A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer

A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer

Hiromu Suzuki[1], Edward Gabrielson[1,2], Wei Chen[1,3], Ramaswamy Anbazhagan[1,2], Manon van Engeland[1,4], Matty P. Weijenberg[5], James G. Herman[1] & Stephen B. Baylin[1,3,6]

Published online: 6 May 2002, DOI: 10.1038/ng892

Aberrant hypermethylation of gene promoters is a major mechanism associated with inactivation of tumor-suppressor genes in cancer. We previously showed this transcriptional silencing to be mediated by both methylation and histone deacetylase activity, with methylation being dominant. Here, we have used cDNA microarray analysis to screen for genes that are epigenetically silenced in human colorectal cancer. By screening over 10,000 genes, we show that our approach can identify a substantial number of genes with promoter hypermethylation in a given cancer; these are distinct from genes with unmethylated promoters, for which increased expression is produced by histone deacetylase inhibition alone. Many of the hypermethylated genes we identified have high potential for roles in tumorigenesis by virtue of their predicted function and chromosome position.We also identified a group of genes that are preferentially hypermethylated in colorectal cancer and gastric cancer. One of these genes, *SFRP1*, belongs to a gene family; we show that hypermethylation of four genes in this family occurs very frequently in colorectal cancer, providing for (i) a unique potential mechanism for loss of tumor-suppressor gene function and (ii) construction of a molecular marker panel that could detect virtually all colorectal cancer.

Introduction

A frequent epigenetic change in cancer involves aberrantly hypermethylated CpG islands in gene promoters, with loss of transcription of the genes[1]. Recognition of this promoter hypermethylation has fostered a growing effort to screen the cancer genome to identify such loci. These search strategies, including the identification of hypermethylated CpG islands in regions of high-frequency loss of heterozygosity[2] and throughout the genome[3–5], have all proven useful for identifying hypermethylated CpG islands that are tumor-specific. However, each strategy is hindered by one or more factors: (i) the identification of sites that are not associated with gene promoters, (ii) the potential bias of methylation-sensitive restriction sites for CpG island subsets, (iii) the lack of utilized restriction sites in many islands and (iv) the need to laboriously search for nearby genes once the altered locus is identified.

We describe a new microarray-based strategy that combines gene expression status and epigenetic regulation. The approach is based upon our recent observation that silencing of hypermethylated genes in cancer is dependent on both methylation of dense CpG islands and histone deacetylase (HDAC) activity[6]. We show that this procedure robustly identifies new genes for which transcriptional repression might mediate tumorigenesis and helps define the nature of gene promoters controlled by either a combination of methylation and HDAC activity or only by HDAC. Our gene screening technique has led to the identification of gene hypermethylation events that cluster within specific tumor types, and can simultaneously involve several members of a single gene family.

Results

Microarray analysis and categorization of upregulated genes

We used cDNA microarray technology to identify genes upregulated in the colorectal cancer (CRC) cell line RKO, after cells were treated with low-dose 5-aza-2′ deoxycytidine (DAC), which minimally blocks DNA methylation, and trichostatin A (TSA) to inhibit HDAC. In our first study with this drug combination, we showed that the low dose of DAC used and the short treatment time for the cells resulted in only a few alleles being demethylated, and these alleles may have led to the upregulation of gene expression[6]. This situation could diminish the sensitivity of a

[1]*The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins and* [2]*Department of Pathology, The Johns Hopkins University School of Medicine, 1650 Orleans Street, Baltimore, Maryland 21231, USA.* [3]*Predoctoral Training Program in Human Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.* [4]*Department of Pathology and* [5]*Department of Epidemiology, University Maastricht, The Netherlands.* [3]*Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. Correspondence should be addressed to S.B.B. (e-mail: sbaylin@welchlink.welch.jhu.edu).*

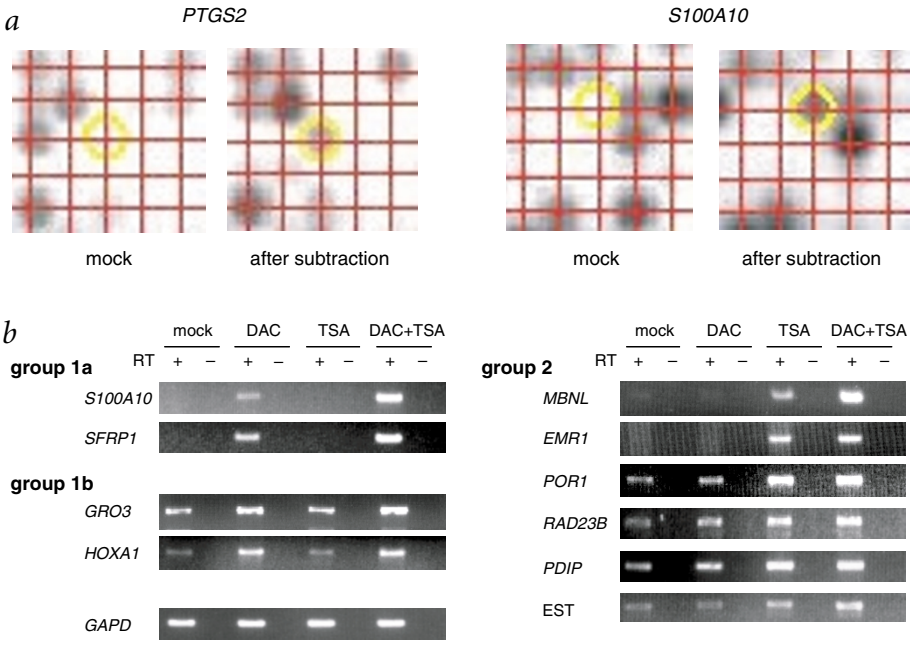

Fig. 1 Microarray analysis and categorization of genes by semi-quantitative RT–PCR. ***a***, Examples of subtraction microarray hybridization in RKO. Paired results of microarrays probed with non-subtracted RKO cDNA from mock-treated cells and the same array probed with post-subtraction cDNA from combination-treated cells are shown. Yellow circles indicate genes appearing only with the subtraction probe. ***b***, Representative results of semi-quantitative RT–PCR of RKO cells. Group 1a are genes without expression in mock- (lane 1) and TSA- (lane 5) treated RKO cells, upregulated after low-dose DAC treatment (lane 3) and showing greater re-expression after the combination of DAC and TSA (lane 7). Group 1b genes show similar patterns, although basal expression was detected in mock cells. Group 2 genes can be upregulated by TSA alone, whereas their basal expression and upregulation with DAC vary among genes. The gene *GAPD* serves as a positive control to ensure cDNA quality and loading accuracy. The accession number for the EST is W46439 (see Table 1).

microarray screen. Indeed, we initially failed to detect control genes arrayed on our filters, which were known to be synergistically reactivated by the drug combination, in the cell line studied[6,7]. We therefore increased the sensitivity of our screen by carrying out an initial cDNA subtraction step between mock-treated and DAC- and TSA-treated RKO cells, using the mock-treated cells as the driver and the drug-treated cells as the tester populations. The PCR product after the second round of subtraction was then used as a probe for microarray hybridization.

Of the four control genes (*MLH1*, *CDKN2A*, *TIMP3*, *PTGS2*) arrayed on the filters that were known to be methylated in RKO cells, we could not detect *MLH1* re-expression; however, we successfully detected, as validated by PCR, the other three control genes (Fig. 1*a*). For unknown genes, we selected those that showed no expression in the mock filter (genes having the same intensity as empty spots when probed with non-subtracted cDNA from mock-treated cells) and showed detectable expression after probing with the subtraction products from mock-treated and drug-treated cells (Fig. 1*a*). We analyzed these genes by semi-quantitative RT–PCR in cells subjected to mock treatment and DAC alone, TSA alone, or a combination of the two drugs. Of a total of 10,814 genes examined by subtraction microarray, 74 were upregulated by treatment with DAC and/or TSA. We divided these genes into two groups. Group 1 ($n = 51$) genes showed no change in expression with TSA alone, minimal increase in expression after low-dose DAC alone, and stronger induction by the combined DAC and TSA treatment (Table 1 and Fig. 1*b*). Group 1 genes could be subdivided into two groups: group 1a genes ($n = 24$) are completely inactivated in mock cells, whereas group 1b genes ($n = 27$) show some basal expression detected by RT–PCR (Fig. 1*b*). Group 2 genes ($n = 23$), by contrast, are upregulated in expression by TSA alone and have a variable initial expression or response to DAC alone (Fig. 1*b*).

Of all of the non-EST (expressed sequence tags) genes (Table 1), 56 had known chromosomal positions. We were able to identify a putative transcription start site for 46 of the genes by searching all available genome databases. We were also able to identify 5′ CpG islands (GC content > 60%, ratio of CpG to GpC > 0.6 and minimum length 200 bp)[8] for 27 of the 56 genes. Failure to find CpG islands in the putative upstream regions of the remaining genes could indicate either that a CpG-rich proximal promoter was not present, that a CpG island contained a control region located further upstream than available genomic data allowed us to analyze or that the region identified was not the true transcription start site.

Methylation analysis of 5′ CpG islands

We analyzed the methylation status of the CpG islands by using bisulfite-PCR in combination with methylated CpG site–specific restriction enzymes[9] and methylation-specific PCR (MSP)[10]. We compared the results with the gene expression status. All 12 of the group 1a genes (including 3 positive control genes) with identifiable 5′ CpG islands contained dense methylation of these regions in RKO cells (Fig. 2*a*,*b* and Table 1) and showed no basal expression detectable by RT–PCR (Fig. 1*b*). Of the five group 1b genes, for which we identified 5′ CpG islands, three showed partial methylation in these regions (Table 1), in agreement with their low basal expression levels (Fig. 1*b*). However, the other two genes did not show any methylation. None of the ten group 2 genes, whether they showed basal expression or not, showed any methylation of their 5′ CpG islands (Fig. 2*c*,*d* and Table 1).

Methylation and expression of group 1a genes

We first studied the methylation status and expression of group 1a genes in a series of eight CRC cell lines (Figs 2*b* and 3). We found hypermethylation of the genes *SFRP1*, *SEZ6L*, *PCDH8* and *FOLH1* in all CRC lines investigated. Of the eight lines, five showed total or predominant methylation of *KIAA0786*. The gene *CXX1* is of particular interest, because it is located on the X chromosome and is normally inactivated and methylated on one allele and active and unmethylated on the other in female cells. However, only methylated or predominantly methylated *CXX1* alleles were found in five of the eight CRC lines, including RKO, which, except for HT29, were all derived from males with cancer. *SNRPN* is also unusual in that it is known to be maternally imprinted in humans and to have a hypermethylated CpG island in the promoter region of the silenced allele. Thus, as expected, normal peripheral blood lymphocytes showed partial methylation of the CpG island around the transcription start site[11]. However, RKO, HCT116 and SW480 cells showed complete methylation and lack of basal expression

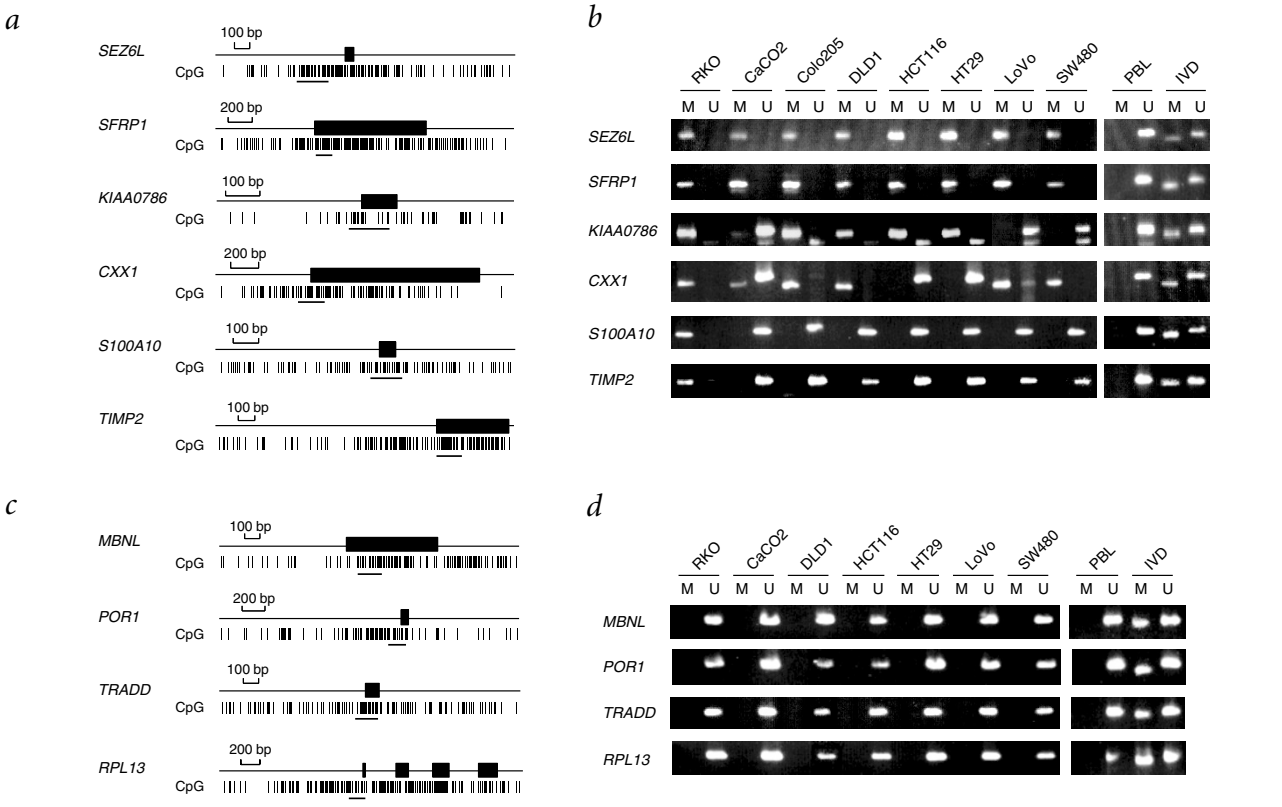

Fig. 2 Methylation analysis of group 1a and group 2 genes in a series of CRC cell lines. ***a***, Schematic representations of 5′ regions of group 1a genes. Black bars indicate first exons. Vertical bars indicate CpG sites. Regions analyzed by MSP are shown by black bars below the CpG sites. ***b***, Examples of MSP analysis in a series of CRC cell lines. Bands in 'M' lanes are PCR products for methylation-specific primers, and those in 'U' lanes are products with unmethylated-specific primers. Peripheral blood lymphocytes (PBL) and *in vitro* methylated DNA (IVD) serve as negative and positive controls, respectively. ***c***, Schematic representations of 5′ regions of group 2 genes. First exons, CpG sites and regions analyzed by MSP are indicated. ***d***, MSP analysis in a series of CRC cell lines.

(data not shown). *S100A10* and *TIMP2* methylation was seen only in RKO cells. Notably, each of the above genes lacked basal expression in methylated lines, which was restored by DAC (Fig. 3). We found, unexpectedly, that *KIAA0786* was not basally expressed in SW480 cells, despite lack of methylation, and yet was reactivated by DAC.

Methylation of group 1a genes in primary CRC tissues

We next investigated the methylation status of group 1a genes in primary colon cancers and normal colon tissues. We found a high frequency (17/20) of *SFRP1* methylation in primary CRC samples, but did not find methylation in 6 of 17 normal tissues from the same individuals with the tumors and from 3 individuals whose tumors showed no methylation (Fig. 4). In 11 individuals, *SFRP1* methylation occurred in both tumors and normal cells, but tumors showed stronger methylation signals (Fig. 4). We found no methylation of *SFRP1* in normal colon tissues from two individuals without CRC (data not shown).

The genes *SEZ6L* and *KIAA0786* also showed a high frequency of hypermethylation in primary CRC (13 of 20 and 8 of 20 affected individuals, respectively). These genes showed no methylation in normal colon of individuals whose tumors harbored no methylated genes, and some methylation in normal colon of individuals with tumors that contained methylated genes (2 of 13 and 4 of 8 individuals, respectively). The tumors showed stronger methylation signals than the normal tissues (Fig. 4). As expected, all tissue samples, including normal colon mucosa from females, showed partial methylation of *CXX1*. However, 3 of 14 males showed *CXX1* methylation in a tumor-specific manner (Fig. 4). We did not find methylation of *S100A10* and *TIMP2* in any primary CRC sample (Fig. 4). Notably, *FOLH1* and *PCDH8* were equally methylated in every CRC sample and normal counterpart (data not shown).

Methylation patterns of group 1a genes link CRC and gastric cancers

Our findings suggested that *SFRP1*, *SEZ6L*, *CXX1*, *KIAA0786*, *S100A10* and *TIMP2* might influence tumor development and progression. We therefore studied these genes in tumor cell lines of other types of cancer. A pattern of tumor profiling emerged: complete hypermethylation of *SFRP1*, *SEZ6L*, *LPPH1* and *CXX1* was common in CRC and gastric cancers, but only partial or no methylation was seen in all other types of cancer studied (Fig. 5). We found exceptions to this pattern for *SFRP1*. This gene has been shown to induce apoptosis in a breast cancer cell line, MCF7, which did not express the gene in the basal state[12]. We found complete methylation of the CpG island region in this cell line, as well as in MDA MB231 breast cancer cells and two of four prostate cancer cell lines studied (Fig. 5).

Methylation and expression of *SFRP* family members

So far, five *SFRP* genes have been identified. We found that four of these five genes have dense CpG islands around their first exons (Fig. 6*a*). The gene without a CpG island, *SFRP3*, was expressed basally in all seven CRC cell lines tested (data not shown). However, the other four *SFRP* genes are all hypermethylated with a high frequency in CRC cell lines; this hypermethylation is associated with a lack of basal expression, which is restored by DAC treatment (Fig. 6*b*,*c*).

Table 1 • Genes upregulated by DAC and TSA treatment in RKO cells

Group 1a

Acc no.[a]	Gene name	Symbol	Location	CpG island[b]	Methylation[c]
R80217	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)[d]	*PTGS2*[d]	1q25.2–q25.3	yes	yes
AA877595	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)[d]	*CDKN2A*[d]	9p21	yes	yes
AA099153	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)[d]	*TIMP3*[d]	22q12.3	yes	yes
AA444051	S100 calcium-binding protein A10	*S100A10*	1q21	yes	yes
N32514	secreted frizzled-related protein 1	*SFRP1*	8p12–p11.1	yes	yes
W72596	CAAX box 1	*CXX1*	Xq26	yes	yes
H29013	seizure-related gene 6 (mouse)-like	*SEZ6L*	22q11.2–12.1	yes	yes
W74533	latrophilin	*KIAA0786*	1p31.1	yes	yes
AA486280	tissue inhibitor of metalloproteinase 2	*TIMP2*	17q25	yes	yes
H29216	protocadherin 8	*PCDH8*	13q14.3–q21.1	yes	yes
N64840	folate hydrolase (prostate-specific membrane antigen) 1	*FOLH1*	11p11.2	yes	yes
AI017332	human SNRPN mRNA, 3′ UTR, partial sequence	*SNRPN*	15q12	yes	yes
N54793	pregnancy specific β-1-glycoprotein 6	*PSG6*	19q13.2	no	
H87471	kynureninase (L-kynurenine hydrolase)	*KYNU*	2p23.3–q14.3	no	
AA001432	laminin, α3 (nicein (150 kD), kalinin (165 kD),BM600 (150 kD), epilegrin)	*LAMA3*	18q11.2	no	
AA034939	laminin, alpha 2 (merosin, congenital muscular dystrophy, LAMA2)	*LAMA2*	6q22–q23	no	
AI298976	small inducible cytokine subfamily C, member 1 (lymphotactin)	*SCYC1*	1q21–q25	no	
AA291484	cytochrome P450, subfamily IVB, polypeptide 1	*CYP4B1*	1p34–p12	no	
R62603	Collagen, type VI, α3	*COL6A3*	2q37	no	
T73558	deoxyribonuclease I-like 3	*DNASE1L3*	3p21.1–3p14.3	no	
AA404246	*Homo sapiens* hypothetical protein MGC13047	10	no		
AA156424	EST				
H16554	EST				
N67972	EST				

Group 1b

Acc no.[a]	Gene name	Symbol	Location	CpG island[b]	Methylation[c]
AA173290	homeo box A1	*HOXA1*	7p15.3	yes	partial
AA935273	GRO3 oncogene	*GRO3*	4q21	yes	partial
AA256304	distal-less homeobox 7	*DLX7*	17q21.33	yes	partial
H17115	stromal antigen 3	*STAG3*	7	yes	no
AA454880	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA-binding protein 1, 37 kD)	*HNRPD*	4q21.1–q21.2	yes	no
AA496149	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	*HMGCS2*	1p13–p12	no	
AA176491	myogenic factor 6 (herculin)	*MYF6*	12q21	no	
H16793	chromosome 8 open reading frame 4	C8orf4	8p11.2	no	
H10079	KIAA0751 gene product		8	no	
H59614	similar to putative insulin-like growth factor II associated protein		11p15.5	uk	
AA457731	SNARE protein	*YKT6*	6	uk	
AA419251	interferon induced transmembrane protein 1 (9-27)	*IFITM1*	11	uk	
N48178	KIAA0403 protein		6	uk	
AA027147	hypothetical protein MGC3040		3	uk	
H18646	hypothetical protein FLJ10697		10	uk	
AA013268	*Homo sapiens* mRNA containing (CAG)4 repeat, clone CZ-CAG-7		UK	uk	
AA039857	EST				
AA101632	EST				
AA464518	EST				
AA427754	EST				
H16733	EST				
H88953	EST				
N90849	EST				
N22486	EST				
T62979	EST				
R53558	EST				
R39555	EST				

Group 2

Acc no.[a]	Gene name	Symbol	Location	CpG island[b]	Methylation[c]
AA425908	partner of RAC1 (arfaptin 2)	*POR1*	11p15	yes	no
AA405717	muscleblind (*Drosophila melanogaster*)-like	*MBNL*	3	yes	no
AA916906	TNFRSF1A-associated via death domain	*TRADD*	16q22	yes	no
AA404394	for protein disulfide isomerase-related	*PDIP*	3	yes	no
AA489678	RAD23 (*S. cerevisiae*) homolog B	*RAD23B*	3p25.1	yes	no
AA447514	ribosomal protein L13	*RPL13*	16q24.3	yes	no
AA071330	guanine nucleotide binding protein (G protein), α-inhibiting activity polypeptide 2	*GNAI2*	3p21	yes	no
AA669126	protein phosphatase 1, regulatory (inhibitor) subunit 12A	*PPP1R21A*	12q15–q21	yes	no
R38619	fucose-1-phosphate guanyltransferase	*FPGT*	1	yes	no
AA055503	tripartite motif-containing 32	*TRIM32*	9q32–q34.11	yes	no
T66981	egf-like module containing mucin-like, hormone receptor-like sequence 1	*EMR1*	19p13.3	no	
AA480906	protein kinase C binding protein 1	*PRKCBP1*	20q12	no	
N45318	phosphoglycerate mutase 2 (muscle)	*PGAM2*	7p13–p12	no	
N30096	glutathione S-transferase A3	*GSTA3*	6p12	no	
AA427733	advillin	*ADVIL*	12	no	
N92901	fatty acid binding protein 4, adipocyte	*FABP4*	8q21	no	
T60149	hypothetical protein FLJ13449		13	uk	
AA453578	human DNA sequence from clone RP11-3J10 on chromosome 9–12-13.3		9p12–p13.3	uk	
W81520	*Homo sapiens* gene from PAC 106H8, similar to Dynamin		1	uk	
AA446486	EST				
AA447992	EST				
H94605	EST				
W46439	EST				

[a]GenBank accession number. [b]Yes: CpG island was found around presumed transcription start site or near upstream region; no: no CpG island was found around presumed transcription start site or near upstream region; uk: upstream genomic sequence is unknown. [c]Yes: fully methylated; partial: partially methylated; no: no methylation. [d]Positive control genes.

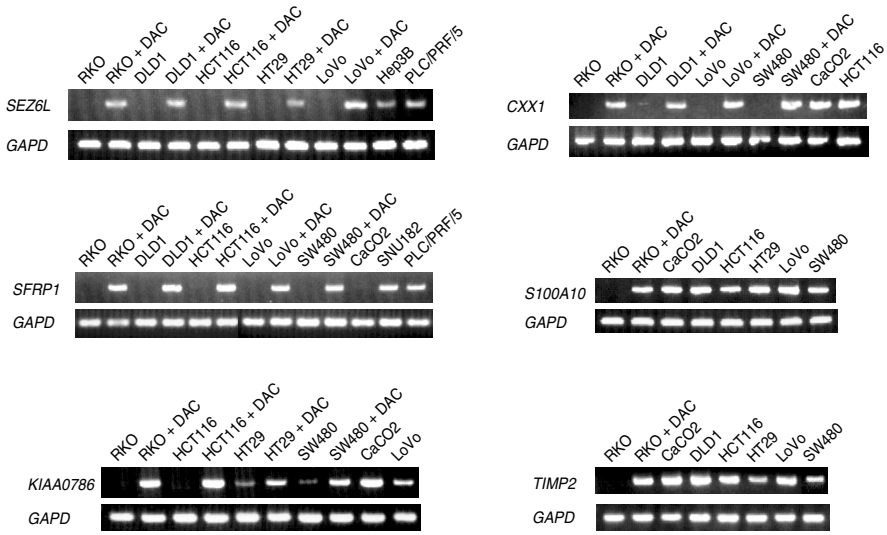

Fig. 3 RT–PCR analysis in various CRC cell lines. For *SEZ6L* and *SFRP1*, hepatocellular carcinoma cell lines without methylation (Hep3B, SNU182 and PLC/PRF/5) are also shown, as examples of cells where these genes are unmethylated. The gene *GAPD* serves as a positive indicator for RNA quality and loading. 'DAC' indicates cell lines treated with 5 μM DAC.

We analyzed methylation of these genes in primary CRC tissues (n = 124). The genes are not hypermethylated in normal colon, except for trace methylation of *SFRP2* in an individual who has a colon cancer in which this gene is hypermethylated. In addition, normal colon and cell lines derived from other tissues express the genes in the absence of promoter methylation (Fig. 6*c*). However, we found hypermethylation of all three genes in primary CRC tumors (Fig. 6*d*). The frequencies differ in this large analysis, which includes expanded data for *SFRP1* (*SFRP1*, 118/124, 95.1%; *SFRP2*, 111/124, 89.5%; *SFRP4*, 36/124, 29.0%; and *SFRP5*, 73/124, 58.9%). Notably, 24.1% of affected individuals (30/124) show methylation of all four *SFRP* genes with CpG islands, and at least one of the four is methylated in 123 of 124 (99.2 %) of the tumors (Fig. 7).

These data show that logical mining of the initial microarray data can markedly extend the gene discovery consequences. The findings also reveal the involvement in epigenetic silencing of a gene family which, in CRC, could abrogate a block to WNT oncogene activity. To our knowledge, this hypermethylation of a single gene family provides the highest molecular marker coverage yet for a common human cancer[13].

Discussion

Our previous studies suggest that the transcriptional silencing of hypermethylated genes in cancer cells depends on synergy between the methylation and the activity of HDACs, with methylation having the dominant effect[6]. Our findings seem to validate this concept concerning the nature of chromatin associated with such genes and to provide a strategy with high efficiency for identifying genes with high potential for a role in tumorigenesis.

From the standpoint of transcriptionally repressive chromatin, the strategy we have used has provided information about the promoters of genes with various responses to inhibitors. Our findings with respect to group 1a genes confirm that densely methylated genes will not re-express if exposed to HDAC inhibition alone. In contrast, the results for group 2 genes indicate that those genes that do re-express or have upregulated expression after HDAC inhibition alone lack promoter methylation, even when CpG islands are present in their 5′ regions. Our study also identifies genes that were upregulated after treatment of cells with DAC, despite the fact that their promoters seemed to be unmethylated. Similar findings have recently been reported[14]. Perhaps methylation of upstream genes, such as those encoding transcription factors, could secondarily result in activation of these genes. Another possibility, however, is that inhibitors of DNA methyltransferases, such as DAC, might affect these proteins in ways other than simply blocking their methylating capacities. Recent studies showed that DNA methyltransferases have the potential to repress transcription independently of their methylating activities[15–19], both directly and through interaction with HDACs and other corepressor proteins.

That our approach begins with established cell lines could lead to bias toward detection of genes that are either altered only in culture but not in primary tumors, or for which promoter hypermethylation is not tumor-specific. Excessive CpG island hypermethylation in cancer cell lines, compared with primary cancer tissues, has been described[20]. However, analysis of paired primary tumors and normal tissues suggests that our method efficiently identifies genes for which altered expression is associated with hypermethylated 5′ CpG islands in primary as well as cultured cells. Seven of the 12 genes detected by the microarray approach, *CDKN2A*, *COX2*, *TIMP3*, *SEZ6L*, *SFRP1*, *KIAA0786* and *CXX1* were methylated specifically in primary tumors or only in regions of normal colon from individuals with CRC who

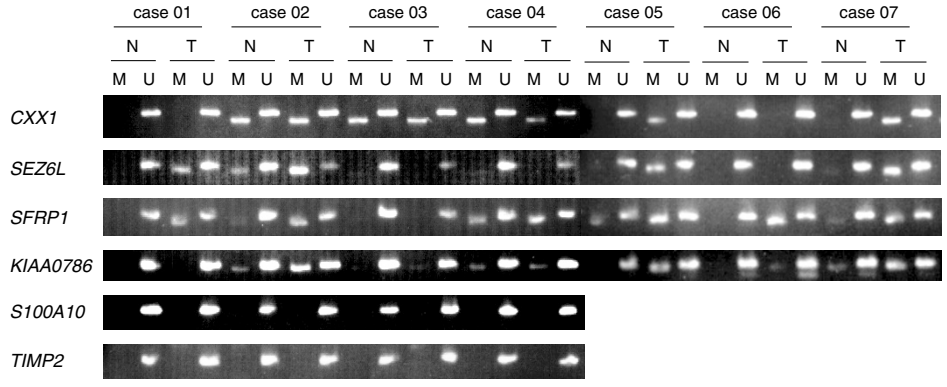

Fig. 4 MSP analysis of group 1a genes in primary CRC tissues (T) and normal colon mucosa (N) from the same affected individuals. Cases 1, 5, 6 and 7 are tumors from affected males; the others are from females. Bands in 'M' and 'U' lanes are as indicated in Figs 2 and 3.

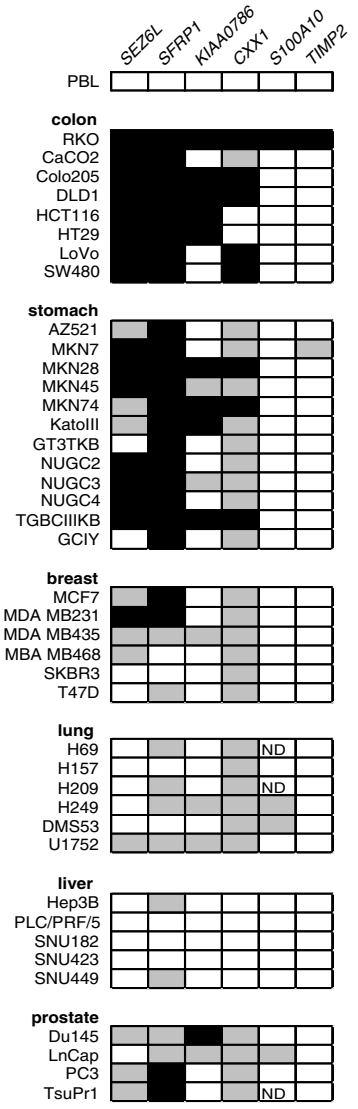

Fig. 5 Summary of MSP analyses of six genes from group 1a in a series of human cancer cell lines of various origins. Gene names are indicated at the top, and cell line names are indicated on the left. Closed boxes indicate full methylation; gray boxes and open boxes indicate partial methylation and no methylation, respectively. ND, not determined (because of lack of amplification in MSP).

showed methylation of those genes. Another gene, *TIMP2*, although not methylated in normal colon, primary CRCs or peripheral blood lymphocytes, has subsequently proven to be very frequently hypermethylated in malignant lymphomas (data not shown). In addition, *SNRPN*, an imprinted gene, was methylated in the promoter of the silenced allele. Two genes were methylated in both normal colon and primary CRCs. Only *S100A10* has not been found to be methylated in primary tissues in our study so far, but our analysis is not extensive and this gene has been reported to be downregulated in prostate cancer[21].

Our microarray-based approach identified a large number of genes that are hypermethylated in a tumor-specific manner. Some genes, such as *SFRP1*, were found to be methylated in some, but not all, normal colon mucosa tissues from individuals with, but not without, CRC. This methylation of genes in normal tissues could reflect a 'field effect', in which pre-malignant changes are found over a broad region of the colon, or could indicate a tendency for certain CpG islands to become methylated with age in normal colon, as has been found for a group of genes frequently hypermethylated in CRC[22]. The field effect seems more likely, as the age of individuals with no methylation in normal tissues ranged from 53 to 64 years, and one 46-year-old individual showed methylation in both normal and tumor tissues.

One advantage of this approach is that most of the genes we identified have known properties or implied functions that are involved in tumorigenesis. Most of the group 1a genes, and

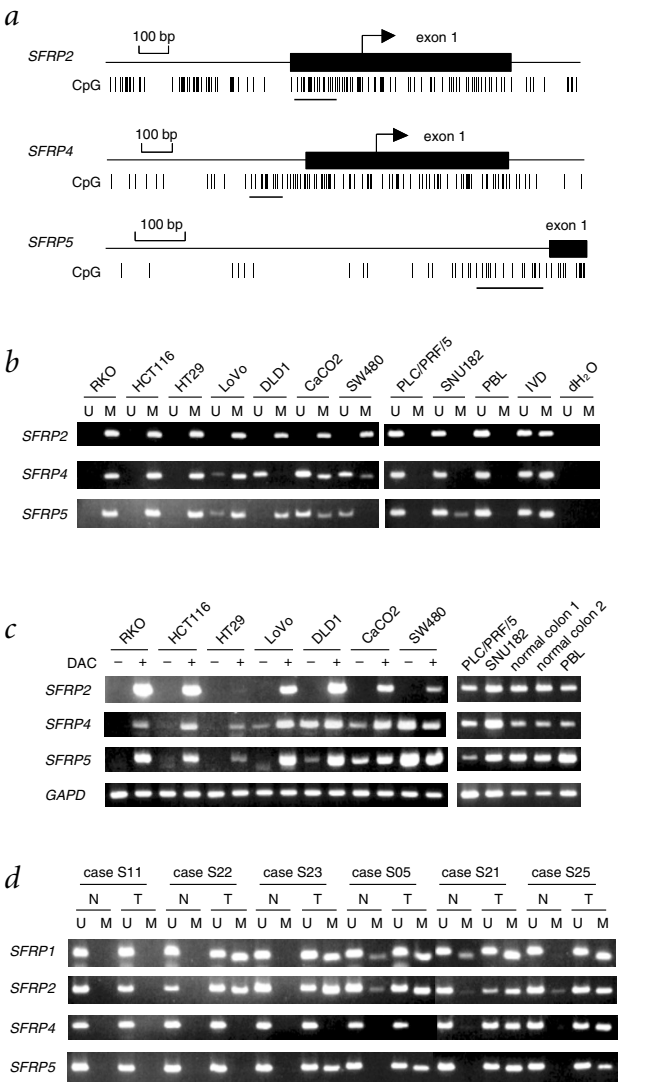

Fig. 6 Methylation and expression analysis of the *SFRP* genes. ***a***, Schematic representations of 5′ regions of *SFRP2*, *4* and *5*. The large black bars indicate first exons. Vertical bars indicate CpG sites. Regions analyzed by MSP are shown as black bars below the CpG sites. ***b***, Examples of MSP analysis in a series of cell lines. Bands in 'U' and 'M' lanes are as indicated in Figs 2 and 3. MSP analyses for cell lines PCL/PRF/5 and SNU182, and of normal peripheral blood lymphocytes (PBL), are shown as controls for unmethylated promoters; the *in vitro* methylated DNA IVD reactions are shown as positive controls for partially methylated promoters. ***c***, RT–PCR analysis of the *SFRP* genes. PLC/PRF/5, SNU182, two normal colon mucosa samples and PBL are shown as positive controls for expression of each gene in a setting of a non-methylated promoter region. 'DAC' indicates PCR for cell lines treated with 5 μM DAC. ***d***, MSP analysis in primary CRC tissues (T) and normal colon mucosa (N) from the same affected individuals.

many genes in the other groups, are located in chromosomal regions known to undergo frequent loss of heterozygosity in cancers, including *SFRP1* at chromosome 8p12, *SEZ6L* at 22q11 and *TIMP2* at 17q25 (Table 1). In addition, many of the genes identified encode components of pathways that have been implicated in cancer. For example, among the group 1a genes, *SFRP1* antagonizes Wnt oncogene signaling[23], and breast cancer cells transfected with *SFRP1* showed increased sensitivity to pro-apoptotic stimuli[12]. Underexpression of *SFRP1* has been reported in the majority of breast carcinomas[24,25]. Expression of mouse *SEZ6* and rat latrophilin is limited to brain; however, their human homologs (*SEZ6L* and *KIAA0786*) were identified in frequently deleted regions in lung and breast cancers, respectively, although their functions in humans remain unclear[26,27]. *TIMP2* is a member of the tissue inhibitor of matrix metalloprotainase (TIMP) family; another member, *TIMP3*, is known to be frequently inactivated by hypermethylation in various malignancies[28]. The S100A10 protein, also termed annexin II light chain or p11, forms a heterotetrameric complex with another calcium-binding protein, annexin II heavy chain (p36)[29]. A recent study reported frequent loss of p36 and p11 protein expression in prostate cancers and suggested that methylation could be responsible for the silencing of *p36* (ref. 21). CXX1 is a putative prenylated protein, but its function remains unclear[30]. *SNRPN* is located on 15q11-q13, a region that is implicated in Prader-Willi syndrome and Angelman syndrome, and is thought to be involved in pre-mRNA splicing[31].

The genes *FOLH1* and *PCDH8* also have notable characteristics. Folate metabolism affects DNA methylation, and recent reports suggested that one of the folate metabolic enzymes, methylenetetrahydrofolate reductase, may affect susceptibility to human malignancies[32,33]. The FOLH1 protein is involved in folate uptake and may have a role in DNA methylation in cancers[34]. The gene *PCDH8* encodes a member of a cell–cell adhesion molecule family[35] for which loss of function is known to be important in invasion and metastasis. However, these two genes did not show tumor-specific or tumor-predominant methylation. The FOLH1 protein was originally characterized as a prostate-specific membrane antigen, is strongly expressed in prostate cancers[34] and has not been studied in colorectal tumors. In normal tissues, *PCDH8* is expressed exclusively in fetal and adult brain[35]. Thus, methylation of *FOLH1* and *PCDH8* might be a tissue-specific phenomenon that seems to be related to gene expression, as these genes are silent in CRC cell lines and treatment of these cells with DAC leads to re-expression.

One unexpected finding is the frequent hypermethylation of many genes involved in gastrointestinal tumors—in particular, hypermethylation of the *SFRP* gene family. This finding raises several possibilities. First, a common defect in chromatin constitution may bias several genes, and particularly a family of related genes, toward epigenetic silencing in association with promoter hypermethylation. Second, all of the *SFRP* genes are thought to counter WNT/frizzled signaling[23,36–38]; loss of function of these genes could represent abrogation of an entire tumor-suppressor pathway. For example, mutations of *APC* are common in colon cancer and could lead to constitutive WNT pathway action[39,40]. Our initial data indicate that *APC* mutations are frequent throughout CRC tumors with all combinations of hypermethylation of the *SFRP* genes (data not shown). However, APC has additional postulated functions[41]; thus, loss of inhibition of WNT activity through other mechanisms indicates a new functional pathway important to colorectal tumorigenesis.

Our approach could allow the identification of the entire spectrum of genes silenced by epigenetic mechanisms in cancers. Our finding that the methylation patterns for the newly identified genes map with the specific cancer type initially screened and a related tumor type (Fig. 5) underlines the importance of promoter hypermethylation for profiling of human cancers. Notably, CRC and gastric tumors have another previously demonstrated hypermethylation pattern. These are among the few kinds of tumors that show the microsatellite instability phenotype due to loss of mismatch repair function; in each case, the link is a hypermethylation event involving the promoter of *MLH1* (ref. 1). Thus, panels of such markers have potential for facilitating basic studies of the pathways that regulate tumorigenesis. Moreover, the mechanisms involved in linking the hypermethylated genes that we have identified in a single colon cancer line to gastrointestinal tumors are important to explain at a biological level. Our findings further demonstrate how a limited number of hypermethylated genes may be used to compose comprehensive marker panels for sensitive detection of specific types of human cancer, and suggest that our technique may be especially suited for identifying such gene panels.

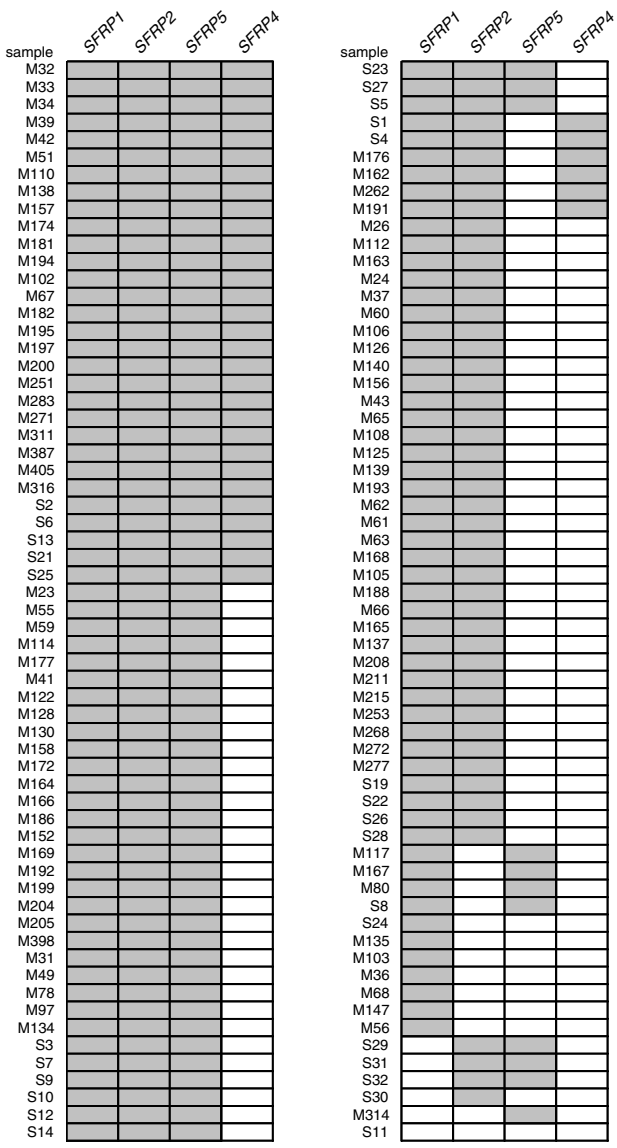

Fig. 7 Summary of MSP analyses of the *SFRP* genes in 124 primary CRC samples. Gene names are indicated at the top. Each row represents a primary CRC tumor. Gray boxes and open boxes indicate methylation and no methylation, respectively.

Methods

Cell culture and tissue samples. We cultured cell lines in RPMI 1640 or Minimal Essential Medium (GIBCO BRL) supplemented with 10% fetal bovine serum. Tissue samples of colorectal cancer and normal colon mucosa were all from specimens obtained at the time of clinically indicated surgical procedures.

DAC and TSA treatment and RNA preparation. We treated RKO cells with DAC (Sigma) and TSA (Wako) as described previously[6]. Briefly, the treatment consisted of DAC (200 nM) for 48 h, with drug and medium replaced 24 h after the beginning of treatment, followed by addition of TSA, to a final concentration of 300 nM from a 1.5-mM ethanol-dissolved stock, for a further 24 h. We also treated cells with mock, DAC alone and TSA alone by using the same volumes of PBS and/or ethanol, and/or same amount of the drugs. We also treated some CRC cell lines for RT-PCR analysis, to assess more robust levels of gene expression, with 5 μM of DAC for 72 h, replacing drug and medium every 24 h. We extracted total RNA used for microarray analysis, cDNA subtraction and RT–PCR by using the TRIZOL Reagent (Gibco/BRL).

cDNA subtraction. Before cDNA subtraction, we isolated poly(A)$^+$ RNA from total RNA by using the Message Maker Reagent Assembly kit (Gibco/BRL). We carried out cDNA subtraction with a combination-treated RKO cell line as the tester, and mock treated cells as the driver, by using the PCR-Select cDNA Subtraction Kit (Clontech). We digested synthesized cDNA with *Rsa*I and ligated tester cDNA to adaptors included in the kit. After hybridization, we amplified the subtracted cDNA by using the Advantage cDNA PCR kit (Clontech).

Microarray analysis. We carried out microarray analysis using the Mammalian GeneFilters Microarrays system (Research Genetics). We produced filters for approximately 5,000 of the genes analyzed in the Johns Hopkins Comprehensive microarray core, and we also purchased filters for another 5,000 genes (Human GeneFilters Microarrays Release II) from Research Genetics. We analyzed a total of 10,814 genes and ESTs. We hybridized the filters according to the manufacturer's recommendation. Briefly, we reverse-transcribed and labeled 5 μg of total RNA using oligo $(dT)_{12–18}$ primer and [^{33}P]dCTP with Superscript II reverse transcriptase (Gibco/BRL). We hybridized the filters for 12–18 h. We analyzed the data thus obtained by using the PSCAN program (National Institutes of Health). For subtraction–microarray analysis, we labeled the second PCR product from cDNA subtraction by using the Multiprime DNA Labeling System (Amersham) with ^{33}P. We carried out hybridization and data analysis as described above. We repeated microarray analysis independently at least three times for each condition and compared results for probing the arrays with cDNA for total RNA from mock-treated cells to those for hybridizations with subtraction PCR products.

Semi-quantitative RT–PCR. We reverse-transcribed total RNA (2 μg) treated with DNase I (Ambion) for single-stranded cDNA using oligo $(dT)_{12–18}$ primer with Superscript II reverse transcriptase (Gibco/BRL). We carried out PCR reactions in a volume of 50 μl containing 1× PCR buffer (Gibco/BRL), 1.5 mM of $MgCl_2$, 0.3 mM of dNTP, 0.25 μM of each primer and 2 U of *Taq* polymerase (Gibco/BRL). We used 100 ng of cDNA for PCR amplification, and we amplified all of the genes with multiple cycle numbers (20 cycles to 35 cycles) to determine the appropriate conditions for obtaining semi-quantitative differences in their expression levels. We carried out the RT–PCR analyses in Fig. 3 with 35 cycles. We also carried out PCR with *GAPD* (25 and 28 cycles) to ensure cDNA quality and loading accuracy. Primer sequences are available upon request.

Methylation analysis. We carried out bisulfite modification of genomic DNA as described previously[10]. We studied methylation status with PCR analysis of bisulfite-modified genomic DNA, using two procedures. First, all genes investigated were analyzed by bisulfite-PCR, followed by digestion with several methylated CpG site-specific restriction enzymes (COBRA), as described[9]. The second analysis used MSP for all genes analyzed in several cancer cell lines and tissue samples, as described[10]. We designed all of the bisulfite PCR and MSP primers according to genomic sequences around presumed transcription start sites of investigated genes. Primer sequences and PCR conditions for methylation analysis are available upon request.

Methylation and expression analysis of the *SFRP* genes. We carried out methylation analysis of *SFRP2* and *SFRP4* by using three different MSP primer pairs to cover the 5′ CpG islands of each gene. For *SFRP5* methylation analysis, we used two different MSP primer pairs. For *SFRP2* RT–PCR, we designed sense and antisense primers to amplify exons 2 and 3, respectively. We designed RT–PCR primers for *SFRP4* to amplify exons 2 and 5. RT–PCR primers for *SFRP5* were designed to amplify exons 2 and 3. For each gene, we used the MSP primer pair that best assessed the methylation status of the gene with respect to the expression data in cell lines in Fig. 6; we also used these primers for analysis of primary CRC tissues. Primer sequences for MSP and RT–PCR are available upon request.

Acknowledgments

We thank K. Bachman, K. Schuebel and H. Yoshikawa for providing cell lines; Y. Akiyama, M. Esteller and O. Galm for helpful comments on the manuscript; M. Toyota, M. Kusano, F. Itoh and K. Imai for providing CRC samples and M. House for providing normal colon samples.These studies were supported by grants from the National Cancer Institute.

Competing interests statement

The authors declare competing financial interests: see the Nature Genetics website (http://genetics.nature.com) for details.

Received 21 March; accepted 16 April 2002.

1. Baylin, S.B. & Herman, J.G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16**, 168–174 (2000).
2. Makos-Wales, M. *et al.* p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3. *Nature Med.* **6**, 570–577 (1995).
3. Toyota, M. *et al.* Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* **59**, 2307–2312 (1999).
4. Costello, J.F. *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature Genet.* **24**, 132–138 (2000).
5. Gonzalgo, M.L. *et al.* Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res.* **57**, 594–599 (1997).
6. Cameron, E.E., Bachman, K.E., Myohanen, S., Herman, J.G. & Baylin, S.B. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature Genet.* **21**, 103–107 (1999).
7. Toyota, M. *et al.* Aberrant methylation of the cyclooxygenase 2 CpG island in colorectal tumors. *Cancer Res.* **60**, 4044–4048 (2000).
8. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **20**, 261–282 (1987).
9. Xiong, Z. & Laird, P.W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25**, 2532–2534 (1997).
10. Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. & Baylin, S.B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA* **93**, 9821–9826 (1996).
11. Sutcliffe, J.S. *et al.* Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region. *Nature Genet.* **8**, 52–58 (1994).
12. Melkonyan, H.S. *et al.* SARPs: a family of secreted apoptosis-related proteins. *Proc. Natl Acad. Sci. USA* **94**, 13636–13641 (1997).
13. Esteller, M., Corn, P.G., Baylin, S.B. & Herman, J.G. A gene hypermethylation profile of human cancer. *Cancer Res.* **61**, 3225–3229 (2001).
14. Soengas, M.S. *et al.* Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature* **409**, 207–211 (2001).
15. Rountree, M.R., Bachman, K.E. & Baylin, S.B. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genet.* **25**, 269–277 (2000).
16. Bachman, K.E., Rountree, M.R. & Baylin, S.B. Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* **276**, 32282–32287 (2001).
17. Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature Genet.* **24**, 88–91 (2000).
18. Fuks, F., Burgers, W.A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **20**, 2536–2544 (2001).
19. Robertson, K.D. *et al.* DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nature Genet.* **25**, 338–342 (2000).
20. Smiraglia, D.J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10**, 1413–1419 (2001).
21. Chetcuti, A. *et al.* Loss of annexin II heavy and light chains in prostate cancer and its precursors. *Cancer Res.* **61**, 6331–6334 (2001).
22. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA* **96**, 8681–8686 (1999).
23. Finch, P.W. *et al.* Purification and molecular cloning of a secreted, Frizzled-related antagonist of Wnt action. *Proc. Natl Acad. Sci. USA* **94**, 6770–6775 (1997).
24. Ugolini, F. *et al.* Differential expression assay of chromosome arm 8p genes identifies Frizzled-related (FRP1/FRZB) and fibroblast growth factor receptor 1 (FGFR1) as candidate breast cancer genes. *Oncogene* **18**, 1903–1910 (1999).

25. Ugolini, F. *et al.* WNT pathway and mammary carcinogenesis: loss of expression of candidate tumor suppressor gene SFRP1 in most invasive carcinomas except of the medullary type. *Oncogene* **20**, 5810–5817 (2001).
26. Nishioka, M. *et al.* Identification of a 428-kb homozygously deleted region disrupting the *SEZ6L* gene at 22q12.1 in a lung cancer cell line. *Oncogene* **19**, 6251–6260 (2000).
27. White, G.R., Varley, J.M. & Heighway, J. Isolation and characterization of a human homologue of the latrophilin gene from a region of 1p31.1 implicated in breast cancer. *Oncogene* **17**, 3513–3519 (1998).
28. Bachman, K.E. *et al.* Methylation-associated silencing of the tissue inhibitor of metalloproteinase-3 gene suggest a suppressor role in kidney, brain, and other human cancers. *Cancer Res.* **59**, 798–802 (1999).
29. Kube, E., Weber, K. & Gerke, V. Primary structure of human, chicken, and *Xenopus laevis* p11, a cellular ligand of the Src-kinase substrate, annexin II. *Gene* **102**, 255–259 (1991).
30. Frattini, A., Faranda, S., Zucchi, I. & Vezzoni, P. A low-copy repeat in Xq26 represents a novel putatively prenylated protein gene (*CXX1*) and its pseudogenes (*DXS9914*, *DXS9915*, and *DXS9916*). *Genomics* **46**, 167–169 (1997).
31. Nicholls, R.D., Saitoh, S. & Horsthemke, B. Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.* **14**, 194–200 (1998).
32. Matsuo, K. *et al.* Association between polymorphisms of folate- and methionine-metabolizing enzymes and susceptibility to malignant lymphoma. *Blood* **97**, 3205–3209 (2001).
33. Song, C., Xing, D., Tan, W., Wei, Q. & Lin, D. Methylenetetrahydrofolate reductase polymorphisms increase risk of esophageal squamous cell carcinoma in a Chinese population. *Cancer Res.* **61**, 3272–3275 (2001).
34. Heston, W.D. Characterization and glutamyl preferring carboxypeptidase function of prostate specific membrane antigen: a novel folate hydrolase. *Urology* **49** (Suppl.), 104–112 (1997).
35. Strehl, S., Glatt, K., Liu, Q.M., Glatt, H. & Lalande, M. Characterization of two novel protocadherins (PCDH8 and PCDH9) localized on human chromosome 13 and mouse chromosome 14. *Genomics* **53**, 81–89 (1998).
36. Rattner, A. *et al.* A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *Proc. Natl Acad. Sci. USA* **94**, 2859–2863 (1997).
37. Chang, J.T. *et al.* Cloning and characterization of a secreted frizzled-related protein that is expressed by the retinal pigment epithelium. *Hum. Mol. Genet.* **8**, 575–583 (1999).
38. Abu-Jawdeh, G. *et al.* Differential expression of frpHE: a novel human stromal protein of the secreted frizzled gene family, during the endometrial cycle and malignancy. *Lab. Invest.* **79**, 439–447 (1999).
39. Morin, P.J. *et al.* Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC. *Science* **275**, 1787–1790 (1997).
40. Behrens, J. *et al.* Functional interaction of an axin homolog, conductin, with β-catenin, APC, and GSK3β. *Science* **280**, 596–599 (1998).
41. Mimori-Kiyosue, Y. & Tsukita, S. Where is APC going? *J.Cell Biol.* **154**, 1105–1109 (2001).