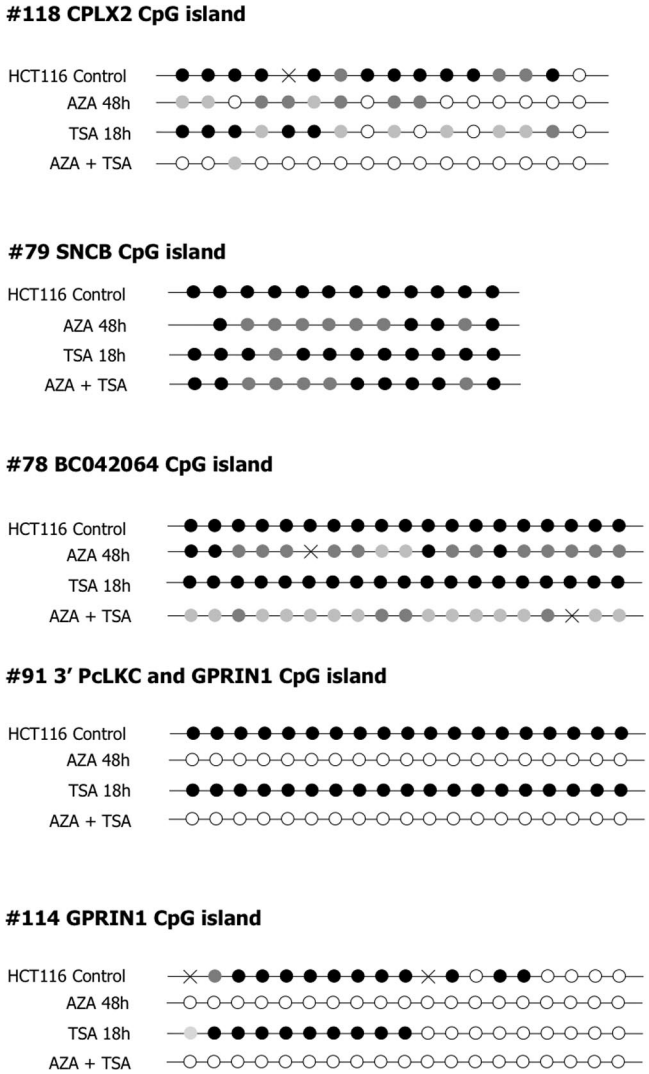

Fig. S6. DNA methylation data of methylated CpG islands after 5-azaC, TSA, and 5-azaC/TSA treatments. Individual CpG residues are represented by white circles (unmethylated), light gray circles (25% methylated), dark gray circles (50% methylated), and black circles (over 75% methylated). Data has been obtained by direct sequencing of sodium bisulfite treated DNAs. Not determined CpG residues are represented by a cross.

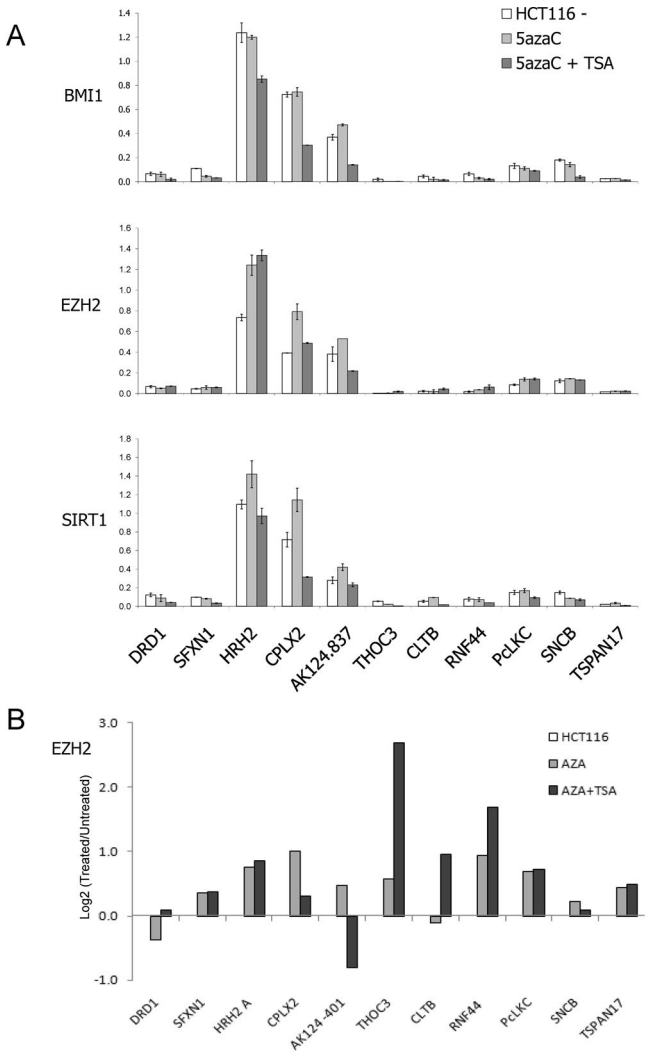

Fig. S7. Mapping of the PcG components EZH2 and BMI1 and the SIRT1 HDAC across gene promoters in 5q35.2. (*A*) Absolute enrichment levels over the input fraction are shown for the untreated control HCT116 cell line (white bars), 5-AzaC-treated (light gray), and 5-AzaC/TSA-treated cells (dark gray). (*B*) Relative increase in EZH2 enrichment levels over the input fraction for the 5-AzaC treated (light gray) and 5-AzaC/TSA-treated HCT116 cells (dark gray) represented as the $\log_2$ ratio in regard to the values in untreated control cells (0).

Table S1. Bivalent domains and CTCF binding across 5q35.2 in human cancer cells, human and murine ESCs, and human differentiated cells

Gene	CpGi status*	HCT116 cells†	Human ESC‡	Human ESC§	Differentiated human cells¶	Murine ESC‖	CTCF boundaries**
DRD1	Methylated	Bivalent	Bivalent	Bivalent	H3K4me3	Bivalent	DRD1 and SFXN1
SFXN1	Unmethylated	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	SFXN1 and HRH2
HRH2	Methylated	Bivalent	Bivalent	None	Bivalent	Bivalent	HRH2 and CPLX2
CPLX2	Methylated	Bivalent	Bivalent	H3K4me3	H3K4me3	Bivalent	
AK124.837	Methylated	Bivalent	Bivalent	H3K4me3	None	Bivalent	AK124.837 and THOC3
THOC3	Unmethylated	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	THOC3 and BC042064
CLTB	Unmethylated	H3K4me3	H3K4me3	Bivalent	H3K4me3	H3K4me3	CLTB CpG island
RNF44	Unmethylated	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	RNF44 and PcLKC
PcLKC	No CpGi	H3K4me3	ND	None	None	H3K4me3	PcLKC and GPRIN1
GPRIN1	Methylated	ND	Bivalent	H3K4me3	Bivalent	Bivalent	GPRIN1 and SNCB
SNCB	Methylated	Bivalent	Bivalent	Bivalent	H3K4me3	Bivalent	
TSPAN17	Unmethylated	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	3′ TSPAN17

ND, not determined.

*Data summarized from Figure 1.

†Data summarized from Figure 3.

‡Data summarized from Pan G, *et al*. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1:299–312.

§Data summarized from Zhao XD, *et al*. (2007) 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.

¶Data summarized from Barski A, *et al*. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.

‖Data summarized from Mikkelsen TS, *et al*. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560.

**Data summarized from CTCFBSDB (http://insulatordb.utmem.edu/).

Table S2. Characteristics and positions of CpG islands analyzed across 5q35.2

Gene	CpG island	CpG island position*	Length, bp	% G + C	% CpG†	Ratio O/E‡	Relative position to CpG118§
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
None	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

*Base positions based on March 2006 build of the human genome.
†Ratio of the observed CpG dinucleotides and the CpG island length
‡Ratio of the observed (O) and expected (E) CpG dinucleotides in each CpG island.
§Total distance between most distal ends CpG145 (DRD1) and CpG47 (TSPAN17): 1,203,869 bp.