Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores

Rafael Irizarry[1,2,*,†], Christine Ladd-Acosta[2,3,*], Bo Wen[2,3], Zhijin Wu[6], Carolina Montano[2,3], Patrick Onyango[2,3], Hengmi Cui[2,3], Kevin Gabo[2,3], Michael Rongione[2,3], Maree Webster[7], Hong Ji[2,3], James Potash[2,4], Sarven Sabunciyan[2,5] and Andrew P. Feinberg[2,3,*,†]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA; [2]Center for Epigenetics, Institute for Basic Biomedical Sciences, and Departments of [3]Medicine, [4]Psychiatry, and [5]Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [6]Center for Statistical Sciences, Brown University, Providence, RI; [7]Stanley Laboratory of Brain Research, Uniform Services University of Health Sciences, Bethesda, MD 20892, USA

[*]Equal contribution from these authors

[†]To whom correspondence should be addressed. Email: rafa@jhu.edu and afeinberg@jhu.edu

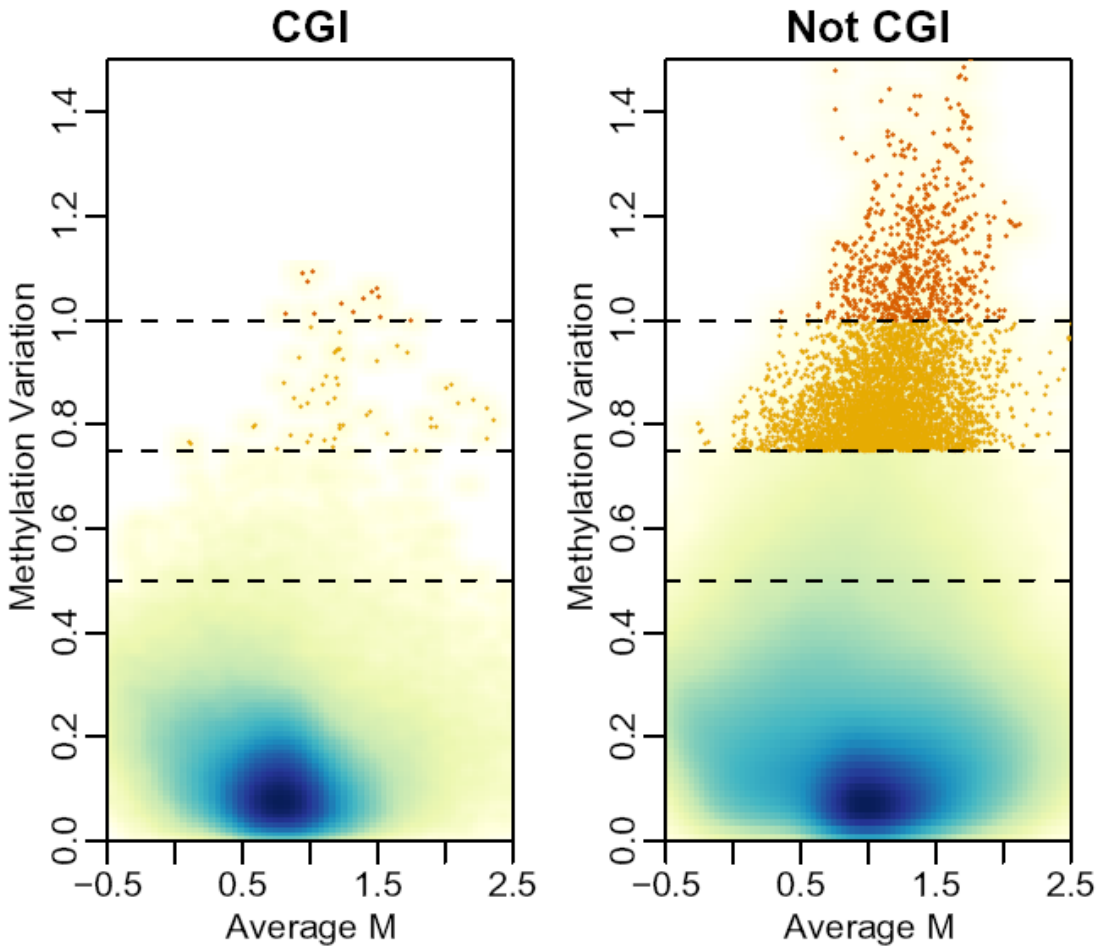
Supplementary Figures

This document contains the following Supplementary Figures:

Supplementary Figure 1
Supplementary Figure 3
Supplementary Figure 5
Supplementary Figure 6
Supplementary Figure 7
Supplementary Figure 8
Supplementary Figure 9

Supplementary Figures 2 and 4 have been uploaded as separate files (Supplementary Figure 2.pdf and Supplementary Figure 4.pdf) as Supplementary information because of their size, each containing 50 images.

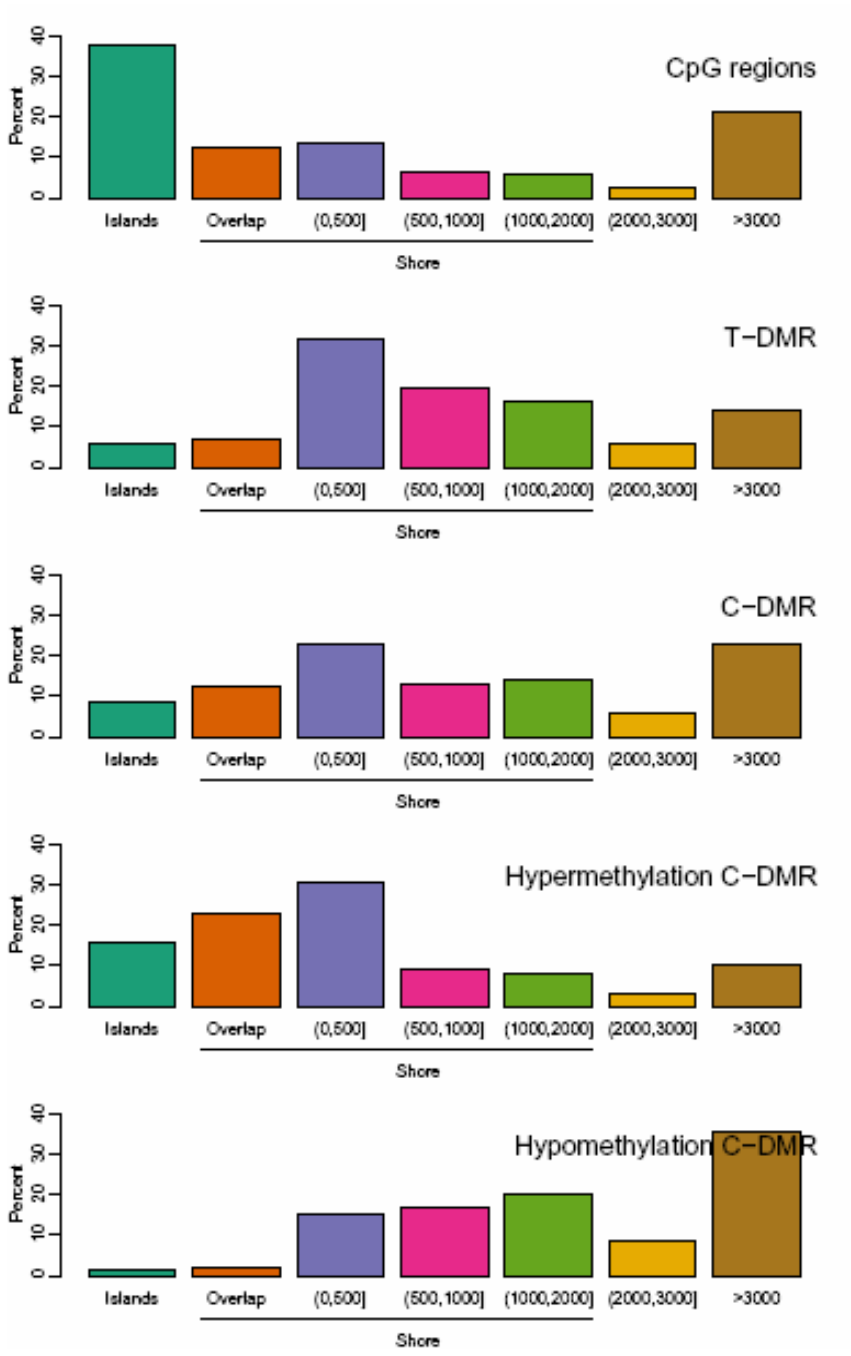
Supplementary Figure 1



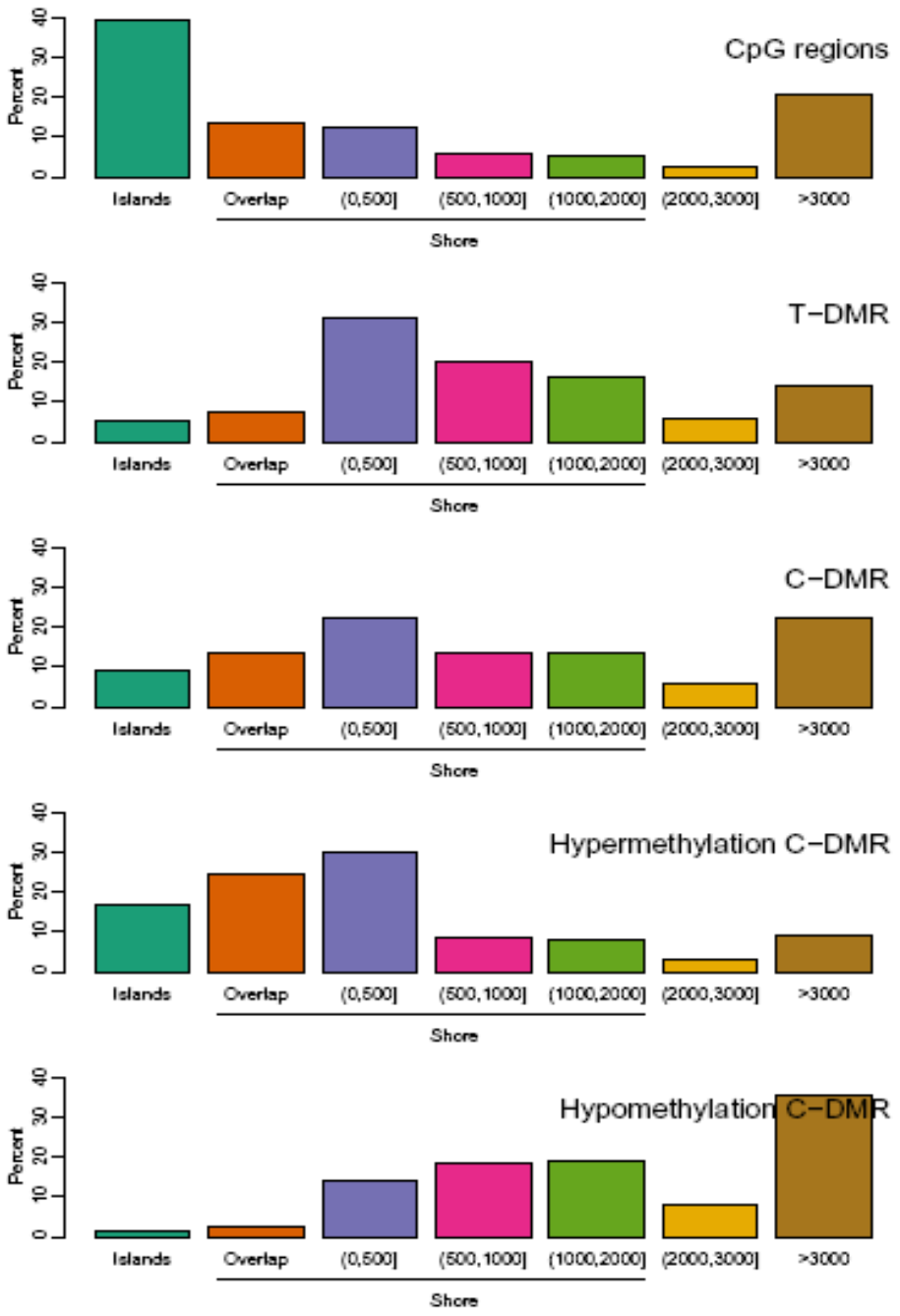
Genome-wide DNAm analysis shows methylation variation is more common outside of CpG islands than within them. We computed the average M value for four sets of normal brain, liver, spleen and colon. Methylation variation across tissues was determined using the standard deviation across tissues of within tissue-averaged M. For all CHARM probes located within a CGI, left plot, and all CHARM probes located outside of a CGI, right plot, we used a smooth scatter plot to represent distribution of values from high (blue) to low (yellow) number of points. Where tissue M variation was greater than 0.75 we used yellow points, and where greater than 1.0, orange points.

Supplementary Figure 3

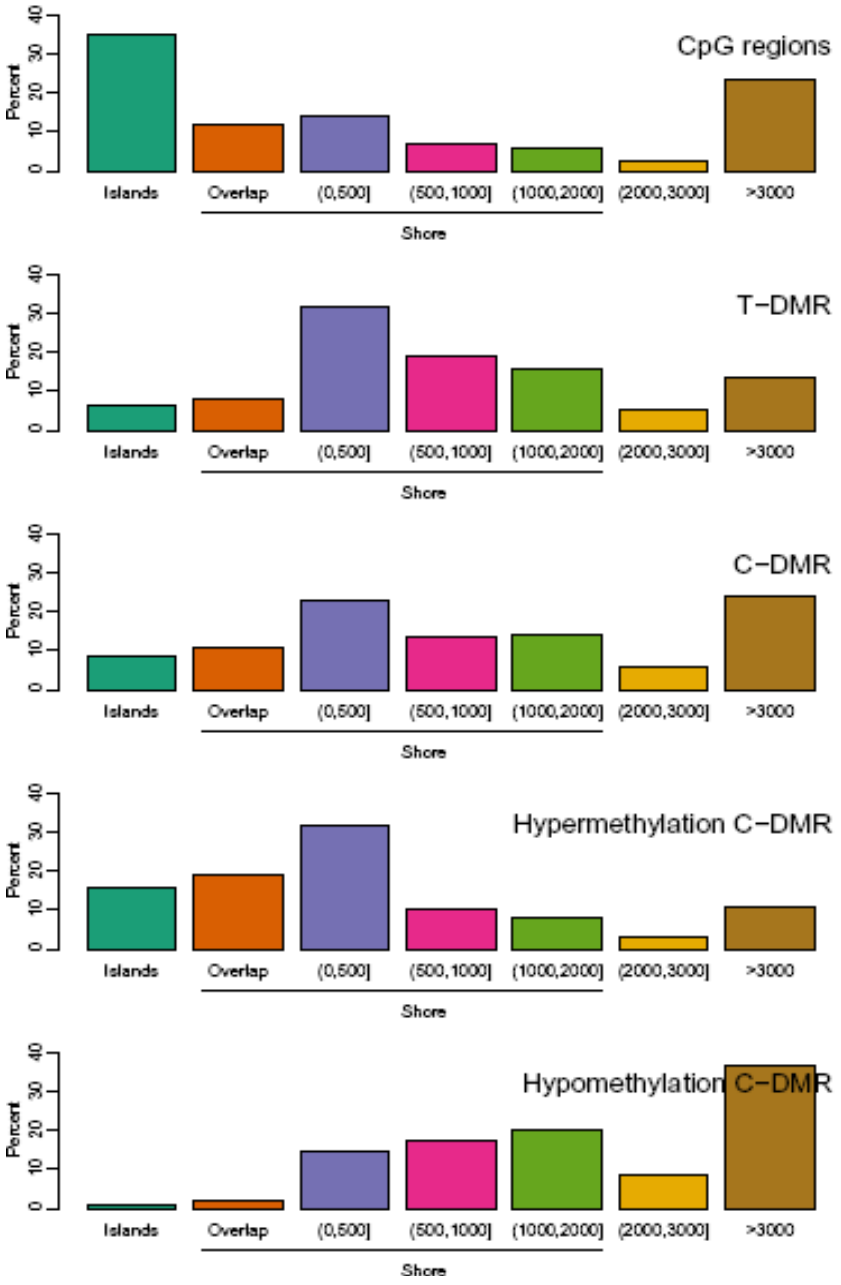
a



b

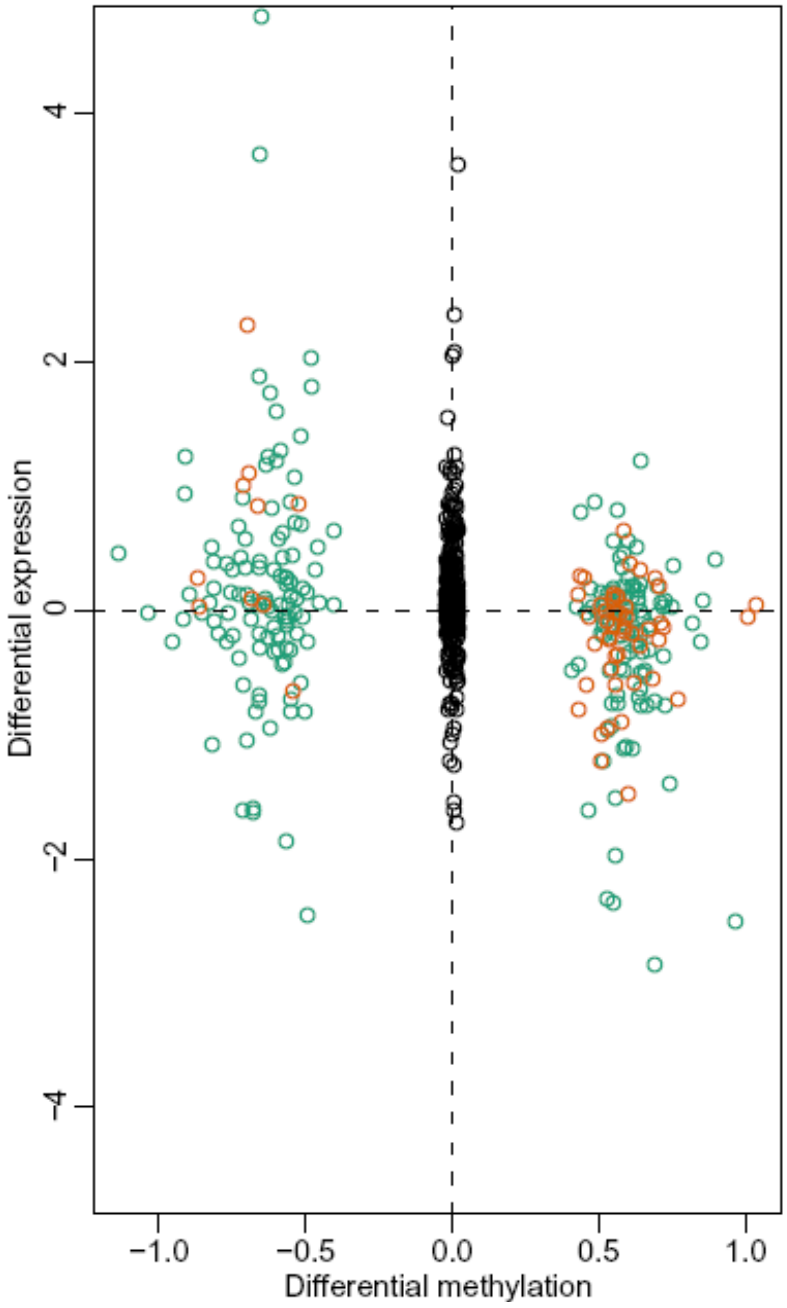

CpG regions
T−DMR
C−DMR
Hypermethylation C−DMR
Hypomethylation C−DMR
Percent
0
10
20
30
40
Islands
Overlap
(0,500]
(500,1000]
(1000,2000]
(2000,3000]
>3000
Shore

c



Different FDR cutoffs do not substantively alter the distribution of distance of T-DMRs and C-DMRs from CpG islands. (**a**) using an FDR cutoff of 0.01, (**b**) using an FDR cutoff of 0.05 (**c**) using an FDR cutoff of 0.10**.** *Islands* (teal) are regions which cover or overlap more than 50% of a CpG island. *Overlap* (orange) are regions which overlap 0.1-50% of a CpG island. *(0,500]* (purple) are regions which do not overlap islands but are located ≤ 500 bp of islands. *(500,1000]* (magenta) are regions located > 500 and ≤ 1000 bp from an island. *(1000,2000]* (green) are regions > 1000 bp and ≤ 2000 bp from an island. *(2000,3000]* (yellow) are regions located > 2000 bp and ≤ 3000 bp from an island. Greater than 3000 (brown) are regions > 3000 bp from an island. The percentage of each class is provided for CpG regions (the CHARM arrays themselves, null hypothesis), tissue-specific differentially methylated regions (T-DMRs), cancer-specific differentially methylated regions (C-DMRs), and the latter subdivided into regions of cancer-specific hypermethylation and hypomethylation.

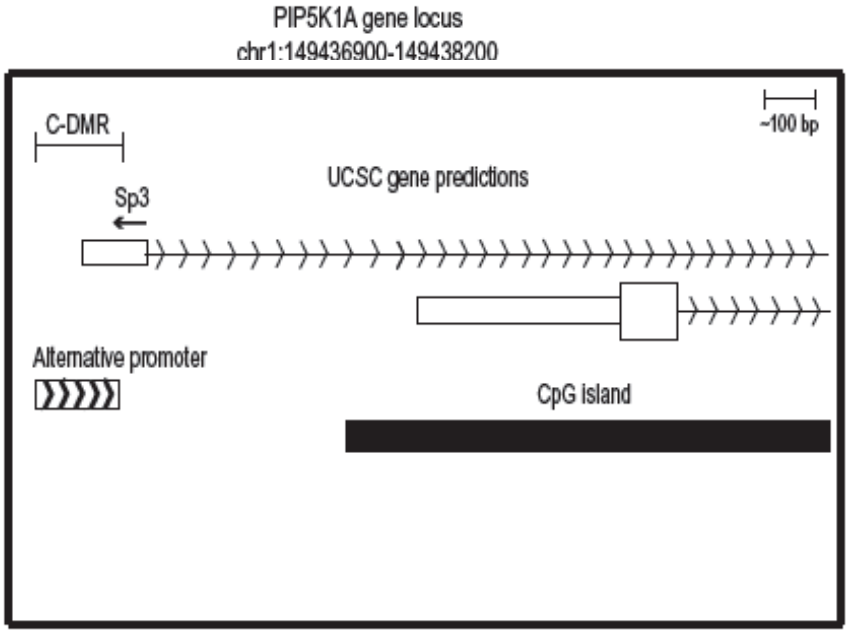
Supplementary Figure 5



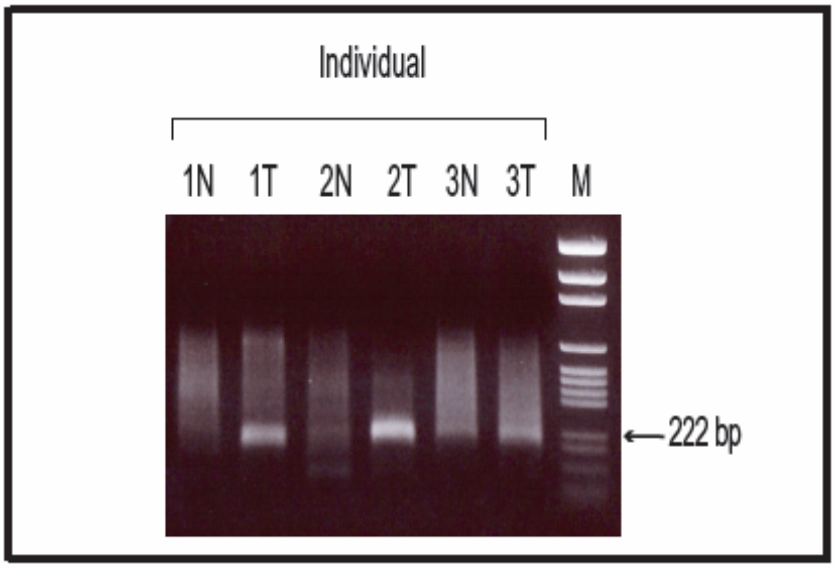
Gene expression is strongly correlated with C-DMRs at CpG island shores. For each colon tumor versus normal mucosa C-DMR we found the closest annotated gene on the Affymetrix HGU133A microarray, resulting in a total of 650 gene/C-DMR pairs. Plotted are log (base 2) ratios of colon tumor to normal expression against delta M values for colon tumor and normal DNAm. Orange dots represent C-DMRs located within 300 bp from the corresponding gene's transcriptional start site (TSS). Green dots represent C-DMRs that are located from 300-2000 bp from the TSS of an annotated gene. Black dots, in the middle, represent log ratios for all genes further than 2 kb from an annotated TSS.

Supplementary Figure 6

a

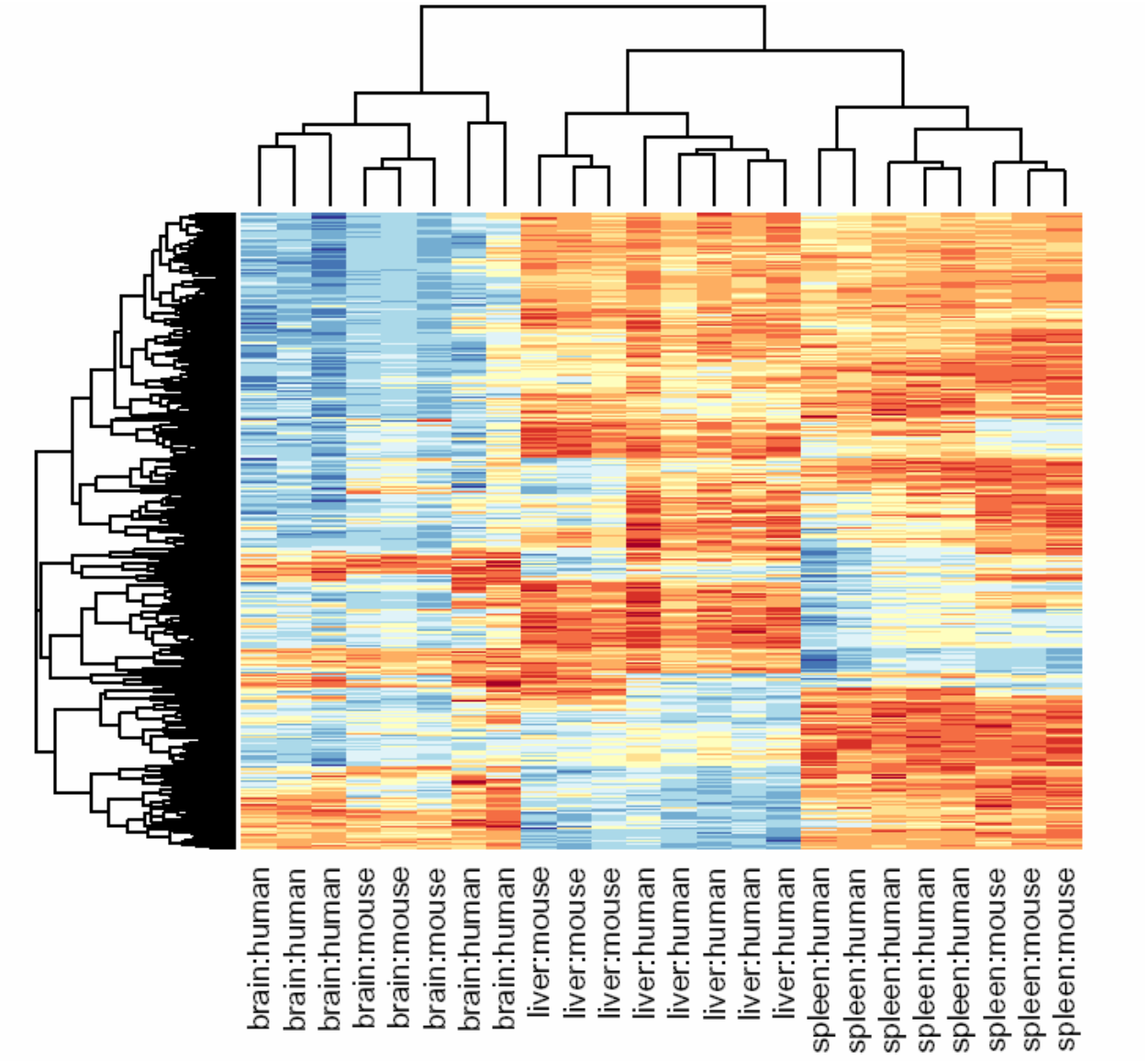


b



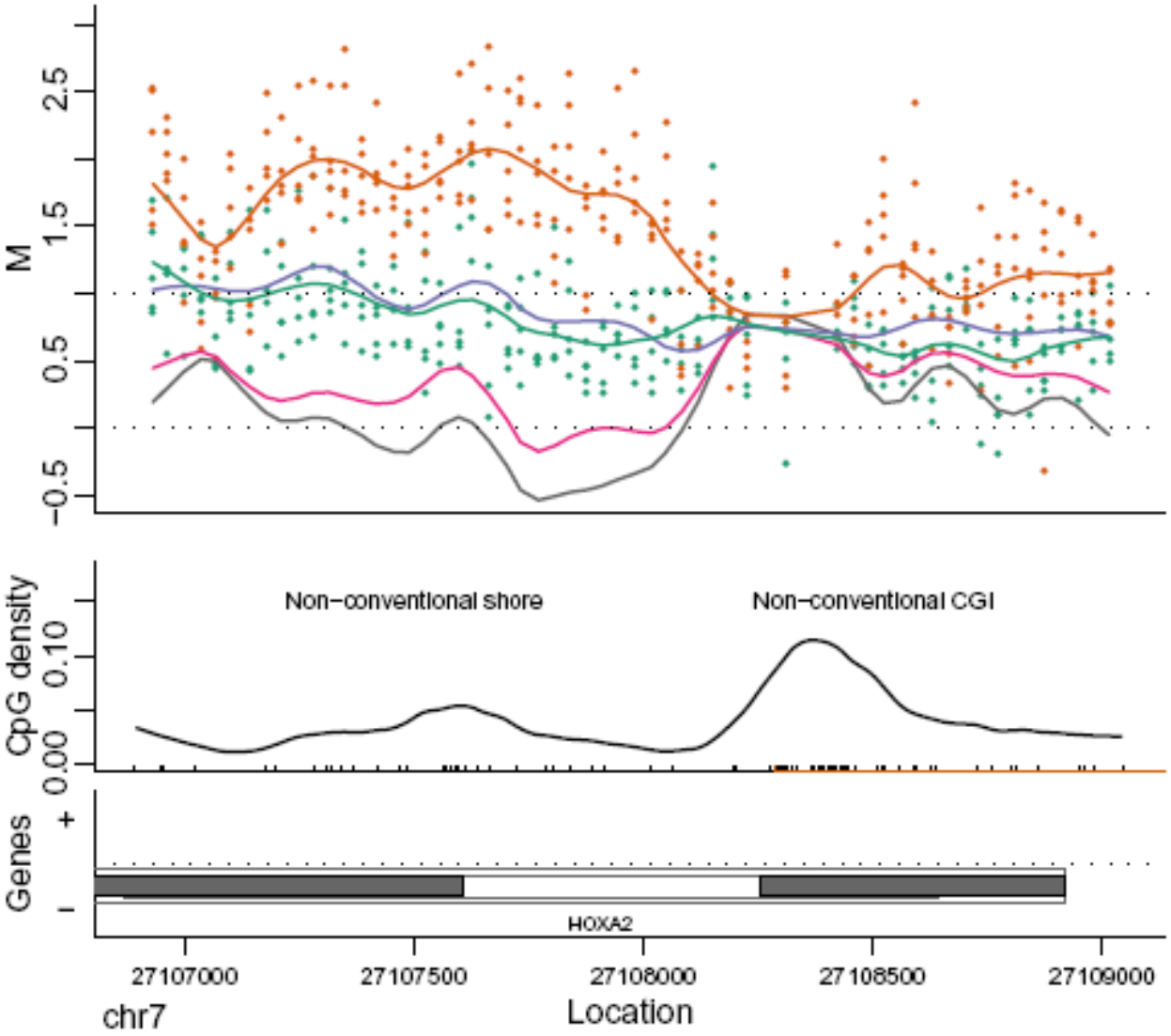
Alternate gene transcription origin in colon tumors at a CpG island shore. (**a**) schematic overview of the *PIP5K1A* gene locus. C-DMR represents the genomic region that is hypomethylated in colon tumors as compared to normals. (**b**) the arrow indicates a 220bp fragment present in 3 colon tumors but not normal mucosa from the same individual amplified by the Sp3 primer and the 5' RACE anchor primer.

Supplementary Figure 7



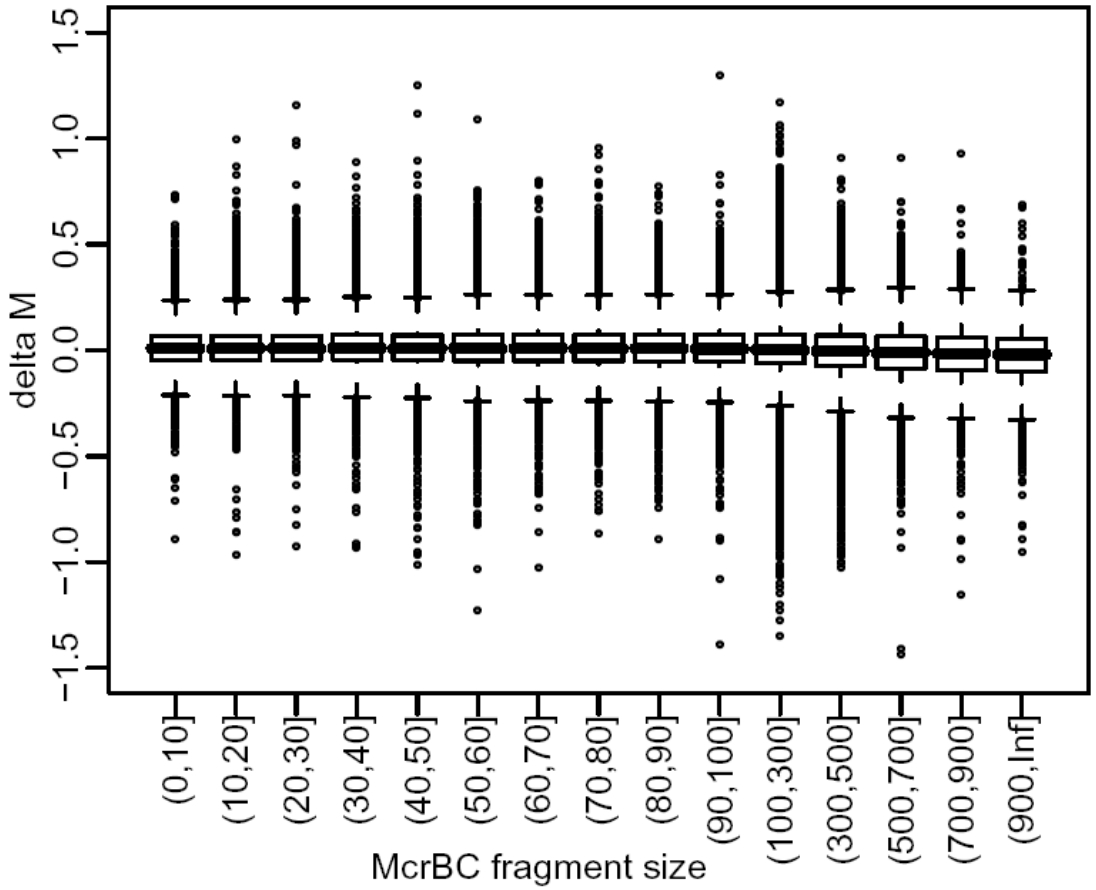
Methylation of mouse T-DMRs completely discriminates tissues regardless of species of origin, even for T-DMRs > 2 kb from an annotated gene. The *M* values of all tissues from the 957 regions corresponding to mouse T-DMRs that mapped to the human genome and were greater than 2 kb from an annotated gene were used for unsupervised hierarchical clustering. By definition, the mouse tissues are segregated. Surprisingly, all of the human tissues are also completely discriminated by the regions that differ in mouse tissues. The three major branches in the dendrograms correspond perfectly to tissue type regardless of species. Columns represent individual samples, and rows represent regions corresponding to mouse T-DMRs. The heatmap displays *M* values, with red being more methylated and blue less.

Supplementary Figure 8



Example of non-conventional CpG island. A C-DMR that is not adjacent to a conventionally defined CpG island. Upper panel: plot of M value versus genomic location for brain (grey), liver (pink), spleen (purple), normal colon (green), colon tumor (orange). Each point represents the methylation level of an individual sample for each probe across the region. The smoothed line represents averaged, smoothed M values for each tissue, described in detail in the methods. M values which range from -0.5 to 0.5 represent unmethylated points and values from 0.5 to 1.5 represent baseline levels of methylation due to the scale and standardization used. The middle panel provides the location of CpG dinucleotides with black tick marks on the x-axis. CpG density was calculated across this region using a standard density estimator and is represented by the smoothed black line. The location of non-conventional CpG islands, defined using Hidden Markov modeling, is denoted on the x-axis as an orange line. The lower panel provides gene annotation for the genomic region. The thin outer grey line represents the start of transcription, the thin inner lines represent the start and end of a coding region. Filled in grey boxes represent exons. On the y-axis, plus and minus marks denote sense and antisense gene transcription respectively.

Supplementary Figure 9



Fragment size has no effect on ΔM. For each probe we computed the expected DNA fragment size based on McrBC recognition sites. We then stratified the delta M values for each probe from the colon tumor and normal mucosa comparison by fragment size. The box-plots show delta M values.

Supplementary Tables

This document contains the following Supplementary Tables:

Supplementary Table 1. Bisulfite pyrosequencing confirms differential DNA methylation at CpG island shores but not the associated island.

Gene	Location[a]	Region	Tissue[b]	CG1	CG2	CG3	CG4	CG5	CG6	CG7	CG8	CG9	CG10	CG11	CG12	CG13	CG14	CG15	CG16	CG17	CG18
PCDH9	+3,338	Shore	Brain	32	26	12	39	19	22												
			Spleen	91	71	31	76	66	60												
			P value	<.001	<.001	<.001	<.001	<.001	0.003												
	-267	Island	Brain	2	3	4	2	3	5	2	3	2	3	3	3	5	5	4	3	3	4
			Spleen	2	3	3	2	3	6	2	4	3	3	3	3	3	5	4	3	4	3
			P value	0.032	0.298	0.336	0.108	0.475	0.150	0.393	0.141	0.011	0.661	0.265	0.208	0.420	0.051	0.133	0.885	0.783	0.270
HEY1	+3,381	Shore	Brain	54	53	51	51														
			Liver	70	84	87	71														
			P value	.023	<.001	<.001	<.007														
	+2,207	Island	Brain	4	7	3	4	4	5	1	8	5	5	7	7	4					
			Liver	3	6	3	4	4	6	2	9	23	26	26	8	8					
			P value	0.349	0.309	0.226	0.460	0.630	0.252	0.017	0.336	0.255	0.179	0.238	0.432	0.001					
HAGH	+2,192	Shore	Liver	26	30	22	18	7	6	23	33										
			Spleen	93	93	82	56	20	20	86	95										
			P value	<.001	<.001	<.001	<.001	<.001	<.001	0.017	0.017										
	+206	Island	Liver	2	1	2	3	2	7	3	2	3	1	2	2	1	1	1	2	1	4
			Spleen	2	2	3	4	2	2	2	2	4	1	4	2	2	3	2	4	1	8
			P value	0.608	0.207	0.433	0.803	0.058	0.342	0.262	0.529	0.504	0.782	0.060	0.832	0.366	0.074	0.307	0.073	0.141	0.015
SLMO2	+1,125	Shore	Normal	89	63	85	46	68	30	78	81	75	82	40	85	81	43	65	76	76	87
			Tumor	37	28	34	19	30	13	34	40	35	36	18	38	36	19	30	39	37	46
			P value	<.001	<.001	<.001	0.005	0.002	<.001	<.001	<.001	0.002	<.001	0.036	<.001	<.001	<.001	<.001	<.001	0.003	<.001
	+40	Island	Normal	4	2	3	3	6	4	3	2	3	2	3	3	4	2	7	5		
			Tumor	4	1	3	3	3	4	3	2	2	4	3	2	4	2	7	6		
			P value	0.619	0.233	0.293	0.546	0.302	0.364	0.461	0.204	0.586	0.263	0.173	0.369	0.253	0.928	0.230	0.509		

[a] Location represents distance in base pairs, + denoting upstream and – downstream, from the transcriptional start site to the closest CpG site measured by bisulfite pyrosequencing (CG1). CG1-18 denote individual CpG site measured by bisulfite pyrosequencing. Values are percent methylation. The coordinates for each CpG site measured by pyrosequencing are provided in Supplementary Table 7.

[b] Brain, spleen, and liver tissues are from the same individuals. Normal and tumor represent matched colon tumor and mucosa from the same individuals.

Supplementary Table 2. Bisulfite pyrosequencing confirms differential DNA methylation at 4 additional CpG island shores

Gene	Location[a]	Region	Tissue[b]	CG1	CG2	CG3	CG4	CG5	CG6	CG7	CG8	CG9	CG10	CG11	CG12
SEMA3C	-409	Shore	Tumor	31	32	32	34	31	26	24	21	23	9	9	1
			Normal	65	71	68	64	63	52	58	41	43	21	17	2
			P value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.007	<0.001	0.001	0.003	<0.001	0.048
DSCAML1	+2,875	Shore	Brain	34	30	29	28	33	26	33	26	38	23	26	
			Liver	86	79	87	89	76	71	88	82	98	70	77	
			P value	<0.001	<0.001	<0.001	<0.001	<0.001	0.004	<0.001	<0.001	<0.001	<0.001	<0.001	
GPT2	-2,984	Shore	Spleen	93	93	82	56	20	20	86	95				
			Liver	26	30	22	18	7	5	23	33				
			Pvalue	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001				
ZNF532	+1,127	Shore	Liver	97	96	92	93	79	96	98	89				
			Brain	34	38	26	30	23	24	22	21				
			Pvalue	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001				

[a] Location is distance in base pairs, + denoting upstream and − downstream, from the transcriptional start site to closest CpG site measured by bisulfite pyrosequencing (CG1). CG1-12 denote individual CpG sites measured by bisulfite pyrosequencing. Values are percent methylation. The coordinates for each site are provided in Supplementary Table 7.

[b] Brain, spleen, and liver tissues are from the same individuals. Normal and tumor represent matched colon tumor and mucosa from the same individuals.

Supplementary Table 3. Bisulfite pyrosequencing confirms differential DNA methylation at 9 C-DMRs.

C-DMR	Gene	CHARM methylation[a]	Pyro normal[b]	Pyro cancer[b]	Number normal[c]	Number cancer[c]	Number matched[d]	*P*
Chr7: 96491680-96494485	*DLX5*	C > N	8%	22%	53	61	42	$<10^{-6}$
Chr14: 41147653-41148722	*LRFN5*	C > N	23%	44%	61	65	55	$<10^{-14}$
Chr7: 27129188-27131713	*HOXA3*	C > N	52%	72%	52	59	44	$<10^{-16}$
Chr13: 83352750-83353823	*SLITRK1*	C > N	36%	46%	56	51	26	0.0009
Chr3: 62335415-62336514	*FEZF2*	C > N	16%	40%	42	48	39	$<10^{-14}$
Chr6: 52637426-52638797	*TMEM14A*	N > C	67%	45%	57	64	53	$<10^{-11}$
Chr8: 846979-849372	*ERICH1*	N > C	71%	56%	30	34	23	$<10^{-3}$
Chr13: 113615559-113616275	*FAM70B*	N > C	67%	50%	38	36	29	$<10^{-3}$
Chr20: 55705356-55707713	*TMEPAI*	N > C	29%	17%	48	41	30	$<10^{-6}$

[a] Relative methylation reported by CHARM (C, cancer; N, normal)
[b] Mean methylation level reported by bisulfite pyrosequencing
[c] Number of samples analyzed by bisulfite pyrosequencing
[d] Number of matched sample pairs from the same individuals

Supplementary Table 4. Relationship between DNA methylation and gene expression for 2 T-DMRs and 6 C-DMRs

Gene	Methylation[a]	Region	Type	Distance[b]	Distance[c]	Fold change[d]	*P*	Expression[e]
FZD3	Brain < Liver	Upstream Shore	T-DMR	844 bp	7 bp	28.2	0.002	Brain > Liver
	Brain < Spleen			844 bp	7 bp	34.0	<0.001	Brain > Spleen
RBM38	Brain > Spleen	Intragenic Shore	T-DMR	1937 bp	69 bp	7.0	0.005	Brain < Spleen
	Liver > Spleen			1937 bp	69 bp	4.0	0.136	Liver < Spleen
NDN	Tumor > Normal	Promoter Shore	C-DMR	47 bp	0 bp	4.2	0.025	Normal > Tumor
TRAF1	Tumor > Normal	Intragenic Shore	C-DMR	724 bp	0 bp	2.8	0.025	Normal > Tumor
ZNF804A	Tumor > Normal	Promoter Shore	C-DMR	105 bp	185 bp	2.5	0.047	Normal > Tumor
CHRM2	Tumor > Normal	Promoter Shore	C-DMR	433 bp	0 bp	2.0	0.346	Normal > Tumor
NQO1	Normal > Tumor	Promoter Shore	C-DMR	146 bp	216 bp	2.8	0.004	Tumor > Normal
SEMA3C	Normal > Tumor	Promoter Shore	C-DMR	143 bp	0 bp	4.8	0.025	Tumor > Normal

[a] Methylation level reported by CHARM from greatest to least.
[b] Base pairs to canonical transcriptional start site from the DMR.
[c] Base pairs to an alternative transcriptional start site from the DMR.
[d] Fold change = $2^{-\Delta\Delta CT}$; $\Delta\Delta C_T$ is equal to $(C_{T\ \text{tissue A target gene}} - C_{T\ \text{beta actin}}) - (C_{T\ \text{tissue B target gene}} - C_{T\ \text{beta actin}})$.
[e] Tissue expression from greatest to least.
P was computed using a paired, two-tailed, t-test.

Supplementary Table 5. Distribution of C-DMRs and T-DMRs.

		Total number	Percent
Cancer DMRs		2,707	100%
Colon DMRs			
	Tumor hypermethylated	1,508	56%
	Tumor hypomethylated	1,199	44%
Tissue DMRs		16,379	100%
Brain DMRs		8200	
	Brain hypermethylated	1717	21%
	Brain hypomethylated	6483	79%
Liver DMRs		3511	
	Liver hypermethylated	1763	50%
	Liver hypomethylated	1748	50%
Spleen DMRs		3186	
	Spleen hypermethylated	1208	38%
	Spleen hypomethylated	1978	62%
Brain:Liver:Spleen DMRs		1482	

Brain:Liver:Spleen DMRs are regions that vary in methylation across all three tissues.

Supplementary Table 7. Gene ontology functional categories enriched in hyper- and hypomethyated C-DMRs (P < 0.01)

C-DMR	GO BP ID	P value	Odds Ratio	ExpCount	Count	Size	Term	C-DMR Region
Hypermethylated	GO:0065007	1.75E-15	1.43	1049.09	1221	3840	biological regulation	Inside
	GO:0006355	2.40E-13	1.50	504.3282	632	1846	regulation of transcription, DNA-dependent	Inside/promoter
	GO:0022008	1.49E-12	2.56	66.11453	117	242	neurogenesis	Inside/promoter
	GO:0032774	2.03E-12	1.47	517.4418	641	1894	RNA biosynthetic process	Inside/promoter
	GO:0019219	1.13E-11	1.44	545.5814	667	1997	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	Inside/promoter
	GO:0006350	2.42E-11	1.43	552.1383	672	2021	transcription	Inside/promoter
	GO:0010468	9.44E-11	1.41	569.6231	687	2085	regulation of gene expression	inside
	GO:0007399	3.96E-10	2.24	69.90378	116	263	nervous system development	inside
	GO:0019222	7.12E-10	1.37	622.0776	737	2277	regulation of metabolic process	inside
	GO:0007275	2.63E-09	1.91	100.0965	152	395	multicellular organismal development	inside/promoter
	GO:0050794	2.78E-09	1.37	612.7722	721	2297	regulation of cellular process	inside
	GO:0007155	3.54E-09	1.74	143.9767	204	527	cell adhesion	inside
	GO:0048667	1.03E-08	3.00	30.59846	59	112	neuron morphogenesis during differentiation	inside/promoter
	GO:0031175	2.06E-08	2.79	33.87686	63	124	neurite development	inside/promoter
	GO:0006366	5.56E-07	1.57	154.0851	206	564	transcription from RNA polymerase II promoter	inside
	GO:0048598	6.02E-07	3.03	22.67564	44	83	embryonic morphogenesis	inside/promoter
	GO:0006813	7.63E-07	2.39	36.60887	63	134	potassium ion transport	inside
	GO:0048731	2.51E-06	1.51	163.7593	214	629	system development	inside
	GO:0007268	4.80E-06	2.06	46.00111	73	169	synaptic transmission	inside
	GO:0001501	6.53E-06	2.07	43.98528	70	161	skeletal development	inside
	GO:0007166	7.82E-06	1.38	258.195	316	947	cell surface receptor linked signal transduction	inside/promoter
	GO:0048663	8.67E-06	16.03	3.824807	12	14	neuron fate commitment	inside/promoter
	GO:0009887	1.12E-05	1.73	77.15602	110	283	organ morphogenesis	inside
	GO:0048858	1.20E-05	1.98	47.81009	74	175	cell projection morphogenesis	inside/promoter
	GO:0007411	1.37E-05	3.38	14.20643	29	52	axon guidance	inside
	GO:0032989	1.71E-05	1.81	62.01652	91	227	cellular structure morphogenesis	inside
	GO:0007417	1.77E-05	4.03	4.162579	15	210	central nervous system development	promoter/inside
	GO:0045165	1.78E-05	5.12	7.884457	19	29	cell fate commitment	inside

	GO:0050877	3.16E-05	1.46	153.5387	196	562	neurological system process	inside
	GO:0030182	4.77E-05	4.13	3.500752	13	180	neuron differentiation	promoter/inside
	GO:0003002	6.81E-05	3.01	14.43907	28	53	regionalization	inside
	GO:0030198	7.50E-05	4.23	8.469216	19	31	extracellular matrix organization and biogenesis	inside
	GO:0007165	0.000112	1.27	370.8697	428	1393	signal transduction	inside
	GO:0007498	0.000119	3.46	10.65482	22	39	mesoderm development	inside
	GO:0006811	0.000179	1.76	50.34717	73	186	ion transport	inside/promoter
	GO:0016481	0.000212	1.62	69.11973	95	253	negative regulation of transcription	inside
	GO:0008286	0.000212	4.13	7.649615	17	28	insulin receptor signaling pathway	inside
	GO:0007223	0.00022	5.79	5.19081	13	19	Wnt receptor signaling pathway, calcium modulating pathway	inside
	GO:0048518	0.000234	1.31	236.0452	281	864	positive regulation of biological process	inside
	GO:0007156	0.00028	1.97	32.23766	50	118	homophilic cell adhesion	inside
Hypermethylated	GO:0007409	0.000289	2.70	14.67244	27	54	axonogenesis	inside/promoter
	GO:0006928	0.000332	1.54	81.96016	109	300	cell motility	inside
	GO:0000122	0.000357	1.95	32.51086	50	119	negative regulation of transcription from RNA polymerase II promoter	inside
	GO:0051253	0.000364	1.72	49.72249	71	182	negative regulation of RNA metabolic process	inside
	GO:0048869	0.000389	1.88	26.48193	44	1336	cellular developmental process	promoter
	GO:0030901	0.000399	30.11	0.158574	3	8	midbrain development	promoter/inside
	GO:0019933	0.000447	2.43	17.21163	30	63	cAMP-mediated signaling	inside
	GO:0048523	0.000503	1.38	142.3193	176	525	negative regulation of cellular process	inside
	GO:0009790	0.000527	3.01	10.78958	21	40	embryonic development	inside
	GO:0007169	0.000542	3.86	2.834518	10	143	transmembrane receptor protein tyrosine kinase signaling pathway	promoter/inside
	GO:0045941	0.000581	1.57	68.30013	92	250	positive regulation of transcription	inside
	GO:0035107	0.000665	3.17	9.562018	19	35	appendage morphogenesis	inside/promoter
	GO:0060173	0.000665	3.17	9.562018	19	35	limb development	inside/promoter
	GO:0035295	0.00079	2.00	26.22725	41	96	tube development	inside
	GO:0009792	0.000941	2.04	24.04165	38	88	embryonic development ending in birth or egg hatching	inside
	GO:0006817	0.001021	2.34	16.39203	28	60	phosphate transport	inside
	GO:0035270	0.001048	3.65	7.092567	15	26	endocrine system development	inside
	GO:0001709	0.001066	3.64	7.103214	15	26	cell fate determination	inside/promoter

	GO:0030326	0.001093	3.24	8.469216	17	31	embryonic limb morphogenesis	inside/promoter
	GO:0043583	0.00125	2.81	10.65482	20	39	ear development	inside
	GO:0051254	0.001323	1.58	56.00611	76	205	positive regulation of RNA metabolic process	inside
	GO:0001649	0.00135	3.74	6.556812	14	24	osteoblast differentiation	inside
	GO:0016477	0.001648	3.09	3.84543	11	194	cell migration	promoter
	GO:0001756	0.001671	5.34	4.098008	10	15	somitogenesis	inside
	GO:0030154	0.001739	1.24	282.3619	323	1062	cell differentiation	inside
	GO:0009953	0.001796	3.34	7.376414	15	27	dorsal/ventral pattern formation	inside
	GO:0007420	0.0018	2.17	17.64753	29	65	brain development	inside
Hypermethylated	GO:0031325	0.001905	1.42	91.79537	116	336	positive regulation of cellular metabolic process	inside
	GO:0031016	0.002219	7.11	3.005206	8	11	pancreas development	inside
	GO:0030878	0.002221	15.99	1.912404	6	7	thyroid gland development	inside
	GO:0045597	0.002287	2.24	15.57243	26	57	positive regulation of cell differentiation	inside
	GO:0042472	0.002303	3.40	6.830013	14	25	inner ear morphogenesis	inside
	GO:0001654	0.002407	2.67	10.38162	19	38	eye development	inside
	GO:0048534	0.002449	1.72	34.96967	50	128	hemopoietic or lymphoid organ development	inside
	GO:0001764	0.002689	2.84	9.015617	17	33	neuron migration	inside
	GO:0009653	0.002713	1.70	35.04919	50	132	anatomical structure morphogenesis	inside
	GO:0007267	0.002732	1.48	66.00984	86	245	cell-cell signaling	inside
	GO:0045944	0.002991	1.71	34.42327	49	126	positive regulation of transcription from RNA polymerase II promoter	inside
	GO:0048666	0.003085	4.51	4.323974	10	16	neuron development	inside
	GO:0006812	0.003153	1.37	103.8162	128	380	cation transport	inside
	GO:0007507	0.003355	2.05	18.26662	29	67	heart development	inside
	GO:0003007	0.003369	4.45	4.371208	10	16	heart morphogenesis	inside
	GO:0048513	0.003421	1.84	17.14586	29	865	organ development	promoter
	GO:0042311	0.003551	11.57	0.317149	3	16	vasodilation	promoter
	GO:0030855	0.003746	3.11	7.103214	14	26	epithelial cell differentiation	inside
	GO:0045761	0.004025	2.67	9.288818	17	34	regulation of adenylate cyclase activity	inside
	GO:0007595	0.004249	10.74	0.336971	3	17	lactation	promoter
	GO:0007369	0.004843	3.67	5.19081	11	19	gastrulation	inside
	GO:0030902	0.004843	3.67	5.19081	11	19	hindbrain development	inside
	GO:0001935	0.005025	10.03	0.356792	3	18	endothelial cell proliferation	promoter
	GO:0042136	0.005064	5.33	3.278406	8	12	neurotransmitter biosynthetic process	inside

Hypermethylated	GO:0048839	0.005371	6.29	0.713585	4	36	inner ear development	promoter
	GO:0002052	0.005562	Inf	1.092802	4	4	positive regulation of neuroblast proliferation	inside
	GO:0009249	0.005562	Inf	1.092802	4	4	protein lipoylation	inside
	GO:0021871	0.005562	Inf	1.092802	4	4	forebrain regionalization	inside
	GO:0045885	0.005562	Inf	1.092802	4	4	positive regulation of survival gene product activity	inside
	GO:0001505	0.005663	1.97	18.03123	28	66	regulation of neurotransmitter levels	inside
	GO:0048732	0.005765	2.53	9.547686	17	35	gland development	inside
	GO:0006814	0.005976	1.84	22.12924	33	81	sodium ion transport	inside
	GO:0006023	0.006192	3.81	4.644409	10	17	aminoglycan biosynthetic process	inside
	GO:0042127	0.006293	1.35	97.53259	119	357	regulation of cell proliferation	inside
	GO:0030324	0.006442	2.40	10.38162	18	38	lung development	inside
	GO:0002062	0.006821	8.00	2.185604	6	8	chondrocyte differentiation	inside
	GO:0040018	0.006821	8.00	2.185604	6	8	positive regulation of multicellular organism growth	inside
	GO:0007611	0.006957	2.31	11.20122	19	41	learning and/or memory	inside
	GO:0045661	0.007037	13.32	1.639203	5	6	regulation of myoblast differentiation	inside
	GO:0045773	0.007037	13.32	1.639203	5	6	positive regulation of axon extension	inside
	GO:0007154	0.007222	1.48	52.15117	68	2631	cell communication	promoter
	GO:0009880	0.007599	2.89	6.830013	13	25	embryonic pattern specification	inside
	GO:0030321	0.007689	19.97	0.138753	2	7	transepithelial chloride transport	promoter
	GO:0031018	0.007689	19.97	0.138753	2	7	endocrine pancreas development	promoter
	GO:0008015	0.007821	3.29	2.279507	7	115	blood circulation	promoter
	GO:0000165	0.008109	1.64	30.87166	43	113	MAPKKK cascade	inside
	GO:0007215	0.008131	3.26	5.46401	11	20	glutamate signaling pathway	inside
	GO:0051056	0.008464	1.47	50.8153	66	186	regulation of small GTPase mediated signal transduction	inside
	GO:0007204	0.008553	5.44	0.812694	4	41	elevation of cytosolic calcium ion concentration	promoter
	GO:0006936	0.008566	3.23	2.319151	7	117	muscle contraction	promoter
	GO:0040011	0.008827	1.96	16.11883	25	59	locomotion	inside
	GO:0030900	0.008854	2.29	10.64347	18	39	forebrain development	inside
	GO:0031324	0.009247	1.32	98.62539	119	361	negative regulation of cellular metabolic process	inside
	GO:0042592	0.009255	1.34	90.42937	110	331	homeostatic process	inside

	GO:0002009	0.009685	2.92	6.26769	12	23	morphogenesis of an epithelium	inside
	GO:0001656	0.009895	2.91	6.283612	12	23	metanephros development	inside
	GO:0042445	0.00997	2.14	12.29402	20	45	hormone metabolic process	inside
Hypomethylated	GO:0032501	7.53E-07	1.31	504.7175	590	2357	multicellular organismal process	inside
	GO:0007155	2.18E-06	1.60	112.8494	157	527	cell adhesion	inside
	GO:0032502	5.33E-05	1.25	503.4327	572	2351	developmental process	inside
	GO:0006816	9.88E-05	2.42	18.41566	34	86	calcium ion transport	inside
	GO:0007165	0.000127	1.31	258.2146	308	1226	signal transduction	inside
	GO:0009887	0.000132	1.63	64.02653	91	299	organ morphogenesis	inside
	GO:0051056	0.000185	1.81	39.82921	61	186	regulation of small GTPase mediated signal transduction	inside
	GO:0007399	0.000207	1.41	134.4771	171	628	nervous system development	inside
	GO:0048731	0.000231	1.40	134.76	171	641	system development	inside
	GO:0006812	0.000257	1.52	81.37151	110	380	cation transport	inside
	GO:0007154	0.000309	1.64	55.3083	79	271	cell communication	inside
	GO:0006817	0.00031	2.64	12.84813	25	60	phosphate transport	inside
	GO:0048667	0.000336	2.06	23.98318	40	112	neuron morphogenesis during differentiation	inside
	GO:0007409	0.00043	2.07	22.69837	38	106	axonogenesis	inside
	GO:0001525	0.000486	2.08	22.05596	37	103	angiogenesis	inside
	GO:0003015	0.000486	3.16	8.351286	18	39	heart process	inside
	GO:0032989	0.000504	1.66	48.60877	70	227	cellular structure morphogenesis	inside
	GO:0035023	0.000542	2.35	15.41776	28	72	regulation of Rho protein signal transduction	inside
	GO:0030879	0.000937	7.36	2.569627	8	12	mammary gland development	inside
	GO:0008016	0.001128	3.11	7.494744	16	35	regulation of heart contraction	inside
	GO:0006811	0.001488	2.18	15.51374	27	74	ion transport	inside
	GO:0031175	0.001662	1.83	26.55281	41	124	neurite development	inside
	GO:0006820	0.001971	1.81	26.76694	41	125	anion transport	inside
	GO:0048858	0.002059	1.65	37.47372	54	175	cell projection morphogenesis	inside
	GO:0050801	0.002091	1.64	38.33026	55	179	ion homeostasis	inside
	GO:0001568	0.002126	1.77	28.48003	43	133	blood vessel development	inside
	GO:0048661	0.002212	18.39	1.284813	5	6	positive regulation of smooth muscle cell proliferation	inside
	GO:0007268	0.002312	1.54	50.32185	69	235	synaptic transmission	inside
	GO:0006813	0.002495	1.75	28.69416	43	134	potassium ion transport	inside
	GO:0030324	0.00321	2.68	8.137151	16	38	lung development	inside
	GO:0050877	0.003314	1.32	120.3442	147	562	neurological system process	inside

	GO:0006029	0.003629	10.84	0.312344	3	30	proteoglycan metabolic process	promoter
	GO:0051147	0.004404	7.36	1.92722	6	9	regulation of muscle cell differentiation	inside
	GO:0008299	0.005816	5.15	2.569627	7	12	isoprenoid biosynthetic process	inside
	GO:0007265	0.005855	1.53	41.75643	57	195	Ras protein signal transduction	inside
	GO:0005513	0.006372	9.19	1.498949	5	7	detection of calcium ion	inside
	GO:0009395	0.006372	9.19	1.498949	5	7	phospholipid catabolic process	inside
	GO:0003001	0.007507	2.03	13.2764	22	62	generation of a signal involved in cell-cell signaling	inside
Hypomethylated	GO:0043062	0.008118	1.97	14.13295	23	66	extracellular structure organization and biogenesis	inside
	GO:0007528	0.009022	5.52	2.141355	6	10	neuromuscular junction development	inside
	GO:0007613	0.009022	5.52	2.141355	6	10	memory	inside
	GO:0045884	0.009022	5.52	2.141355	6	10	regulation of survival gene product expression	inside
	GO:0048659	0.009022	5.52	2.141355	6	10	smooth muscle cell proliferation	inside
	GO:0007242	0.0098	1.19	242.8297	274	1134	intracellular signaling cascade	inside
	GO:0030949	0.009808	Inf	0.642407	3	3	positive regulation of vascular endothelial growth factor receptor signaling pathway	inside
	GO:0035313	0.009808	Inf	0.642407	3	3	wound healing, spreading of epidermal cells	inside
	GO:0045662	0.009808	Inf	0.642407	3	3	negative regulation of myoblast differentiation	inside
	GO:0045740	0.009808	Inf	0.642407	3	3	positive regulation of DNA replication	inside

Supplementary Table7. Location of CpG sites validated by bisulfite pyrosequencing

Chromosomal coordinates for CG1-CG9

Gene	Region	Chr	CG1	CG2	CG3	CG4	CG5	CG6	CG7	CG8	CG9
SLMO2	Shore	20	57049852	57049859	57049884	57049912	57049933	57049937	57049958	57049972	57050013
SLMO2	Island	20	57051256	57051274	57051277	57051280	57051287	57051293	57051299	57051305	57051308
PCDH9	Shore	13	66698906	66698964	66699018	66699044	66699123	66699126			
PCDH9	Island	13	66702731	66702734	66702745	66702747	66702753	66702766	66702770	66702776	66702779
HEY1	Shore	8	80839112	80839160	80839206	80839272					
HEY1	Island	8	80840367	80840374	80840376	80840382	80840385	80840389	80840391	80840399	80840409
HAGH	Shore	16	1814967	1814974	1814978	1815044	1815048	1815086	1815091	1815095	
HAGH	Island	16	1816837	1816840	1816843	1816846	1816848	1816851	1816868	1816873	1816875
SEMA3C	Shore	7	80387012	80387355	80387019	80387340	80387067	80387071	80387080	80387092	80387102
DSCAML1	Shore	11	117170021	117170072	117170149	117170209	117170224	117170231	117170236	117170239	117170250
GPT2	Shore	16	45474293	45474335	45474389	45474480	45474502	45474508	45474539	45474564	
ZNF532	Shore	18	54683963	54683972	54683997	54684040	54684060	54684099	54684127	54684162	

Chromosomal coordinates for CG10-CG18

Gene	Region	Chr	CG10	CG11	CG12	CG13	CG14	CG15	CG16	CG17	CG18
SLMO2	Shore	20	57050041	57050045	57050047	57050057	57050099	57050109	57050116	57050141	57050171
SLMO2	Island	20	57051310	57051317	57051305	57051327	57051329	57051337	57051347		
PCDH9	Shore	13									
PCDH9	Island	13	66702783	66702789	66702798	66702802	66702816	66702825	66702828	66702838	66702841
HEY1	Shore	8									
HEY1	Island	8	80840425	80840435	80840441	80840446					
HAGH	Shore	16									
HAGH	Island	16	1816895	1816904	1816909	1816913	1816927	1816929	1816933	1816935	1816953
SEMA3C	Shore	7	80387104	80387107	80387120						
DSCAML1	Shore	11	117170304	117170311							

Supplementary Table 8. Primer sequences and annealing temperatures used for bisulfite pyrosequencing

Gene	Region	Primer	Sequence (5'→3')	Annealing temperature (°C)	Nested annealing temperature (°C)
SLMO2	Island	Forward	ATAATGAGGTATAGAGGTTATA	44	
		Reverse	AACATCTATATCAACAAACTAA		
		Nested forward	Bio-GAGGTTATATTTGTTTTTGTTT		45
		Nested reverse	Bio-AACTCTACCCAAAAATCAAAA		
		Sequencing1 (F)	GGTTTTGTTTTAGTTTTG		
		Sequencing2 (R)	CAAAACTAAAACAAAACC		
	Shore	Forward	GATATAGTAGGTTTTAGGATGTGT	49	
		Reverse	TTACCACACTATTTTAATTAATATAACCT		
		Nested forward	AGGATGTGTTTTATTGAGTATA		43
		Nested reverse	Bio-AAAACCATTTATATTTTTAAAACT		
		Sequencing1 (F)	TGTTTTATTGAGTATAAATG		
		Sequencing2 (F)	GTGGTTTATATTTGTAATTT		
		Sequencing3 (F)	GAATTATTTGAGGTTAGGTG		
		Sequencing4 (F)	TGATTAATATGGTGAAATTT		
		Sequencing5 (F)	TTAAAAATATAAAAATTAGT		
		Sequencing6 (F)	GGAGGTTAAGGTAGGAGAAT		
		Sequencing7 (F)	GTTGTAGTGAGTTAAGAATA		
HEY	Island	Forward	GAGGTGATTATAGGGAGTAT	46	
		Reverse	AACCCTAAAATTTTTCTTTTATTC		
		Nested forward	GTTTTGGGGTAGTAATAG		41
		Nested reverse	Bio-CTAAACATCATTAAAAAACTA		
		Sequencing (F)	GGTTTTTTAGGGAATGTGTT		
	Shore	Forward	AAAGAGGTATTATTATTTATATATTTTGTGG	49	
		Reverse	TATTAAACTTAAACCTAAAATTTCACATC		
		Nested forward	GGTAGTTTTAGGAAAATTAGG		45
		Nested reverse	Bio-AACTATTAATAACCCTAAATCC		
		Sequencing1 (F)	GTATTATTTAATTGATTATT		
		Sequencing2 (F)	ATATTTGTGAATTTGAGATT		
		Sequencing3 (F)	TTGGGGTTGGTAAATGTAGG		
		Sequencing4 (F)	AATGAGATTTAATTTATTAG		
HAGH	Island	Forward	TAGGTTTGGTTTTGTTTATTTAG	46	
		Reverse	CAATAACCTAAATACTACCATAATT		
		Nested forward	Bio-GTTGTTTAGGATTGTAAAATAT		44
		Nested reverse	Bio-AAAAAAACCAACTACCTC		
		Sequencing1 (F)	TTGTTTTTAGTTAATTAG		
		Sequencing2 (F)	GTATAGTGGATTTTTGGAGGT		
		Sequencing3 (R)	CTAATTAACTAAAAACAA		
		Sequencing4 (R)	ACCTCCAAAAATCCACTATAC		
	Shore	Forward	AAGGAGTATAATAAGTAGAGTGTG	49	

		Reverse	AAAACACCTCCCTAAATTATCAA		
		Nested forward	GTAGAAGGGTTGTGATAGGAT		48
		Nested reverse	Bio-CCTCCCTAAATTATCAACTTC		
		Sequencing1 (F)	GAAGGGTTGTGATAGGATTT		
		Sequencing2 (F)	ATAAATAAAAATATTGTTTA		
		Sequencing3 (F)	ATTTTAGTATTTTGGGAGGT		
		Sequencing4 (F)	AAATATAAAAATTAGTTGGG		
		Sequencing5 (F)	GGAGGTTGAGGTAGGAGATT		
		Sequencing6 (F)	AGGTAGAGGTTGTAGTGAGT		
PCDH9	Island	Forward	GTTGATTGTTTTTTAGTTTTTTT	45	
		Reverse	CCCAACCCAAAAATAACTATA		
		Nested forward	Bio-AAATTTGATTTTGGTTTTAGGAGA		49
		Nested reverse	Bio-ACTCCCCCATCTATACATTTTAA		
		Sequencing1 (F)	GTATTGAGTATGTTTTGTAGGGTT		
		Sequencing2 (R)	CTTCACTTAACAAAAAAATAT		
	Shore	Forward	GAAAATGATTATGAGTAAATTTGGG	49	
		Reverse	CTTAAAAATAAAAATAACAACCCACC		
		Nested forward	GAGTAAATTTGGGGTTATTGT		48
		Nested reverse	Bio-CCAATTTTCAACCAACCTATA		
		Sequencing1 (F)	AGAATAGTAATAATTATAGT		
		Sequencing2 (F)	TAGTTATTGTAAAAATGAAT		
		Sequencing3 (F)	TTGTTAAAATTTTTGTTTTT		
		Sequencing4 (F)	TAGTTGTTAAGTATAATTTA		
SEMA3C	Shore	Forward	TGTATTTTTAGTAGAGATAGGGTTAG	49	
		Reverse	TCTATTAACATAACTCAAAACAACC		
		Nested forward	TTAGTAGAGATAGGGTTAGG		46
		Nested reverse	Bio-CAAAACAACCTCTCCACATAA		
		Sequencing1 (F)	GATTTTTTGATTTAATGGTT		
		Sequencing2 (F)	TTAAAATAGGTTAAGATAAA		
		Sequencing3 (F)	TAATTTTGTTGATTTTTTTA		
		Sequencing4 (F)	TATTTTTAAATATAATTATA		
DSCAML1	Shore	Forward	TAGATATTGAATAGAATGTTGGAGA	49	
		Reverse	ATTATTTCCATTCCTCTCAAAATAC		
		Nested forward	AGAATGTTGGAGATTTTTTTAATG		45
		Nested reverse	Bio-CCTTTTTTTTATAATACCTAAACT		
		Sequencing1 (F)	AGATTTTTTTAATGTTTTGA		
		Sequencing2 (F)	GAAGAATTATGAGTTTTTAT		
		Sequencing3 (F)	GAAAATAGGAATTTTAGTGT		
		Sequencing4 (F)	AGGTAATATAGATGTTGGTA		
		Sequencing5 (F)	TAAGGTAGTTATTGGAGATT		
GPT2	Shore	Forward	ATTGGGGAAGATTTTTATTTAGAG	49	
		Reverse	CTAAATCCCAAATTCTCCATATAC		

		Nested forward	AGATTGAATATTTGGTTATTAAG		44
		Nested reverse	Bio-CTCTAAAATCCTTCCCCTTAA		
		Sequencing1 (F)	ATTTAGTGTAGATAAAGGTG		
		Sequencing2 (F)	GAGTTTAAATAATTTTTTAG		
		Sequencing3 (F)	GGAATATTTAAAGATATTTT		
		Sequencing4 (F)	TAGGTGTTATTTTGATTTTA		
		Sequencing5 (F)	AAATTTAGTTAAGTAGTGTA		
ZNF532	Shore	Forward	TTTAGAGTGGAGAAGAAATGTT	49	
		Reverse	ATATTCCACATTAAACATATACCAC		
		Nested forward	GTTTAATGGGATTAGAGTGATTT		46
		Nested reverse	Bio-ATAAAAAACCTTTCAAATTAACAC		
		Sequencing1 (F)	GTGATTTTATGATGGTATTG		
		Sequencing2 (F)	GAGTTAGGTTTGGAAGGAAG		
		Sequencing3 (F)	ATGTGTTGTTTTGTATAAGA		
		Sequencing4 (F)	AAGGAAGGGTTTTATTAAAT		
DLX5C	Shore	Forward	TATTTAGGGTTATTTGGTTTTTTTT	48	
		Reverse	AACCTAACTCCCTACCCACTTATCT		
		Nested forward	Bio-GGATTGTATTAGAAAAATATAGT		43
		Nested reverse	ACCCATATTTCCCTCCTAT		
		Sequencing (R)	CTACAACTCTATTTACCC		
ERICH1	Shore	Forward	TATTTTATTGTGGGAGTTTTTGGAG	51	
		Reverse	AATAATCACATTTCTCACTTTTACCACTA		
		Nested forward	TGGGAGTTTTTGGAGTATAGT		42
		Nested reverse	TATATTTACCTATATTCCTATCT		
		Sequencing (R)	AATAAAATACATTTATTATCATT		
FAM70B	Island	Forward	TTTTGTGTTTTGTTGTGGTGTG	52	
		Reverse	CTAACTCCAAACTCCAAAACCATTA		
		Nested forward	GGAAATGAGATTTATTGAGAG		46
		Nested reverse	Bio-CTCTCCTTCTATTACAACTAA		
		Sequencing (F)	TAGTTATTGGTAATTTTTAG		
SLITRK1	Shore	Forward	TTGATTTTGATTTGTTAGTTGTTTG	48	
		Reverse	TATTCCAATATATACCCATCACCC		
		Nested forward	Bio-GGGGTTTAGGAGTAAAGGTT		45
		Nested reverse	CTACCCTATAAAAAAATCTTAAA		
		Sequencing (R)	ATTCCCAAAAATACCCTAAT		
LRFN5	Shore	Forward	TTGTTGTGGAGGAGTTTGTTAG	48	
		Reverse	TCCAACCTACTCCTTATAAATC		
		Nested forward	TGGTTTGTATGAAAGGGAATAT		49
		Nested reverse	Bio-CCTACTCCTTATAAATCAAAACACC		
		Sequencing (F)	TTTAGTTGTATTGTTTT		
FEZF2	Shore	Forward	GTGTGGTTAGAGGTATAAGTAGA	50	
		Reverse	TCAACCCTCTCAAAACTTATTCCTA		

		Nested forward	GTAGAGGGAAGAAAAGATTTTTTTT		49
		Nested reverse	Bio-CCCAATCCTCCCCCTTTC		
		Sequencing (F)	TTGTAGATTATTTTATTTG		
TMEM14A	Island	Forward	TGGGTGGGTGTAGATATTTTGTTAT	54	
		Reverse	ACAATCCTACACACACAAACCTTTA		
		Nested forward	Bio-TAGTGAAAGTTTTGGGAAATTTA		48
		Nested reverse	CTACACACACAAACCTTTAATAA		
		Sequencing (R)	ATTCATTTTAAAAAATAATCC		
TMEPAI	Shore	Forward	TATTAATTAAATTGTTTTTAGGAAGGTAAT	49	
		Reverse	AAATTTACATAAAACCACAACAAAC		
		Nested forward	GTTTTTAGGAAGGTAATTAGAA		45
		Nested reverse	Bio-TACAAAAACTTACCAAATCTATAT		
		Sequencing (F)	AAATTTTAAGAAGTTAGTA		
HOXA3	Island	Forward	GATTAATGAGTTATAGAGAGATGTTG	48	
		Reverse	AAGGAGTTAAAAGTTTTTGGAG		
		Nested forward	GTTTAGGTTTTTTATTTTATAATG		43
		Nested reverse	TAAGATTTGGTGAGGGTTTGT		
		Sequencing (F)	GGGTGATTTATGAA		

Bio = 5’ biotin added
F = forward
R = reverse

Supplementary Table 9. Real time quantitative RT-PCR assays

Gene	Assay ID (Applied Biosystems)
FZD3	Hs00184043_m1
RBM38	Hs00766686_m1
NDN	Hs00267349_s1
TRAF1	Hs00194638_m1
ZNF804A	Hs00290118_s1
CHRM2	Hs00265208_s1
NQO1	Hs00168547_m1
SEMA3C	Hs00170762_m1

Supplementary Methods

Genomic DNA isolation and McrBC fractionation. We performed genomic DNA isolation using the MasterPure DNA purification kit (Epicentre) as recommended by the manufacturer. For each sample, 5 μg of gDNA was digested, fractionated, labeled, and hybridized to a CHARM microarray as previously described[1].

CHARM microarray design. CHARM microarrays were prepared as previously described[1] and additionally included a set of 4,500 probes, totaling 148,500 base pairs across 30 genomic regions as controls. Since these control probes represent genomic regions without CpG sites and hence can not be methylated, we used them to normalize and standardize array data. We standardized the observed *M* values so that the average in the control regions was 0. Therefore, *M* values of 0 for other probes on the array are associated with no methylation.

Bisulfite pyrosequencing. Isolation of gDNA for all bisulfite pyrosequencing validation was performed using the MasterPure DNA purification kit (Epicentre) as recommended by the manufacturer. For validation of shore regions, 1 μg of genomic DNA from each sample was bisulfite treated using an EpiTect kit (Qiagen) according to the manufacturer's specifications. Converted gDNA was PCR amplified using unbiased nested primers, and quantitative pyrosequencing performed using a PSQ HS96 (Biotage) to determine percent methylation at each CpG site. Primer sequences and annealing temperatures are provided in Supplementary Table 8. For bisulfite pyrosequencing of C-DMRs in 34-65 colon tumor and 30-61 normal mucosa samples, 500ng of genomic DNA was bisulfite treated using the EZ-96 DNA Methylation Gold kit (Zymo Research) as specified by the manufacturer. Converted gDNA was PCR amplified using unbiased nested primers (Supplementary Table 8) followed by pyrosequencing using a PSQ HS96 (Biotage). Percent methylation at each CpG site was

determined using the Q-CpG methylation software (Biotage). The genomic coordinates for CpG sites measured in colon tumor and normal samples representing DLX5C, ERICH1, FAM70B, SLITRK1, FEZF2, LRFN5, TMEM14A, TMEPAI, and HOXA3 are chr7:96493826,96493847; chr8: 847868,847870; chr13:113615615; chr13:83352935; chr3:62335457,62335482,62335504; chr14:41148241,41148246,41148252; chr6:52638087; chr20:55707264; and chr7:27130446,27130448,27130450, respectively.

Total RNA isolation. Total RNA for Affymetrix microarray analysis was isolated from all human tissues using the RNeasy Mini kit (Qiagen) as specified by the manufacturer. All samples were DNase treated using the on-column DNase digestion kit (Qiagen) as recommended. Total RNA concentration was measured and quality determined using an RNA 6000 Nano Lab chip kit and running the chip on a 2100 Bioanalyzer (Agilent).

Real-time PCR. Quantitative real time PCR was performed on high quality total RNA samples, determined using Agilent Bioanalyzer and RNA Nano 6000 chips, using pre-optimized Taqman assays (Applied Biosystems). A summary of assay identification numbers can be found in Supplementary Table 9. RNA for quantitative real time PCR was prepared using Trizol (Invitrogen) for *FZD3*, *RBM38*, *NDN*, and *SEMA3C* and the RNeasy mini kit (Qiagen) was used to isolate total RNA for *ZNF804A*, *CHRM2*, and *NQO1*. We prepared complementary DNA, and eliminated genomic DNA, using the QuantiTect RT kit (QIAGEN) with Turbo DNase as recommended by the manufacturer. TaqMan assays (Applied Biosystems) were used to determine relative gene expression and experiments analyzed on a 7900HT detection system. Human *ACTB* was used as an endogenous control. Relative expression differences were calculated using the ΔΔCt method[2].

Affymetrix Microarray Expression Analysis. Genome-wide transcriptional analysis was perfomed on a total of 5 liver and 5 brain samples from the same individuals using Affymetrix U133A GeneChip microarrays. The raw microarray gene expression data for the brain tissue was obtained from The Stanley Medical Research Institute (SMRI) online genomics database www.stanleygenomics.org [3]. The five individuals selected were unaffected controls from the SMRI Array collection. The microarray expression data for these samples were submitted to the Stanley Medical Research Institute microarray database by Dr. Tadafumi Kato, Dr. Sabine Bahn and Dr. Seth Dobrin. We also carried out genome-wide transcriptional profiling on 4 colon tumor and 4 matched normal mucosa using Affymetrix U133 Plus 2.0 microarrays. One microgram of high-quality total RNA was amplified, labeled, and hybridized according to the manufacturer's (Affymetrix) specifications and as previously described [4,5].

5' RACE PCR. 5' RACE experiments were performed using a 2nd generation RACE kit (Roche Applied Science) as specified by the manufacturer's protocol. RACE PCR products were directly sequenced with 3100 Genetic Analyzer (AB Applied Biosystems). The sequences for our gene specific primers are: PIP5K1A cDNA synthesis (PIP5K1A-Sp1): TCCTGAGGAATCAACACTTC; first round PIP5K1A primer (PIP5K1A-Sp2): CAGATGCCATGGGTCTCTTG; second round PIP5K1A primer (PIP5K1A-Sp3): ACGTCGAGCCGGCTCCTGGA.

Statistical significance of DMRs. Contiguous regions composed of probes with z-scores associated with p-values smaller than 0.001 were grouped into regions. The area of each region (length multiplied by ΔM) was used to define statistical significance [6]. We used a permutation test to form a null distribution for these areas and the empirical Bayes approach described [7]. We tested for effects of fragment length on M values observed using CHARM by computing the

expected DNA fragment size based on McrBC recognition sites. Next, we stratified the delta M values for each probe, from the colon tumor and normal mucosa comparison, by fragment size. The results are in Supplementary Fig. 9, and shows no relationship between fragment size and ΔM.

GO Annotation. GO annotation was analyzed using the Bioconductor [8] Gostats package [9] to find enriched categories ($P < 0.01$)

References

1. Irizarry, R.A. et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* (2008).
2. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).
3. Higgs, B.W. et al. An online database for brain disease research. *BMC Genomics.* **7**, 70 (2006).
4. Irizarry, R.A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).
5. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-93 (2003).
6. Bullmore, E.T. et al. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans Med Imaging* **18**, 32-42 (1999).

7. Efron, B., Tibshirani, R., Storey, J.D. & Tusher, V. Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc* **96**, 1151-1160 (2001).
8. Gentleman, R.C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
9. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-8 (2007).