Supplementary Information for:

Regions of Focal DNA Hypermethylation and Long-Range Hypomethylation in Colorectal Cancer Coincide with Nuclear Lamina-associated Domains

Benjamin P. Berman[1], Daniel J. Weisenberger[1], Joseph F. Aman[1], Toshinori Hinoue[1], Zachary Ramjan[1], Yaping Liu[1], Houtan Noushmehr[1], Christopher P.E. Lange[2,3], Cornelis M. van Dijk[4], Rob A.E.M. Tollenaar[3], David Van Den Berg[1], Peter W. Laird[1]*

[1] USC Epigenome Center, University of Southern California, Keck School of Medicine, Los Angeles, CA

[2] Department of Surgery, Groene Hart Hospital, Gouda, The Netherlands

[3] Department of Surgery, Leiden University Medical Center, The Netherlands

[4] Department of Pathology, Groene Hart Hospital, Gouda, The Netherlands

* To whom correspondence should be addressed. Email: plaird@usc.edu

Phone: (323) 442-7890, FAX: (323) 442-7880.

Keywords: DNA methylation, epigenome, methylome, colorectal cancer, whole genome bisulfite sequencing, nuclear lamina, gene silencing, replication timing

Abbreviations: hESCs (human Embryonic Stem Cells), iPSCs (induced Pluripotent Stem Cells), MR (Methylation-Resistant), MP (Methylation-Prone), ML (Methylation Loss), PMD (Partially Methylated Domain), LAD (Lamina-Associated Domain), CGI (CpG Island), TSS (Transcription Start Site), base pair (bp), kilobase (kb), megabase (Mb)

Supplementary Note:

Bisulfite-seq library construction and bisulfite conversion

DNA libraries from each tissue sample were prepared using a customized DNA sample preparation protocol. Sample genomic DNA (5 µg) was fragmented using a S2 ultrasonicator (Covaris, Woburn, MA). Sizing of 50-100 ng fragmented DNA was performed using an Experion automated electrophoresis system and Experion 1K DNA Analysis kit (Bio-Rad, Hercules, CA) to confirm the correct average size and size range expected for each application. Fragmented DNA was repaired to blunt ends using the END-It kit (Epicentre Biotechnologies, Madison, WI). The end-repaired DNA was purified and recovered using a Qiagen MinElute PCR purification column. Addition of an 'A' base to the 3' end of blunt DNA was then accomplished using Klenow exo (3'- >5' exonuclease minus) (New England Biolabs, Ipswich, MA). The A-tailed DNA was then purified using the Qiagen MinElute PCR purification columns. Adapters with a 3' 'T' overhang were ligated to the end-modified DNA (see below).

For whole genome bisulfite sequencing, modified Illumina paired-end (PE) adapters were used in which cytosine bases in the adapter are replaced with 5-methylcytosine bases (Integrated DNA Technologies, Coralville, IA) to prevent conversion of the adapter sequence during bisulfite treatment. Double stranded annealed adapters were prepared by annealing and then ligated to the end-repaired DNA samples using ultrapure, rapid T4 ligase (Enzymatics, Beverly, MA) and purified on a Qiagen PCR purification column.

DNA libraries were bisulfite converted with the Zymo EZ DNA Methylation kit (Zymo Research, Orange, CA) according to the manufacturer's specifications. We then performed a series of MethyLight-based, quality control steps to ensure complete

bisulfite conversion as described in the report from Campan and colleagues (PMID: 18987824). Specifically, each bisulfite-converted DNA sample was subject to four MethyLight assays to measure bisulfite conversion completeness and the amount of bisulfite converted DNA from each sample.

Quantitative PCR (qPCR) was performed on a fraction (1 ul) of the Adapter-ligated library using primers from Illumina (PE 1.0/2.0), master mix (Applied Biosystems 2X SYBR Green) and a serial dilution curve of PhiX library DNA as a standard using an ABI 7900HT real-time PCR machine (Applied Biosystems, Foster City, CA). Based on the amplification curve from the qPCR, the optimal cycle number was determined for each library as plateau cycle, where the reaction exits the linear phase of amplification, minus 3 cycles. Amplification of the Adapter-ligated library was performed using *Pfu* Cx Turbo Taq Polymerase (Stratagene, Carlsbad, CA) according to manufacturers instructions and Illumina primers (PE 1.0/2.0). A total of 30% of the bisulfite converted adapter-ligated library was added to a total of 450 ul of reaction mix, vortexed repeatedly, then divided into 9 separate wells (50uL each) for thermal cycling. Following PCR the 9 wells are pooled and PCR products are purified using a QIAquick PCR Purification kit (Qiagen, Valencia, CA). Size fractionation and removal of adapter dimers was performed using 10% TBE acrylamide gels and a 25bp ladder size standard (Invitrogen, Carlsbad, CA). An acrylamide gel slice containing the target library (typically 200-300bp) was excised from the gel, crushed and sheared to break up the acrylamide, soaked in a Tris/Ammonium Acetate buffer (10mM/2.5mM) and precipitated with ethanol. Cluster density was estimated using qPCR with both PhiX library serial dilutions and other libraries previously run with known cluster density. Amplified library

DNA was denatured with NaOH prior to cluster generation. The denatured DNA was diluted to a 4pM concentration.

Illumina Genome Analyzer cluster generation and sequencing.

Attachment of the adapted DNA to the flow cell for Genome Analyzer sequencing was performed on an Illumina Cluster Station fluidics device. We use the Illumina Cluster Generation Kit to prepare each flow cell for sequencing. The protocol was performed for Amplification/Linearization/Blocking on one day and the flow cell is stored in a 50ml conical tube in storage buffer at 4°C for up to two weeks prior to sequencing. Single-end DNA sequencing (76 bp reads) of the bisulfite-converted DNA libraries from the colon tumor and normal adjacent tissues was performed using the Illumina Genome Analyzer IIx as described in [1].

Gene expression assay (Illumina HumanRef-8 v3.0 Expression BeadChip)

All expression data are available at the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE25070. Total RNA from 25 normal pairs of colorectal tumor and non-tumor adjacent tissue samples was isolated using the TRIZOL® Reagent (Invitrogen, Burlington, ON) according to the manufacturer's protocol. The concentrations of RNA samples were measured using the NanoDrop 8000 (Thermo Fisher Scientific Inc., Waltham, MA). The quality of the RNA samples was assessed using the Experion RNA StdSens analysis kit (Bio-Rad, Hercules, CA). Expression analysis was performed using the Illumina Ref-8 whole-genome expression BeadChip (HumanRef-8 v3.0, 24,526 transcripts) (Illumina). Briefly,

RNA samples were processed using the Illumina TotalPrep RNA Amplification Kit (Illumina). Total RNA (500ng) from each sample was subject to reverse transcription with an oligo(dT) primer bearing a T7 promoter. The cDNA then underwent second strand synthesis and purification. Biotinylated cRNA was then generated from the double-stranded cDNA template through *in vitro* transcription with T7 RNA polymerase. The biotinylated cRNA (750ng) from each patient was then hybridized to the BeadChips. The hybridized chips were stained and scanned using the Illumina HD BeadArray scanner (Illumina). Scanned image and bead-level data processing were performed using the BeadStudio 3.0.1 software (Illumina). The summarized data for each bead type were then processed using the *lumi* package in Bioconductor. The data were $\log_2$ transformed and normalized using Robust Spline Normalization (RSN) as implemented in the *lumi* package.

CpH methylation

Recent reports have identified significant DNA methylation occurring in mammalian cells at non-CpG cytosine (CpH) contexts, especially CpA [1-3]. We found that nuclear DNA had slightly greater CpH retention than mitochondrial DNA (0.65% retention vs. 0.25%), suggesting that there is indeed some low level methylation of nuclear CpH DNA. Bisulfite-seq reports of ES and iPSs have indicated that a significant fraction of CpHs are methylated in more than 25% of reads. When we searched for CpHs with >15 reads and at least 25% methylation, we found almost no difference between the normal tumor and colon, and both were at frequencies of approximately 0.1%, a level much more

consistent with that of somatic cell lines than hESCs or iPSCs, which generally had frequencies around 1% (Supplemental Figures S4-S5).

Bisulfite-seq identification of focal methylation elements: Methylation-Prone (MP), Methylation-Resistant (MR), and Methylation Loss (ML)

Unmethylated regions (UMRs) were identified by scanning all windows of at least 10 individual cytosines contained within CpG dinucleotides (each CpG dinucleotide contains two, one on each strand). Only those covered by at least 3 cytosine or thymine reads were counted, and each CpG dinucleotide was assigned a weighting factor defined as the span (in base pairs) between the next CpG dinucleotide upstream and the next CpG dinucleotide downstream. A weighted average was calculated for each window, and those with an average methylation of less than 5% in both tumor and adjacent normal tissue were categorized as methylation resistant (MR). Those with less than 5% in the adjacent normal tissue and greater than 35% in the tumor were characterized as methylation prone (MR), and those with greater than 35% in the adjacent normal tissue and less than 5% in the tumor as methylation loss (ML). Two or more overlapping UMRs of a single class were merged into one. For “clustered” UMR loci, all UMRs of a single class within 500 bp of one another were merged into one.

Homer *de novo* motif searches

De novo Motif discovery was performed using HOMER (script v2.6 (10-22-10)), an algorithm previously described[4]. Briefly, MR, MP and ML genomic sequences were extracted using Galaxy[5] and were divided into ‘target’ and ‘background’ sets for each

application of the algorithm (HOMER perl script 'findMotifs.pl'). When MP or MR were classified as 'target', known CpG Islands (CGI) overlapping domains were used for each application. Motifs of length 6, 7, 8, 10, 11, and 12 bp were identified separately for enrichment in 'target' compared to 'background' set using the cumulative hypergeometric distribution to score enrichment. To increase sensitivity of the method, up to two mismatches were allowed in each oligonucleotide sequence and distributions of CpG content in 'target' and 'background' sequences were selectively weighted to equalize the distributions of CpG content in both sets. To identify de novo motifs enriched in MP, 'target' was CGI overlapping MP and 'background' was CGI overlapping MR. To identify de novo motifs enriched in MR, 'target' was CGI overlapping MR and 'background' was CGI overlapping MP. To identify de novo motifs enriched in ML, 'target' was ML and 'background' was MR. Raw outputs from HOMER can be found in Supplementary Materials and Methods. HOMER perl script 'annotatePeaks.pl'[4] and R software plus 'ggplot2' package[6] were used to generate genomic distribution of each identified motifs.

Bisulfite-seq identification of Partially Methylated Domains (PMDs) in tumor and IMR90

Partially methylated domains were identified by scanning all windows of at least 10 kb and containing at least 10 individual cytosines contained within CpG dinucleotides (each CpG dinucleotide contains two, one on each strand). Only those covered by at least 3 cytosine or thymine reads were counted, and each CpG dinucleotide was assigned a weighting factor defined as the span (in base pairs) between the next CpG dinucleotide

upstream and the next CpG dinucleotide downstream. A weighted average was calculated for each window, and those with an average methylation between 20-60% were categorized as partially methylated. All overlapping partially methylated windows were collapsed into a single partially methylated domain (PMD), and those of less than 100 kb in length were discarded.

External data: ENCODE ChIP-seq

ChIP-seq data was downloaded from the UCSC ENCODE portal. (http://genome.ucsc.edu/ENCODE). Peaks defined for transcription factor binding clusters were highly specific, but peaks for histone modifications were overly broad so we used only the top 30% of peaks for a given mark, based on the average signal value in the ENCODE peaks files (supplemental fig X).

External data: Hi-C chromatin conformation capture data (GM12678 lymphoblastoid cells)

We used HiC compartments from GM06990 lymphoblastoid cells taken from[7] (GEO #GSE18199). We downloaded 100 kb bin values based on the first principal component (eigenvector), and removed chromosomes 2 and 10 as the article supplement indicated that those two chromosomes did not follow the A/B compartment structure of the others. We identified compartment boundaries using the definition of two adjacent 100kb windows with negative values, followed by two adjacent windows with positive values (yielding 1,095 boundaries total, see Supplemental Data File).

External data: Replication Timing (BJ Fibroblasts)

We downloaded replication timing data from the supplemental materials of ref[8]. Based on a personal communication with the author Scott Hansen, we calculated the ratio of cells dividing at the late (S4 and G2) time points vs. the early (G1 and S1) timepoints for each locus in the genome, and identified those regions with a late:early log2-ratio of greater than 1.5. This cutoff was chosen based on a bimodal distribution of ratios, with a clear peak of early-replicating with values below -1.5, and a clear peak of late-replicating with values above 1.5 (data not shown). We merged any matching positions within 2 kb, and selected those regions larger than 100 kb as late replicating domains (yielding 1,029 late-replicating regions, see Supplemental data file).

Supplementary Table 1. Bisulfite-seq samples

Sample	ID	PCR cycles	GAIIx	Unique mapped reads	Total uniquely mapped sequence (Gb)	Avg coverage	Genome covered	CpH retention (nuclear genome)	CpG retention (nuclear genome)	CpH retention (mitochondrial genome)	CpG retention (mitochondrial genome)	CpGs covered by at least 1 read	CpGs covered by at least 5 non-duplicate reads (either strand)	CpGs covered by at least 10 non-duplicate reads (either strand)	Per cent CpGs filtered out as possible genetic variants
Normal colon mucosa	OTB14838N	5	Single end 75bp	1,205,774,297	86.7	32.22	94.11%	0.61%	73.14%	0.25%	0.25%	87.17%	81.33%	75.74%	2.4%
Colon tumor	OTB14838T	7	Single end 75bp	1,050,881,599	75.8	28.13	94.25%	0.65%	61.98%	0.26%	0.27%	87.09%	79.43%	72.31%	3.8%

Supplementary Table 2. Bisulfite-seq sample details

Normal Colon Including Duplicates

											High-confidence CpGs							High-confidence CpHs			
Chrom	minMapQ	Cytosine maximum oppisite strand non-G	CpG maximum next base non-G	Minimum converted CpHs for 5' conversion filter		total uniquely mapped reads	total genomic coordinates covered	total sequence (bp)	CpGs removed 5' conversion filter	CpGs retained 5' conversion filter	Unique bases covered (both strands)	Total C/T sequence, single-stranded (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)
chr1	30	0.101	0.101	1		98,974,836	211,260,609	7,031,554,208	5,455,533	64,942,250	3,870,875	59,892,578	43,093,105	3,306,768	2,616,016	3,711,576	3,485,912	71.19%	79,939,278	1,303,055,382	7,439,216	68,714,099
chr2	30	0.101	0.101	1		105,049,371	226,185,997	7,563,554,712	5,510,420	63,804,488	3,727,898	59,778,945	44,576,932	3,227,640	2,596,842	3,592,105	3,402,780	73.81%	82,936,553	1,361,875,019	8,664,095	71,550,989
chr3	30	0.101	0.101	1		86,695,031	187,499,838	6,242,037,408	4,221,470	47,953,236	2,830,916	44,701,386	33,360,036	2,441,877	1,957,583	2,726,158	2,578,774	74.29%	68,192,899	1,113,397,015	7,228,723	58,790,198
chr4	30	0.101	0.101	1		80,034,844	179,589,443	5,759,231,968	3,943,783	42,417,875	2,527,897	39,344,815	29,290,370	2,165,013	1,717,004	2,431,283	2,295,350	74.22%	62,671,877	989,047,918	7,238,957	53,283,680
chr5	30	0.101	0.101	1		77,258,879	168,214,522	5,562,639,288	3,832,098	43,695,917	2,584,242	40,896,990	29,996,714	2,233,203	1,790,576	2,489,399	2,357,194	73.16%	60,790,971	988,588,905	6,506,959	52,313,913
chr6	30	0.101	0.101	1		74,162,466	159,991,399	5,326,983,568	3,902,141	44,225,908	2,553,557	41,350,886	30,467,734	2,204,846	1,768,450	2,460,361	2,327,991	73.32%	57,786,847	945,178,618	6,265,843	49,585,888
chr7	30	0.101	0.101	1		65,779,667	145,713,284	4,720,866,136	3,657,985	42,805,651	2,618,936	39,985,279	29,738,451	2,228,060	1,757,112	2,505,312	2,349,544	73.56%	53,349,894	852,090,136	5,421,948	45,494,091
chr8	30	0.101	0.101	1		63,431,077	136,375,388	4,567,037,544	3,340,552	38,445,236	2,267,074	36,078,962	26,795,919	1,968,736	1,579,681	2,190,453	2,076,673	73.52%	49,988,505	818,866,215	5,214,684	43,203,107
chr9	30	0.101	0.101	1		48,702,816	106,140,129	3,506,602,752	2,691,118	32,153,883	1,984,975	30,243,086	22,107,900	1,692,709	1,335,463	1,898,646	1,781,566	71.84%	39,775,169	645,307,789	3,812,255	34,187,384
chr10	30	0.101	0.101	1		58,707,349	123,587,643	4,221,358,568	3,509,072	40,890,421	2,334,717	38,354,884	28,488,153	2,032,505	1,643,625	2,252,765	2,136,201	73.23%	46,690,092	780,612,140	4,562,697	40,552,141
chr11	30	0.101	0.101	1		57,945,386	125,493,735	4,172,067,792	3,112,092	37,139,524	2,260,644	34,897,778	25,023,153	1,945,287	1,540,477	2,177,420	2,051,071	70.72%	47,464,763	771,090,097	4,367,904	40,877,132
chr12	30	0.101	0.101	1		57,525,873	125,548,987	4,141,862,856	3,137,631	36,543,217	2,229,246	34,393,717	24,993,506	1,910,618	1,516,747	2,141,979	2,016,230	72.38%	46,506,429	748,890,079	4,582,362	39,923,847
chr13	30	0.101	0.101	1		41,622,010	92,492,888	2,996,784,720	2,132,131	23,911,747	1,413,159	22,376,386	16,739,596	1,223,298	981,309	1,364,609	1,293,853	74.47%	32,566,589	519,677,265	3,744,646	27,862,300
chr14	30	0.101	0.101	1		38,790,772	84,532,488	2,792,935,584	2,087,383	24,582,847	1,496,336	23,120,958	16,783,432	1,283,523	1,017,156	1,437,709	1,353,467	71.92%	31,420,691	508,215,112	3,103,098	27,044,061
chr15	30	0.101	0.101	1		35,495,619	75,213,041	2,555,684,568	2,029,717	24,238,826	1,445,317	22,826,010	16,724,475	1,240,848	993,090	1,385,813	1,303,723	72.50%	28,651,367	477,881,244	2,685,195	24,782,043
chr16	30	0.101	0.101	1		34,592,508	72,753,681	2,490,660,576	2,446,980	30,086,603	1,847,550	28,534,640	20,718,259	1,579,438	1,237,132	1,769,143	1,653,468	70.88%	29,020,356	482,819,319	2,357,998	25,200,946
chr17	30	0.101	0.101	1		32,819,358	72,618,925	2,362,993,776	2,337,374	29,618,564	1,952,328	28,285,594	19,818,996	1,644,477	1,269,056	1,860,254	1,725,811	68.75%	29,335,234	465,344,584	2,188,699	25,183,361
chr18	30	0.101	0.101	1		34,017,588	72,157,224	2,447,431,328	1,842,766	21,100,558	1,199,901	19,662,924	14,694,330	1,045,196	846,509	1,160,287	1,100,745	74.09%	26,231,819	437,111,882	2,841,862	22,694,438
chr19	30	0.101	0.101	1		22,657,201	52,787,824	1,611,461,064	1,994,708	25,767,641	1,825,884	24,491,673	16,341,251	1,507,308	1,113,594	1,735,909	1,591,286	65.67%	22,165,814	330,322,260	1,332,069	18,636,561
chr20	30	0.101	0.101	1		28,175,886	57,625,600	2,028,663,792	1,738,864	21,566,664	1,281,116	20,426,355	14,541,883	1,117,571	898,797	1,238,613	1,171,470	69.92%	23,079,957	393,871,618	1,916,552	20,292,299
chr21	30	0.101	0.101	1		15,496,466	32,918,345	1,115,745,552	969,028	11,316,216	660,845	10,568,765	7,871,080	574,592	459,444	638,221	604,112	73.64%	12,130,273	201,443,831	1,273,137	10,453,023
chr22	30	0.101	0.101	1		14,677,365	32,145,863	1,056,770,280	1,136,219	14,952,889	978,172	14,295,616	10,227,355	831,912	648,310	934,173	869,725	69.91%	13,666,922	219,074,773	876,882	11,863,997
chrX	30	0.101	0.101	1		29,939,501	134,872,696	2,155,644,072	1,408,482	15,613,808	1,870,589	14,690,490	10,913,577	1,281,091	682,888	1,707,934	1,415,232	73.95%	46,581,686	379,239,566	2,471,664	31,263,406
chrY	30	0.101	0.101	1		3,222,428	13,954,655	232,014,816	191,238	1,760,858	166,159	1,522,248	1,132,641	108,101	58,028	145,428	118,687	75.07%	4,584,657	40,966,047	243,640	2,893,097
chrM	30	0.101	0.101	1		993,674	16,239	71,544,528	235,873	1,856,826	783	2,040,971	5,030	702	697	757	753	0.25%	5,694	9,636,871	24,149	5,286
nuclear genome totals						1,205,774,297	2,689,674,204	86,662,586,928	66,628,785	779,534,827	47,928,333	730,720,965	534,438,848	40,794,617	32,024,889	45,955,550	43,060,865		995,528,642	##########	96,341,085	846,646,001

Summaries:

Mean reads per position covered (both strands)	32.22	Mean reads per CpG (each strand)	15.25	High-confidence CpGs in sample	47,928,333
Total bp covered in hg18 (from UCSC)	2,858,034,764	Total CpGs in reference	56,328,604	Percent with possible sequence variant	2.4%
Percentage of hg18 covered	94.1%	Total reference CpGs covered:	49,101,921		
		Percent of reference covered (1 read including	87.17%		
		Percent of reference covered (5 read including	81.58%		
Mean CpH retention (nuclear genome)	0.61%	Percent of reference covered (10 reads includi	76.45%	Mean CpG methylation (nuclear genome)	73.14%
Mean CpH retention (mitochondrial genome)	0.25%			Mean CpG methylation (mitochondrial genom	0.25%

Normal Colon Excluding Duplicates

											High-confidence CpGs							High-confidence CpHs			
Chrom	minMapQ	Cytosine maximum oppisite strand non-G	CpG maximum next base non-G	Minimum converted CpHs for 5' conversion filter		total uniquely mapped reads	total genomic coordinates covered	total sequence (bp)	CpGs removed 5' conversion filter	CpGs retained 5' conversion filter	Unique bases covered (both strands)	Total C/T sequence, single-stranded (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)
chr1	30	0.101	0.101	1		85,980,455	211,125,778	6,190,592,760	4,649,722	55,693,355	3,866,815	52,470,518	37,741,984	3,275,861	2,520,319	3,699,062	3,453,680	71.21%	79,869,127	1,146,913,356	6,548,785	68,147,799
chr2	30	0.101	0.101	1		92,354,414	226,073,954	6,649,517,808	4,781,114	55,643,376	3,724,585	52,203,162	38,857,422	3,200,320	2,508,179	3,581,679	3,375,051	73.83%	82,872,455	1,199,399,834	7,648,002	70,982,060
chr3	30	0.101	0.101	1		76,507,611	187,434,938	5,508,543,888	3,695,082	42,161,386	2,828,807	39,343,776	29,335,324	2,420,764	1,890,313	2,717,924	2,557,428	74.31%	68,146,685	983,985,353	6,403,392	58,328,546
chr4	30	0.101	0.101	1		70,585,161	179,506,091	5,082,131,592	3,351,407	37,120,365	2,525,383	34,499,261	25,688,701	2,145,817	1,655,845	2,423,815	2,275,993	74.24%	62,623,587	874,467,075	6,428,509	52,834,557
chr5	30	0.101	0.101	1		68,214,627	168,139,838	4,911,453,144	3,359,784	38,441,263	2,582,159	35,997,496	26,391,290	2,213,972	1,729,437	2,482,327	2,337,913	73.19%	60,747,312	874,037,407	5,768,257	51,898,253
chr6	30	0.101	0.101	1		64,741,748	159,932,089	4,661,405,856	3,319,038	38,022,648	2,551,533	35,622,893	26,190,793	2,185,937	1,707,776	2,453,155	2,308,729	73.34%	57,746,902	826,721,957	5,516,154	49,186,200
chr7	30	0.101	0.101	1		57,809,071	145,633,704	4,162,253,112	3,183,980	37,337,576	2,616,552	35,108,929	26,069,945	2,206,872	1,691,402	2,496,906	2,327,107	73.58%	53,304,520	751,223,193	4,788,523	45,098,316
chr8	30	0.101	0.101	1		55,964,454	136,313,218	4,029,440,688	2,926,053	33,830,771	2,264,996	31,762,707	23,565,128	1,952,481	1,525,562	2,184,601	2,060,037	73.54%	49,952,165	723,685,336	4,614,402	42,868,044
chr9	30	0.101	0.101	1		42,911,927	106,042,601	3,089,658,744	2,360,499	28,329,125	1,982,899	26,681,819	19,476,148	1,677,259	1,286,845	1,892,627	1,765,445	71.86%	39,735,179	569,817,751	3,363,268	33,907,175
chr10	30	0.101	0.101	1		51,234,303	123,527,820	3,688,869,816	2,942,041	35,031,302	2,332,921	32,974,031	24,397,233	2,016,020	1,588,284	2,246,728	2,119,904	73.25%	46,654,320	685,036,195	3,994,413	40,240,249
chr11	30	0.101	0.101	1		51,054,235	125,438,967	3,675,904,920	2,727,249	32,699,549	2,258,884	30,762,355	22,030,622	1,928,483	1,484,391	2,170,979	2,033,708	70.74%	47,429,148	680,954,391	3,857,582	40,549,659
chr12	30	0.101	0.101	1		50,786,765	125,501,918	3,656,647,080	2,754,753	32,203,169	2,227,631	30,323,950	22,015,087	1,892,935	1,461,906	2,135,362	1,997,974	72.39%	46,473,424	662,402,420	4,054,964	39,592,854
chr13	30	0.101	0.101	1		36,811,037	92,462,370	2,650,394,664	1,870,800	21,051,801	1,412,183	19,707,113	14,731,177	1,213,218	948,177	1,360,996	1,284,041	74.49%	32,544,637	460,050,785	3,328,652	27,634,698
chr14	30	0.101	0.101	1		34,246,172	84,491,577	2,465,724,384	1,832,142	21,663,837	1,495,170	20,384,419	14,775,320	1,271,966	980,635	1,433,253	1,341,823	71.94%	31,396,432	449,446,881	2,745,674	26,823,882
chr15	30	0.101	0.101	1		31,175,259	75,162,006	2,244,618,648	1,774,556	21,269,640	1,443,868	20,045,824	14,666,102	1,229,631	958,386	1,381,345	1,291,755	72.52%	28,625,195	420,563,111	2,358,529	24,581,891
chr16	30	0.101	0.101	1		30,180,257	72,686,680	2,172,978,504	2,074,907	25,977,014	1,845,185	24,692,126	17,835,632	1,564,257	1,189,324	1,763,018	1,637,114	70.90%	28,989,434	423,624,023	2,048,616	24,991,628
chr17	30	0.101	0.101	1		28,943,009	72,573,789	2,083,896,648	2,058,945	26,197,097	1,950,581	25,038,815	17,501,269	1,627,579	1,217,143	1,853,445	1,707,498	68.77%	29,307,796	411,636,595	1,924,747	24,960,154
chr18	30	0.101	0.101	1		29,660,782	72,130,349	2,135,576,304	1,575,420	18,168,351	1,199,075	17,056,083	12,750,675	1,036,814	819,335	1,157,162	1,092,595	74.11%	26,213,946	381,728,545	2,498,650	22,521,431
chr19	30	0.101	0.101	1		19,675,967	52,752,677	1,416,669,624	1,713,326	22,437,447	1,824,133	21,643,173	14,378,566	1,489,852	1,060,741	1,729,155	1,571,483	65.68%	22,143,321	290,973,120	1,161,818	18,444,080
chr20	30	0.101	0.101	1		24,751,241	57,607,057	1,782,089,352	1,521,861	18,976,046	1,280,302	17,995,829	12,795,258	1,108,634	868,030	1,235,640	1,162,253	69.94%	23,064,626	347,018,231	1,680,380	20,144,345
chr21	30	0.101	0.101	1		13,432,119	32,902,885	967,112,568	832,274	9,815,348	660,308	9,232,416	6,873,117	569,652	443,206	636,387	599,320	73.67%	12,120,804	175,328,483	1,107,453	10,366,469
chr22	30	0.101	0.101	1		12,949,075	32,118,388	932,333,400	998,180	13,259,041	977,172	12,680,802	9,054,728	824,009	622,670	930,808	861,332	69.92%	13,652,859	193,963,317	767,449	11,765,106
chrX	30	0.101	0.101	1		27,979,534	134,821,922	2,014,526,448	1,307,464	14,534,575	1,869,328	13,681,958	10,167,678	1,259,235	626,825	1,700,290	1,386,803	73.98%	46,550,882	354,528,272	2,310,620	30,733,878
chrY	30	0.101	0.101	1		2,647,881	13,921,689	190,647,432	133,694	1,337,522	165,610	1,212,297	913,207	105,985	52,987	144,402	115,744	75.14%	4,576,897	33,890,444	206,404	2,831,564
chrM	30	0.101	0.101	1		31,783	15,282	2,288,376	6,215	52,997	696	57,455	146	651	627	675	670	0.30%	5,179	398,894	1,158	4,776
nuclear genome totals						1,060,597,104	2,688,302,305	76,362,987,384	57,744,291	681,201,604	47,886,080	641,121,752	468,202,406	40,417,553	30,837,718	45,811,066	42,664,730		994,741,653	##########	85,125,243	839,432,838

Summaries:

Mean reads per position covered (both strands)	28.41	Mean reads per CpG (each strand)	13.39	High-confidence CpGs in sample	47,886,080
Total bp covered in hg18 (from UCSC)	2,858,034,764	Total CpGs in reference	56,328,604	Percent with possible sequence variant	2.5%
Percentage of hg18 covered	94.1%	Total reference CpGs covered:	49,101,921	Percent ref cytosines possible non-cytosine	0.42%
		Percent of reference covered (1 read including	87.17%		
		Percent of reference covered (5 read including	81.33%		
Mean CpH retention (nuclear genome)	0.61%	Percent of reference covered (10 reads includi	75.74%	Mean CpG methylation (nuclear genome)	73.03%
Mean CpH retention (mitochondrial genome)	0.29%			Mean CpG methylation (mitochondrial genom	0.25%

Supplementary Table 2. Bisulfite-seq sample details (cont.)

Cytosines with ambiguous context											Possible non-cytosines in sample genome								
Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	cphConversionPreFilter
53,675,646	78,284,077	75,352,670	0.67%	525,963	4,474,934	538,218	424,266	96,121	512,612	466,473	9.04%	344,323	5,430,985	283,939	277,468	209,690	336,001	299,057	5.58%	98.99%
56,056,924	81,378,446	78,479,617	0.75%	531,826	4,555,765	579,722	431,184	99,980	519,360	475,265	9.13%	359,015	5,770,402	313,147	292,546	222,809	351,348	315,531	5.59%	98.88%
45,944,674	66,992,738	64,604,572	0.77%	432,578	3,663,018	446,632	350,481	81,155	422,957	387,183	8.97%	298,634	4,708,848	250,062	243,369	185,496	292,420	262,851	5.49%	98.86%
40,825,999	61,459,219	59,007,245	0.87%	412,004	3,489,566	437,260	332,671	74,363	402,771	367,084	9.03%	295,085	4,667,412	244,198	239,275	179,945	288,475	258,609	5.41%	98.73%
40,827,251	59,658,992	57,498,126	0.78%	387,259	3,240,381	382,664	313,552	71,424	378,382	346,142	8.86%	263,974	4,105,287	214,117	215,308	163,644	258,425	232,208	5.46%	98.84%
38,561,897	56,709,917	54,613,830	0.79%	376,650	3,338,534	449,147	305,013	71,357	368,303	336,649	9.25%	262,492	4,250,909	242,220	213,673	162,294	256,941	230,522	5.64%	98.84%
35,217,218	52,096,507	49,914,285	0.74%	360,748	3,042,757	399,982	290,538	66,940	350,938	317,105	9.80%	240,672	3,665,456	215,012	192,953	144,756	234,624	208,285	6.03%	98.89%
33,818,853	49,126,447	47,421,152	0.75%	320,986	2,709,710	331,589	260,963	60,559	314,311	288,365	9.16%	215,320	3,397,648	181,782	176,224	135,065	210,874	190,231	5.56%	98.88%
26,725,146	38,858,317	37,368,988	0.69%	264,072	2,244,627	281,122	213,327	49,178	256,788	233,139	9.45%	173,924	2,712,982	151,158	140,087	105,897	169,519	150,890	5.75%	98.95%
32,117,450	45,790,506	44,190,499	0.68%	304,016	2,764,664	401,423	247,566	61,459	296,847	271,704	9.83%	205,876	3,478,133	215,329	167,506	128,337	201,120	180,445	5.99%	98.97%
31,947,026	46,534,527	44,799,316	0.67%	312,448	2,622,110	316,205	252,647	58,415	305,151	278,065	9.14%	204,467	3,226,124	174,576	165,437	125,866	199,828	178,566	5.59%	98.99%
31,060,776	45,596,914	43,874,562	0.72%	308,336	2,563,609	310,892	248,505	56,337	300,896	274,077	9.30%	208,746	3,180,027	173,734	168,308	126,590	203,666	181,533	5.75%	98.92%
21,494,840	31,994,820	30,780,276	0.86%	210,075	1,737,722	207,720	170,164	37,358	205,688	187,782	9.00%	150,904	2,309,320	121,063	122,109	91,844	147,569	132,003	5.35%	98.75%
21,076,037	30,813,571	29,654,301	0.72%	206,614	1,723,800	209,847	166,963	38,291	201,698	183,651	9.36%	138,059	2,112,851	116,596	111,313	83,865	134,925	120,350	5.72%	98.92%
19,605,261	28,031,187	27,001,866	0.65%	186,922	1,600,367	206,383	150,773	36,343	181,755	165,079	9.65%	125,710	1,971,153	113,269	101,336	76,974	122,529	109,210	5.86%	99.00%
19,917,784	28,322,001	27,198,903	0.55%	204,501	1,857,493	285,952	165,853	41,929	198,603	179,349	11.11%	128,295	2,072,126	148,570	102,956	77,733	124,884	110,916	6.71%	99.12%
19,595,852	28,558,747	27,306,140	0.53%	214,059	1,780,576	231,455	171,392	39,569	207,466	185,869	10.40%	127,998	1,890,271	123,884	100,648	74,353	124,157	108,185	6.79%	99.15%
17,827,951	25,780,893	24,898,324	0.78%	167,564	1,516,876	222,011	136,042	32,079	164,030	150,381	9.42%	115,928	2,011,967	110,365	94,833	72,494	113,423	102,084	5.58%	98.86%
14,049,177	21,475,186	20,291,690	0.44%	181,282	1,528,407	211,832	143,197	31,026	175,128	154,228	11.39%	100,409	1,440,878	98,509	76,702	53,822	96,957	82,601	7.39%	99.27%
16,278,347	22,694,199	21,971,274	0.56%	150,187	1,324,337	174,663	122,836	30,877	146,878	134,823	9.98%	93,531	1,525,879	92,678	76,513	59,001	91,433	81,976	6.34%	99.11%
8,160,650	11,877,460	11,419,026	0.74%	83,465	816,904	131,995	67,742	17,047	81,289	73,976	10.68%	56,683	1,053,964	72,525	45,612	34,599	55,250	49,180	6.33%	98.89%
9,333,482	13,311,139	12,759,147	0.44%	99,882	842,111	114,986	80,575	19,442	96,762	86,967	11.00%	55,768	824,565	57,908	43,813	32,615	54,087	47,081	7.29%	99.26%
16,148,051	43,848,622	36,842,706	0.76%	400,449	2,565,065	120,496	295,951	19,386	385,875	308,858	4.83%	203,999	1,688,288	59,762	135,453	71,970	194,053	145,379	3.88%	98.85%
1,463,719	4,078,534	3,310,456	0.73%	42,132	391,477	66,146	30,671	3,932	39,619	31,074	6.42%	20,348	347,454	26,805	13,323	7,727	19,259	14,804	5.32%	98.99%
5,219	5,567	5,559	0.31%	61	5,621	20	50	34	60	59	0.90%	41	79,425	86	37	34	41	40	0.14%	99.46%
651,730,011	973,272,966	930,558,971		6,684,018	56,394,810	7,058,342	5,372,872	1,194,567	6,514,107	5,883,288		4,390,160	67,842,929	3,801,208	3,516,765	2,627,386	4,281,767	3,792,497		

Cytosines with ambiguous context											Possible non-cytosines in sample genome								
Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	cphConversionPreFilter
51,748,431	78,126,873	74,885,253	0.67%	545,753	4,328,047	468,834	440,414	88,949	532,258	483,106	8.73%	346,339	4,702,677	250,538	276,600	203,279	337,573	297,560	5.59%	98.99%
54,091,505	81,232,175	78,023,893	0.75%	553,043	4,432,961	496,415	448,694	92,818	540,319	493,200	8.81%	360,820	4,975,825	268,446	291,544	216,086	352,679	314,043	5.58%	98.88%
44,337,256	66,877,967	64,236,898	0.76%	449,253	3,595,298	395,083	364,134	75,233	439,468	401,078	8.64%	300,004	4,146,994	216,580	242,567	179,896	293,460	261,441	5.48%	98.86%
39,301,993	61,347,958	58,639,098	0.86%	427,265	3,394,428	383,589	344,862	68,612	417,867	379,647	8.73%	296,489	4,023,158	211,446	238,147	173,975	289,553	257,249	5.42%	98.73%
39,387,755	59,555,323	57,166,988	0.78%	402,443	3,201,796	344,831	326,051	66,354	393,400	358,939	8.54%	265,144	3,635,023	188,050	214,398	158,566	259,295	230,877	5.46%	98.84%
37,185,315	56,609,511	54,292,676	0.78%	391,399	3,144,668	361,506	317,008	66,353	382,847	348,990	8.94%	263,794	3,623,868	199,520	212,894	157,281	257,932	229,459	5.63%	98.84%
33,907,839	51,984,047	49,575,443	0.74%	374,228	2,975,299	355,076	301,439	61,978	364,237	328,126	9.48%	241,712	3,225,543	187,732	191,842	140,013	235,294	206,874	6.02%	98.89%
32,636,390	49,045,962	47,158,456	0.75%	333,450	2,674,554	299,542	271,257	56,320	326,681	298,828	8.86%	216,318	3,006,461	160,607	175,675	131,166	211,611	189,304	5.55%	98.88%
25,769,335	38,778,406	37,136,499	0.69%	273,923	2,188,570	247,655	221,437	45,465	266,612	241,436	9.11%	174,970	2,374,998	129,425	139,761	102,924	170,329	150,294	5.73%	98.96%
31,019,696	45,707,517	43,938,932	0.68%	315,885	2,583,241	310,851	257,236	56,960	308,625	281,732	9.45%	206,683	2,863,746	167,522	166,767	124,429	201,697	179,266	5.97%	98.97%
30,802,226	46,448,228	44,531,002	0.66%	324,864	2,588,468	285,744	262,733	54,302	317,421	288,498	8.83%	205,436	2,801,061	148,429	164,857	122,176	200,522	177,559	5.57%	98.99%
29,934,927	45,511,826	43,604,245	0.71%	320,272	2,527,924	280,130	258,151	51,933	312,671	283,895	8.99%	210,011	2,819,895	153,883	167,687	122,703	204,662	180,706	5.73%	98.92%
20,714,602	31,941,345	30,600,476	0.86%	218,652	1,725,108	187,637	177,144	34,693	214,192	195,005	8.67%	151,923	2,051,123	107,042	121,743	89,238	148,403	131,608	5.34%	98.75%
20,320,168	30,754,874	29,474,925	0.71%	214,639	1,706,025	190,601	173,604	35,598	209,708	190,481	9.04%	139,063	1,874,314	102,750	111,195	81,560	135,729	119,990	5.68%	98.92%
18,923,852	27,973,056	26,834,214	0.65%	194,067	1,568,834	184,160	156,693	33,678	188,871	171,157	9.31%	126,249	1,735,451	98,550	100,869	74,720	122,883	108,574	5.84%	99.00%
19,206,941	28,256,050	27,014,465	0.54%	211,592	1,749,220	234,396	171,580	38,594	205,591	185,302	10.71%	128,919	1,760,599	114,867	102,601	75,279	125,333	110,274	6.63%	99.12%
18,849,967	28,487,626	27,104,187	0.53%	221,628	1,750,132	208,937	177,496	36,466	214,916	192,016	10.06%	129,025	1,660,335	109,235	100,496	72,072	124,982	107,775	6.79%	99.16%
17,214,609	25,737,782	24,763,023	0.77%	174,277	1,403,611	161,952	141,662	29,738	170,715	156,120	9.09%	116,657	1,619,775	87,126	94,573	70,507	114,028	101,699	5.57%	98.85%
13,440,047	21,412,875	20,103,879	0.43%	187,096	1,434,522	181,119	147,668	28,187	180,779	158,647	11.05%	101,481	1,190,851	84,712	76,575	51,742	97,858	82,395	7.37%	99.28%
15,739,482	22,655,918	21,854,566	0.56%	156,021	1,287,414	153,346	127,728	28,778	152,589	139,814	9.64%	94,134	1,326,063	79,804	76,382	57,285	91,916	81,724	6.30%	99.12%
7,865,825	11,853,116	11,346,522	0.74%	86,436	741,789	104,294	70,165	15,848	84,257	76,427	10.38%	56,901	820,322	55,277	45,322	33,339	55,355	48,801	6.35%	98.89%
8,985,231	13,278,548	12,670,289	0.44%	103,475	829,057	104,772	83,528	18,061	100,336	90,057	10.66%	56,297	717,186	51,576	43,846	31,678	54,505	46,928	7.31%	99.27%
14,830,022	43,702,471	36,147,963	0.75%	411,443	2,584,389	114,990	303,993	15,879	396,542	314,626	4.72%	205,683	1,579,501	55,669	134,698	67,095	195,330	143,851	3.85%	98.85%
1,333,117	4,053,137	3,231,967	0.73%	42,880	302,877	29,427	31,097	3,215	40,388	31,328	6.19%	20,177	196,843	13,380	13,013	7,078	19,045	14,310	5.38%	98.93%
4,560	5,020	4,968	0.31%	33	370	0	22	7	32	30	0.00%	32	2,639	4	29	29	32	30	0.85%	99.43%
627,546,531	971,332,591	924,335,859		6,933,984	54,718,232	6,084,887	5,575,774	1,104,012	6,761,290	6,088,455		4,414,229	58,731,612	3,242,166	3,504,052	2,544,087	4,299,974	3,772,561		

Supplementary Table 2. Bisulfite-seq sample details (cont.)

Colon Tumor Including Duplicates

											High-confidence CpGs								High-confidence CpHs			
Chrom	minMapQ	Cytosine maximum oppisite strand non-G	CpG maximum next base non-G	Minimum converted CpHs for 5' conversion filter		total uniquely mapped reads	total genomic coordinates covered	total sequence (bp)	CpGs removed 5' conversion filter	CpGs retained 5' conversion filter	Unique bases covered (both strands)	Total C/T sequence, single-stranded (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)
chr1		30	0.101	0.101	1	81,614,082	211,399,502	5,792,171,684	4,308,633	50,244,185	3,791,886	45,131,842	27,837,276	3,004,064	2,069,117	3,576,970	3,222,871	62.50%	79,455,676	1,061,900,161	6,620,003	65,603,900
chr2		30	0.101	0.101	1	93,973,585	226,591,412	6,793,734,156	4,781,793	54,664,409	3,681,869	50,660,599	31,823,635	3,107,078	2,352,909	3,529,624	3,306,554	63.58%	82,900,501	1,218,435,078	8,189,615	71,074,374
chr3		30	0.101	0.101	1	78,360,386	187,799,393	5,664,593,572	3,717,945	41,614,016	2,797,981	38,395,351	23,811,632	2,357,467	1,787,908	2,680,282	2,510,203	63.25%	68,169,770	1,006,365,450	6,907,482	58,578,444
chr4		30	0.101	0.101	1	75,150,675	179,943,698	5,431,779,468	3,565,638	37,949,496	2,501,115	34,916,336	21,250,250	2,106,739	1,600,678	2,396,906	2,247,660	62.16%	62,780,693	928,786,092	7,124,865	53,714,021
chr5		30	0.101	0.101	1	70,282,648	168,530,199	5,080,628,732	3,380,828	38,026,193	2,552,282	35,238,577	21,340,609	2,153,951	1,637,147	2,445,505	2,292,373	61.96%	60,792,652	899,326,735	6,249,233	52,208,717
chr6		30	0.101	0.101	1	66,889,650	160,253,383	4,792,693,236	3,430,369	38,270,569	2,521,548	35,364,446	22,095,857	2,125,099	1,609,161	2,417,386	2,264,913	63.31%	57,765,159	852,445,506	5,949,853	49,371,692
chr7		30	0.101	0.101	1	58,435,236	145,928,194	4,204,763,652	3,130,837	36,156,111	2,582,046	33,265,563	20,869,020	2,117,158	1,545,769	2,449,952	2,258,395	63.70%	53,275,291	754,028,870	5,093,752	44,875,267
chr8		30	0.101	0.101	1	54,753,748	136,550,681	3,958,352,980	2,820,518	32,185,624	2,235,745	29,912,877	18,400,652	1,880,320	1,404,514	2,147,778	2,008,621	62.40%	49,862,276	709,235,349	4,733,388	42,428,679
chr9		30	0.101	0.101	1	42,257,619	106,409,875	3,055,335,248	2,261,496	26,681,711	1,956,970	24,783,328	15,564,922	1,608,252	1,169,438	1,857,841	1,716,117	63.19%	39,729,359	559,689,015	3,524,531	33,615,385
chr10		30	0.101	0.101	1	50,223,334	123,779,493	3,624,682,508	2,962,824	34,112,958	2,302,356	31,602,431	20,150,666	1,939,327	1,454,126	2,207,133	2,064,562	63.88%	46,599,031	669,265,387	4,122,255	39,819,139
chr11		30	0.101	0.101	1	50,934,510	125,703,299	3,682,718,040	2,653,006	31,287,256	2,228,922	29,050,289	17,420,648	1,854,920	1,360,990	2,133,125	1,980,112	60.79%	47,408,055	677,865,234	4,111,076	40,343,250
chr12		30	0.101	0.101	1	50,252,942	125,732,275	3,633,167,968	2,645,714	30,432,804	2,198,055	28,291,997	17,601,782	1,818,899	1,335,170	2,096,653	1,942,651	63.38%	46,426,021	654,198,341	4,275,688	39,347,251
chr13		30	0.101	0.101	1	38,856,727	92,647,750	2,808,654,616	1,921,218	21,228,728	1,397,183	19,701,481	11,999,231	1,189,425	913,037	1,344,395	1,266,185	62.23%	32,599,118	484,484,709	3,675,162	28,050,110
chr14		30	0.101	0.101	1	33,694,527	84,665,011	2,436,155,620	1,760,009	20,454,280	1,475,989	19,025,188	11,612,156	1,224,765	898,009	1,409,276	1,307,594	62.10%	31,367,823	441,527,640	2,862,340	26,640,352
chr15		30	0.101	0.101	1	29,206,355	75,302,002	2,111,994,376	1,642,574	19,446,345	1,422,670	18,094,905	11,414,208	1,171,151	857,287	1,352,437	1,248,642	63.81%	28,544,832	394,801,532	2,333,133	24,070,924
chr16		30	0.101	0.101	1	27,708,632	72,822,511	2,004,217,340	1,933,220	23,540,929	1,815,140	22,034,736	13,957,092	1,472,648	1,026,659	1,722,305	1,572,880	63.11%	28,883,290	388,239,483	2,025,397	24,225,667
chr17		30	0.101	0.101	1	25,112,365	72,651,718	1,816,695,468	1,761,412	22,200,028	1,914,074	20,876,733	13,038,854	1,509,874	1,007,547	1,801,098	1,619,518	62.58%	29,120,205	357,844,852	1,812,892	23,834,470
chr18		30	0.101	0.101	1	30,207,064	72,263,145	2,181,136,208	1,597,776	18,041,167	1,184,617	16,666,061	10,327,394	1,005,618	765,743	1,138,907	1,070,011	63.01%	26,205,883	388,422,265	2,641,981	22,515,540
chr19		30	0.101	0.101	1	16,508,610	52,761,330	1,171,867,876	1,451,883	18,529,565	1,782,643	17,113,019	10,655,986	1,333,583	797,544	1,664,290	1,442,740	62.02%	21,915,394	240,819,902	1,066,070	17,076,810
chr20		30	0.101	0.101	1	22,997,195	57,677,280	1,663,335,072	1,392,591	17,186,725	1,260,059	16,073,301	9,742,617	1,052,084	766,177	1,208,745	1,121,331	61.25%	22,978,812	323,120,976	1,675,496	19,662,121
chr21		30	0.101	0.101	1	13,452,782	32,981,512	970,145,120	801,755	9,291,111	650,757	8,556,444	5,381,612	545,659	401,053	624,377	582,121	63.63%	12,116,690	174,544,433	1,173,227	10,285,368
chr22		30	0.101	0.101	1	10,906,288	32,177,040	789,252,388	847,579	11,137,694	958,686	10,498,688	6,644,642	764,242	514,195	904,591	818,702	63.14%	13,562,203	164,575,608	708,810	11,157,591
chrX		30	0.101	0.101	1	26,183,187	135,071,481	1,892,751,928	1,201,964	13,189,882	1,836,146	12,246,521	7,362,730	1,138,648	493,083	1,639,850	1,249,843	61.22%	46,420,044	332,207,218	2,262,490	29,390,825
chrY		30	0.101	0.101	1	2,919,452	14,061,667	211,043,596	164,229	1,566,682	165,042	1,331,043	854,899	99,459	46,234	141,187	108,054	64.09%	4,612,726	37,703,546	231,041	2,800,754
chrM		30	0.101	0.101	1	1,235,800	16,449	87,653,540	281,146	2,323,051	793	2,518,746	6,755	710	698	763	754	0.50%	5,855	13,234,169	34,726	5,382
nuclear genome totals						1,050,881,599	2,693,703,851	75,771,870,852	56,135,811	647,438,468	47,213,781	598,831,756	371,158,370	38,580,430	27,813,495	44,890,613	41,222,653		993,491,504	###########	89,369,780	830,690,651

Summaries:

Mean reads per position covered (both strands)	28.13	Mean reads per CpG (each strand)	12.68	High-confidence CpGs in sample	47,213,781
Total bp covered in hg18 (from UCSC)	2,858,034,764	Total CpGs in reference	56,328,604	Percent with possible sequence variant	3.8%
Percentage of hg18 covered	94.3%	Total reference CpGs covered:	49,057,680		
		Percent of reference covered (1 read including	87.09%		
		Percent of reference covered (5 read including	79.69%		
Mean CpH retention (nuclear genome)	0.65%	Percent of reference covered (10 reads includi	73.18%	Mean CpG methylation (nuclear genome)	61.98%
Mean CpH retention (mitochondrial genome)	0.26%			Mean CpG methylation (mitochondrial genom	0.27%

Colon Tumor Excluding Duplicates

											High-confidence CpGs								High-confidence CpHs			
Chrom	minMapQ	Cytosine maximum oppisite strand non-G	CpG maximum next base non-G	Minimum converted CpHs for 5' conversion filter		total uniquely mapped reads	total genomic coordinates covered	total sequence (bp)	CpGs removed 5' conversion filter	CpGs retained 5' conversion filter	Unique bases covered (both strands)	Total C/T sequence, single-stranded (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)
chr1		30	0.101	0.101	1	70,975,448	211,263,806	5,133,080,880	3,648,862	42,972,568	3,788,141	39,870,425	24,631,906	2,958,704	1,937,249	3,561,857	3,172,941	62.51%	79,370,902	939,972,672	5,830,874	64,775,480
chr2		30	0.101	0.101	1	82,660,673	226,476,113	5,977,570,076	4,146,302	47,830,228	3,678,586	44,367,204	27,823,029	3,070,717	2,231,039	3,517,907	3,271,847	63.59%	82,819,397	1,073,079,596	7,205,887	70,334,318
chr3		30	0.101	0.101	1	69,138,284	187,726,483	4,999,343,008	3,246,780	36,653,677	2,795,724	33,844,157	20,999,727	2,330,080	1,696,634	2,671,192	2,483,626	63.26%	68,109,123	888,764,991	6,090,150	57,983,792
chr4		30	0.101	0.101	1	66,089,418	179,856,705	4,778,177,208	3,028,232	33,194,822	2,498,546	30,566,690	18,616,666	2,081,664	1,519,412	2,388,598	2,224,032	62.18%	62,720,082	818,055,218	6,279,890	53,153,942
chr5		30	0.101	0.101	1	62,007,059	168,449,241	4,483,655,652	2,955,761	33,492,800	2,550,273	31,041,320	18,821,892	2,129,324	1,554,223	2,437,744	2,268,539	61.98%	60,736,458	794,062,746	5,510,744	51,676,781
chr6		30	0.101	0.101	1	58,240,774	160,186,298	4,211,379,180	2,887,049	32,777,945	2,519,151	30,354,812	18,886,517	2,100,156	1,525,693	2,409,117	2,240,861	63.33%	57,712,657	743,490,283	5,215,142	48,853,525
chr7		30	0.101	0.101	1	51,295,989	145,839,672	3,709,518,640	2,712,507	31,489,484	2,579,371	29,262,772	18,355,856	2,088,837	1,457,351	2,440,147	2,229,241	63.71%	53,217,513	665,072,721	4,488,170	44,359,962
chr8		30	0.101	0.101	1	48,485,101	136,488,737	3,506,127,312	2,475,528	28,471,119	2,233,794	26,456,889	16,281,827	1,857,993	1,329,593	2,140,868	1,986,950	62.41%	49,816,729	628,551,734	4,188,531	41,972,306
chr9		30	0.101	0.101	1	37,354,464	106,305,755	2,701,592,060	1,987,871	23,652,944	1,954,796	21,988,096	13,808,166	1,587,320	1,101,904	1,850,813	1,695,360	63.21%	39,679,802	495,568,582	3,110,218	33,237,480
chr10		30	0.101	0.101	1	43,919,371	123,713,843	3,176,525,088	2,459,837	29,128,166	2,300,287	27,085,526	17,129,894	1,916,618	1,375,490	2,200,154	2,042,631	63.89%	46,552,820	588,472,066	3,607,165	39,395,035
chr11		30	0.101	0.101	1	44,925,937	125,641,750	3,249,218,804	2,322,139	27,623,252	2,227,099	25,666,157	15,403,536	1,831,844	1,284,548	2,125,913	1,957,109	60.80%	47,362,328	598,996,719	3,622,360	39,905,751
chr12		30	0.101	0.101	1	44,487,582	125,679,633	3,217,243,932	2,327,308	26,968,622	2,196,165	25,072,150	15,609,839	1,795,230	1,259,655	2,089,138	1,919,050	63.39%	46,382,145	579,943,797	3,778,433	38,907,509
chr13		30	0.101	0.101	1	34,274,204	92,616,285	2,478,101,684	1,679,049	18,687,866	1,396,032	17,341,667	10,568,126	1,176,084	868,101	1,340,227	1,253,845	62.25%	32,571,525	427,582,369	3,244,264	27,763,043
chr14		30	0.101	0.101	1	30,100,434	84,625,607	2,176,821,956	1,562,588	18,298,173	1,475,322	17,020,760	10,390,758	1,212,155	853,905	1,405,418	1,294,475	62.11%	31,347,461	394,941,691	2,559,007	26,407,623
chr15		30	0.101	0.101	1	26,184,799	75,261,456	1,893,925,420	1,462,459	17,431,305	1,422,158	16,224,940	10,234,352	1,159,524	816,530	1,348,540	1,236,113	63.82%	28,527,361	354,276,486	2,091,347	23,874,692
chr16		30	0.101	0.101	1	24,681,383	72,771,688	1,785,682,828	1,658,388	20,696,346	1,813,810	19,402,205	12,191,908	1,456,545	970,285	1,717,095	1,554,485	63.11%	28,862,191	347,632,593	1,798,178	24,012,208
chr17		30	0.101	0.101	1	22,737,907	72,616,832	1,645,264,720	1,589,083	20,160,965	1,913,425	18,967,107	11,838,044	1,492,209	947,032	1,795,261	1,598,883	62.60%	29,102,777	324,571,269	1,639,552	23,598,789
chr18		30	0.101	0.101	1	26,617,914	72,241,897	1,924,803,952	1,370,084	15,682,141	1,184,278	14,575,035	9,039,593	997,140	733,749	1,136,311	1,061,706	63.03%	26,194,865	342,589,427	2,354,022	22,357,778
chr19		30	0.101	0.101	1	14,592,192	52,737,469	1,056,225,376	1,253,423	16,372,627	1,782,259	15,493,454	9,613,434	1,313,924	737,433	1,658,701	1,418,704	62.02%	21,903,610	217,301,202	956,623	16,865,535
chr20		30	0.101	0.101	1	20,610,284	57,661,953	1,491,044,588	1,240,374	15,419,197	1,259,892	14,437,926	8,755,358	1,042,835	729,042	1,205,989	1,111,371	61.26%	22,969,066	290,095,257	1,499,806	19,513,523
chr21		30	0.101	0.101	1	11,842,429	32,968,769	856,496,128	701,747	8,205,689	650,716	7,625,247	4,800,098	540,899	381,947	622,985	577,205	63.65%	12,109,524	154,070,673	1,037,171	10,201,731
chr22		30	0.101	0.101	1	9,888,433	32,153,963	715,742,476	761,370	10,140,061	958,347	9,556,075	6,045,550	755,605	483,240	901,920	809,017	63.16%	13,552,482	149,465,548	640,053	11,049,114
chrX		30	0.101	0.101	1	24,727,181	135,038,820	1,787,667,908	1,124,894	12,400,466	1,836,475	11,519,507	6,931,382	1,117,348	442,442	1,633,847	1,218,136	61.24%	46,430,397	313,647,793	2,143,820	28,903,422
chrY		30	0.101	0.101	1	2,438,342	14,029,454	176,311,764	119,297	1,202,304	164,513	1,071,453	688,661	97,445	41,516	140,420	105,392	64.11%	4,609,338	31,532,805	198,462	2,745,303
chrM		30	0.101	0.101	1	32,269	15,378	2,378,452	6,204	55,105	701	59,420	99	656	632	687	677	0.14%	5,249	419,974	805	4,841
nuclear genome totals						928,275,602	2,692,352,229	67,131,520,640	48,720,932	568,952,767	47,179,160	528,811,574	327,466,119	38,110,200	26,278,013	44,740,162	40,731,519		992,660,553	###########	79,089,869	821,848,642

Summaries:

Mean reads per position covered (both strands)	24.93	Mean reads per CpG (each strand)	11.21	High-confidence CpGs in sample	47,179,160
Total bp covered in hg18 (from UCSC)	2,858,034,764	Total CpGs in reference	56,328,604	Percent with possible sequence variant	3.8%
Percentage of hg18 covered	94.2%	Total reference CpGs covered:	49,057,680		
		Percent of reference covered (1 read including	87.09%		
		Percent of reference covered (5 read including	79.43%		
Mean CpH retention (nuclear genome)	0.65%	Percent of reference covered (10 reads includi	72.31%	Mean CpG methylation (nuclear genome)	61.92%
Mean CpH retention (mitochondrial genome)	0.19%			Mean CpG methylation (mitochondrial genom	0.17%

Supplementary Table 2. Bisulfite-seq sample details (cont.)

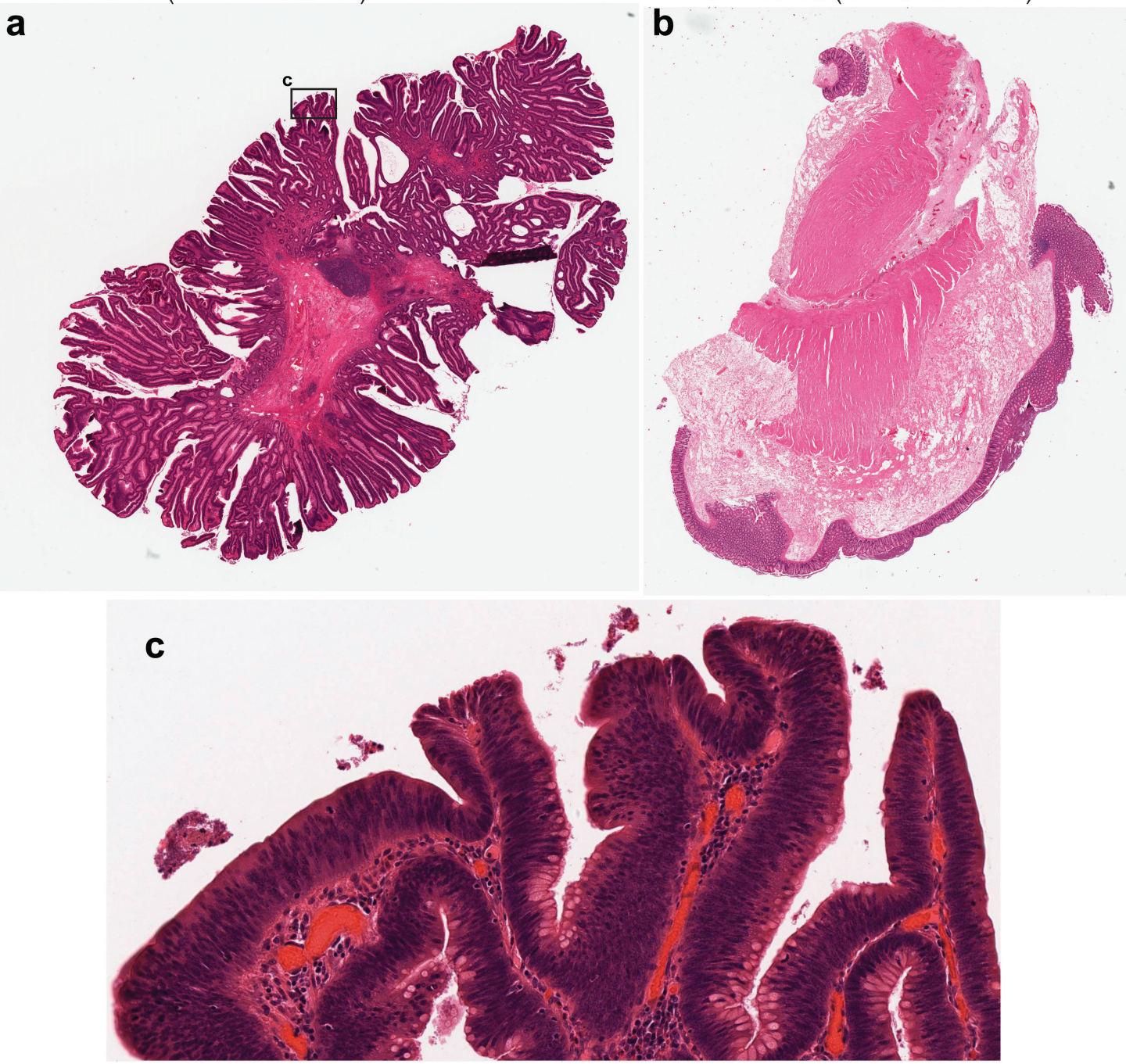
Cytosines with ambiguous context											Possible non-cytosines in sample genome								
Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	cphConversionPreFilter
46,490,825	77,300,669	72,775,588	0.68%	734,621	5,782,107	581,166	584,219	113,061	715,259	638,157	9.19%	570,236	7,743,071	317,916	459,436	327,027	556,171	488,270	4.69%	98.90%
53,561,480	81,185,086	77,996,188	0.75%	698,583	5,813,975	653,339	570,020	133,326	682,942	625,777	9.45%	578,071	8,756,980	366,133	483,499	366,973	566,762	512,523	4.62%	98.81%
44,262,253	66,852,315	64,273,703	0.77%	557,649	4,643,206	491,444	455,311	107,458	545,455	500,305	9.00%	476,094	7,171,247	289,929	399,214	303,915	467,127	423,248	4.45%	98.79%
40,397,567	61,512,220	59,100,534	0.86%	514,437	4,333,489	466,294	419,266	99,616	502,946	461,024	9.00%	459,515	7,113,146	280,200	384,976	292,725	450,475	408,987	4.41%	98.67%
39,480,860	59,563,002	57,262,448	0.78%	499,195	4,121,968	427,799	407,636	95,281	488,310	448,084	9.01%	422,830	6,338,311	251,499	354,890	269,436	414,662	376,173	4.45%	98.77%
37,034,193	56,599,337	54,336,399	0.78%	487,673	4,215,822	499,138	398,027	93,960	477,290	437,653	9.43%	417,166	6,385,305	279,030	348,017	263,333	408,890	369,752	4.64%	98.77%
33,125,606	51,844,964	49,327,024	0.74%	475,415	3,876,669	446,567	383,637	86,503	462,571	417,498	10.07%	381,901	5,463,495	247,216	312,664	231,855	372,959	332,234	4.97%	98.81%
31,423,898	48,868,579	46,779,396	0.75%	430,954	3,494,284	369,464	350,660	78,438	422,187	386,134	9.23%	350,674	5,064,606	210,914	292,436	218,081	344,134	310,078	4.56%	98.82%
24,848,927	38,675,776	36,917,228	0.69%	354,185	2,871,379	323,808	286,231	62,606	344,649	312,456	9.94%	280,260	4,008,390	175,562	230,424	171,067	273,716	244,411	4.83%	98.88%
29,753,386	45,572,090	43,667,419	0.69%	413,882	3,575,521	465,222	337,441	79,349	404,514	369,656	10.10%	329,036	4,991,621	246,473	272,827	204,340	321,904	289,219	5.04%	98.91%
29,978,362	46,365,646	44,357,877	0.67%	417,483	3,404,711	361,405	339,361	76,594	408,118	372,049	9.37%	327,541	4,801,590	193,984	271,257	202,577	320,553	287,715	4.59%	98.91%
29,060,148	45,389,814	43,372,471	0.72%	408,462	3,287,878	358,744	330,011	72,768	398,708	362,474	9.64%	331,480	4,700,016	200,080	272,803	202,301	324,094	289,710	4.82%	98.84%
21,202,483	32,004,515	30,822,047	0.86%	267,907	2,212,804	233,200	218,873	51,587	262,314	241,415	9.17%	237,458	3,578,436	142,476	199,729	151,708	232,885	211,812	4.45%	98.68%
19,672,607	30,678,310	29,328,954	0.72%	276,137	2,223,514	236,245	224,019	49,349	269,839	245,686	9.46%	222,991	3,155,903	132,777	183,849	136,383	218,389	195,394	4.70%	98.85%
17,681,235	27,818,344	26,496,702	0.66%	256,794	2,071,581	232,289	207,091	45,952	250,082	226,037	9.83%	203,099	2,811,011	126,230	165,639	121,826	198,266	176,031	4.90%	98.94%
17,447,286	28,039,258	26,533,284	0.56%	290,481	2,394,683	320,423	234,633	50,804	282,600	253,431	11.42%	205,788	2,807,198	159,600	166,013	118,962	200,461	176,314	5.75%	99.06%
16,411,259	28,157,400	26,351,830	0.55%	310,857	2,365,589	281,941	246,681	45,977	301,706	267,218	11.03%	212,380	2,613,067	138,148	166,915	114,770	206,280	177,399	5.82%	99.09%
16,955,970	25,705,168	24,733,084	0.77%	221,160	1,910,081	223,891	180,890	42,668	216,445	199,108	9.58%	186,202	2,950,955	124,361	156,134	118,638	182,458	165,403	4.56%	98.80%
10,786,600	20,994,448	19,045,639	0.47%	269,647	1,985,109	260,225	208,214	32,711	260,631	223,020	12.35%	164,967	1,831,732	106,509	122,043	75,562	158,849	130,298	6.60%	99.21%
14,579,075	22,510,744	21,541,129	0.57%	212,685	1,752,161	205,471	173,395	38,958	208,137	189,831	10.31%	153,330	2,178,888	101,027	126,634	93,792	149,988	134,048	5.19%	99.05%
7,587,957	11,829,095	11,291,090	0.74%	113,669	1,019,499	137,866	92,457	21,838	110,926	101,024	10.92%	89,327	1,463,353	77,090	73,583	54,899	87,238	78,138	5.31%	98.82%
7,666,583	13,113,038	12,288,915	0.47%	149,929	1,141,967	141,019	119,747	22,542	145,518	129,467	11.60%	93,930	1,134,543	64,640	73,511	50,039	90,883	77,971	6.28%	99.21%
12,781,123	43,226,658	34,229,727	0.76%	498,572	3,125,980	168,428	362,068	22,504	478,972	368,751	5.60%	360,538	2,723,289	82,768	236,507	109,301	344,535	251,726	3.41%	98.79%
1,266,730	4,057,091	3,162,612	0.73%	51,330	434,158	62,397	36,960	4,574	48,262	36,771	6.77%	34,595	435,355	26,133	22,652	11,489	32,741	24,359	4.16%	98.95%
5,324	5,726	5,670	0.29%	72	22,014	42	62	48	71	66	0.34%	60	129,141	200	53	53	60	60	0.15%	99.50%
603,456,413	967,863,567	915,991,288		8,911,707	72,058,135	7,947,785	7,166,848	1,528,424	8,688,381	7,813,026		7,089,409	100,221,508	4,340,695	5,775,652	4,210,999	6,924,420	6,131,213		

Cytosines with ambiguous context											Possible non-cytosines in sample genome								
Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	Unique bases covered (both strands)	Total C/T sequence (bp)	Total C sequence (bp)	Unique bases with at least 5 reads (single-stranded)	Unique bases with at least 10 reads (single-stranded)	Unique bases with at least 5 reads (double-stranded)	Unique bases with at least 10 reads (double-stranded)	Mean Per cent methylation	cphConversionPreFilter
43,832,073	77,098,986	72,046,972	0.67%	759,318	5,584,794	530,996	602,330	96,030	739,597	656,392	8.87%	576,762	6,878,514	286,474	460,090	314,424	561,917	488,402	4.68%	98.90%
50,914,402	81,007,298	77,435,723	0.75%	724,620	5,644,402	573,602	590,287	115,606	708,682	647,096	9.12%	584,351	7,682,050	320,390	484,794	355,276	572,401	513,262	4.62%	98.82%
42,110,417	66,715,572	63,825,848	0.76%	579,408	4,533,349	438,249	472,503	93,410	566,911	518,204	8.65%	481,133	6,372,467	253,567	400,067	294,425	471,650	423,844	4.43%	98.79%
38,419,034	61,381,310	58,679,219	0.85%	534,087	4,187,479	412,979	434,286	86,559	522,311	476,942	8.67%	463,951	6,183,021	245,429	385,303	283,031	454,423	409,071	4.39%	98.67%
37,569,441	59,440,216	56,862,833	0.77%	517,853	4,030,018	386,568	422,229	82,517	506,783	463,231	8.66%	427,456	5,651,851	224,117	355,792	261,149	418,861	376,860	4.43%	98.78%
35,192,035	56,478,926	53,940,430	0.78%	506,349	3,957,149	405,847	412,504	81,658	495,683	452,715	9.12%	421,769	5,518,379	232,663	348,852	255,047	413,040	370,214	4.63%	98.77%
31,385,176	51,708,144	48,899,002	0.74%	491,760	3,769,214	402,658	395,943	74,419	478,631	429,946	9.73%	385,897	4,855,530	219,049	312,814	223,335	376,435	332,191	4.95%	98.81%
29,807,858	48,767,034	46,424,006	0.75%	447,583	3,438,078	338,409	363,546	67,940	438,638	399,394	8.91%	354,524	4,530,970	188,200	293,067	210,638	347,561	310,454	4.53%	98.82%
23,545,264	38,578,711	36,617,201	0.69%	366,539	2,798,234	291,867	295,770	53,911	356,989	322,440	9.60%	283,006	3,562,169	154,394	230,683	164,845	276,137	244,378	4.83%	98.89%
28,225,865	45,470,498	43,337,118	0.68%	429,008	3,333,387	360,301	348,919	68,239	419,403	381,914	9.74%	333,220	4,223,860	194,726	273,753	197,610	325,756	289,994	5.01%	98.91%
28,428,928	46,260,949	44,013,147	0.67%	432,250	3,321,281	328,883	350,569	65,756	422,576	383,701	9.04%	331,204	4,224,603	173,325	271,872	195,829	323,833	288,160	4.57%	98.92%
27,536,078	45,286,294	43,024,533	0.71%	423,393	3,224,150	326,299	341,592	62,781	413,463	374,339	9.29%	335,448	4,213,704	180,249	273,606	195,251	327,667	290,343	4.81%	98.85%
20,176,212	31,942,845	30,613,822	0.85%	278,721	2,174,388	213,094	227,302	44,751	272,956	250,294	8.84%	239,664	3,181,952	126,556	199,739	146,537	234,863	211,795	4.43%	98.68%
18,776,533	30,624,661	29,135,769	0.71%	284,795	2,182,579	217,990	230,812	42,871	278,360	252,625	9.15%	224,700	2,837,292	119,352	183,846	131,901	219,910	195,160	4.68%	98.86%
16,898,041	27,767,564	26,322,557	0.66%	264,261	2,023,669	212,947	212,867	39,820	257,388	231,973	9.53%	204,481	2,532,422	112,897	165,564	117,813	199,415	175,693	4.87%	98.95%
16,615,177	27,982,265	26,333,873	0.56%	298,390	2,285,285	275,545	240,433	43,876	290,418	259,524	11.09%	207,084	2,473,253	129,441	165,732	114,463	201,564	175,812	5.72%	99.06%
15,537,250	28,096,320	26,124,312	0.55%	318,904	2,316,657	260,982	252,326	38,911	309,573	273,158	10.71%	214,551	2,362,148	126,041	167,154	110,377	208,220	177,358	5.82%	99.09%
16,273,958	25,671,273	24,607,355	0.77%	228,556	1,792,094	184,459	186,632	37,406	223,740	205,354	9.26%	187,259	2,477,002	103,082	156,013	115,085	183,371	165,053	4.60%	98.79%
10,088,895	20,939,255	18,812,819	0.47%	275,325	1,908,420	235,289	212,060	27,020	266,199	226,259	12.05%	166,931	1,597,853	96,701	122,218	71,960	160,566	130,152	6.65%	99.21%
13,944,800	22,478,019	21,415,248	0.57%	218,958	1,693,245	182,064	178,349	33,825	214,321	194,842	9.93%	154,827	1,947,674	91,025	126,980	90,916	151,322	134,241	5.20%	99.06%
7,256,918	11,807,485	11,219,862	0.74%	117,068	943,903	115,239	95,127	18,873	114,265	103,761	10.54%	89,859	1,194,669	61,228	73,439	53,088	87,646	77,908	5.25%	98.82%
7,249,196	13,083,308	12,184,032	0.47%	153,520	1,119,939	131,541	122,382	19,108	149,066	132,224	11.27%	95,018	1,019,927	59,294	73,781	48,031	91,815	78,109	6.26%	99.22%
11,537,705	43,139,844	33,511,516	0.76%	510,180	3,116,348	160,489	369,071	16,160	490,125	373,155	5.49%	364,423	2,582,138	78,909	235,694	100,269	348,023	249,632	3.41%	98.79%
1,140,862	4,038,157	3,085,775	0.73%	52,079	352,156	31,466	37,296	3,551	48,978	36,887	6.59%	34,534	285,674	13,710	22,185	10,383	32,664	23,697	4.16%	98.89%
4,573	5,112	5,034	0.21%	37	576	3	24	8	33	30	1.22%	35	3,323	3	32	31	34	33	0.09%	99.64%
572,462,118	965,764,934	908,472,972		9,212,925	69,730,218	7,017,763	7,395,135	1,314,998	8,985,056	8,046,370		7,162,052	88,389,122	3,790,819	5,783,038	4,061,683	6,989,060	6,131,783		

Berman et al., 2011 Supplementary Figure 1

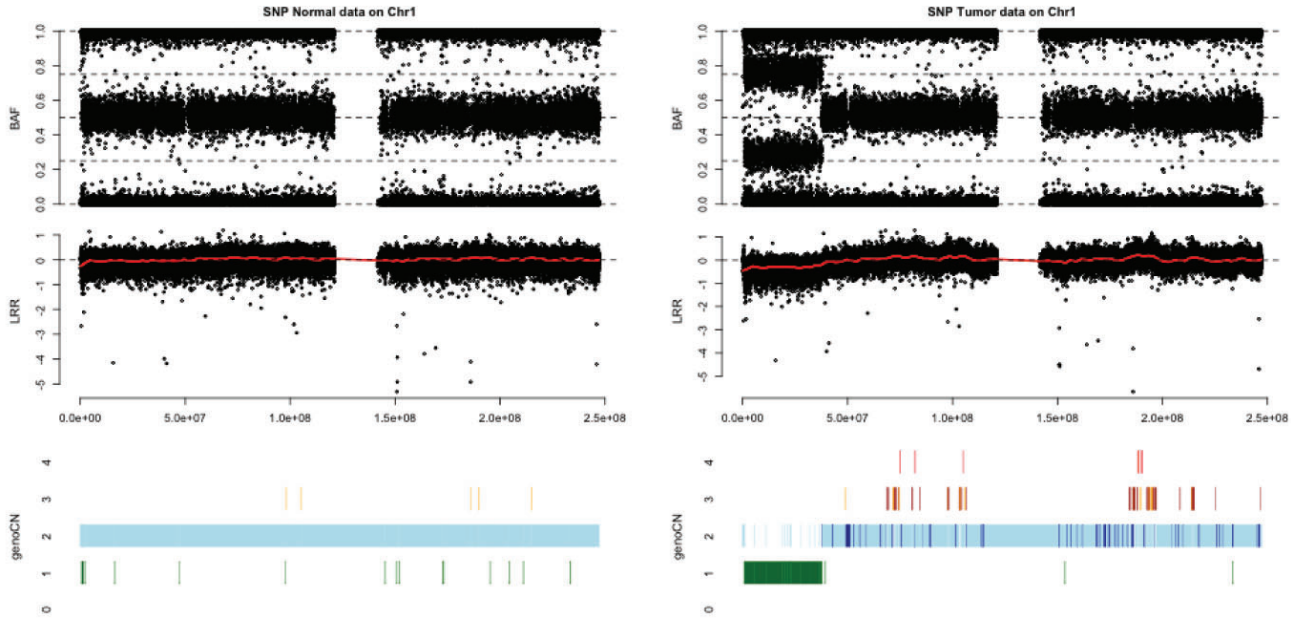
Colon tumor (OTB #B0068109)

Colon mucosa (OTB #B0068110)



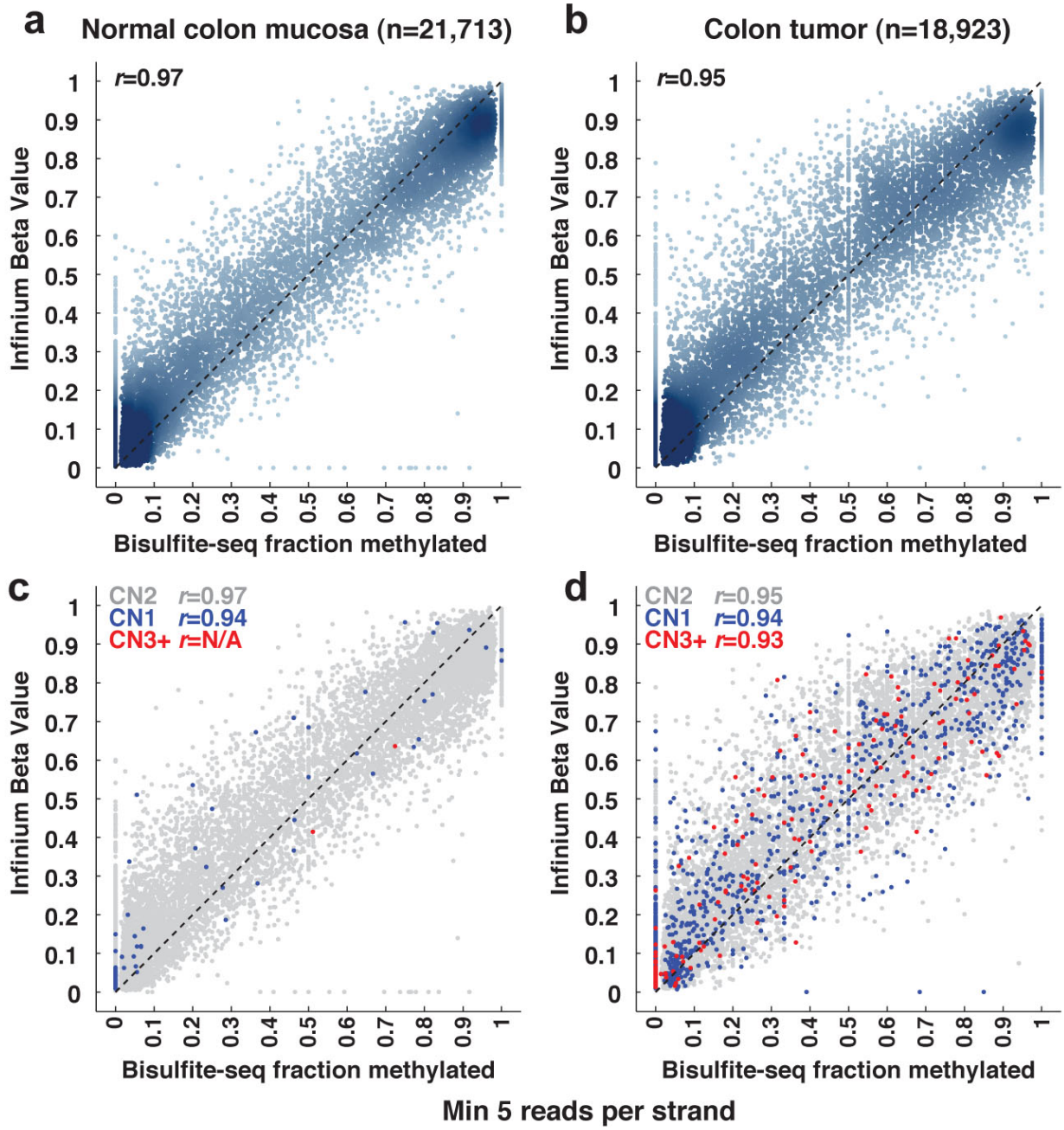
Supplementary figure 1: H&E histological stains show high tumor purity. Colon tumor (a,c) and normal colonic mucosa (b) images were taken directly from Ontario Tumor Bank. Tumor images show high content of transformed epithelial cells within sample, while colon mucosa primarily consists of normal epithelial nuclei.

Berman et al., 2011 Supplementary Figure 2



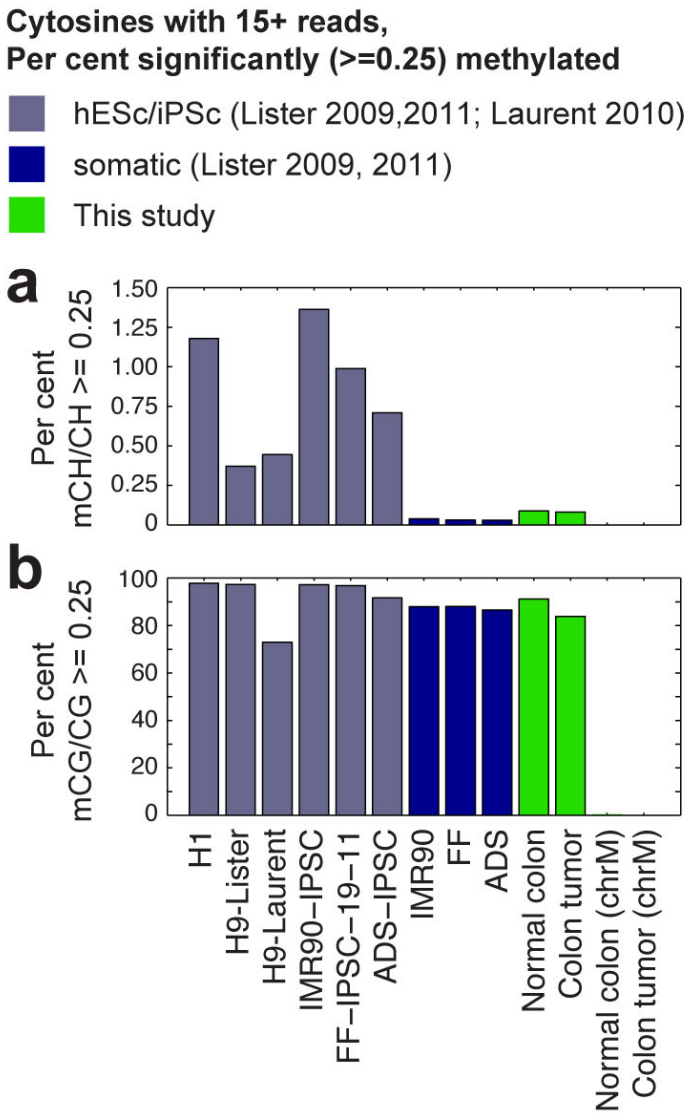
Supplementary Figure 2: Copy number alterations identified using the Infinium 1M SNP array. genoCNV9 plot showing chromosome 1, with a 100 Mbp deletion at the telomeric end of chr1p. Variant copy number calls (genoCN=0, 1, 3, 4) were used to compare Bisulfite-seq and the Infinium DNA methylation values in regions of altered copy number (Supplemental Figure 3).

Berman et al., 2011 Supplementary Figure 3



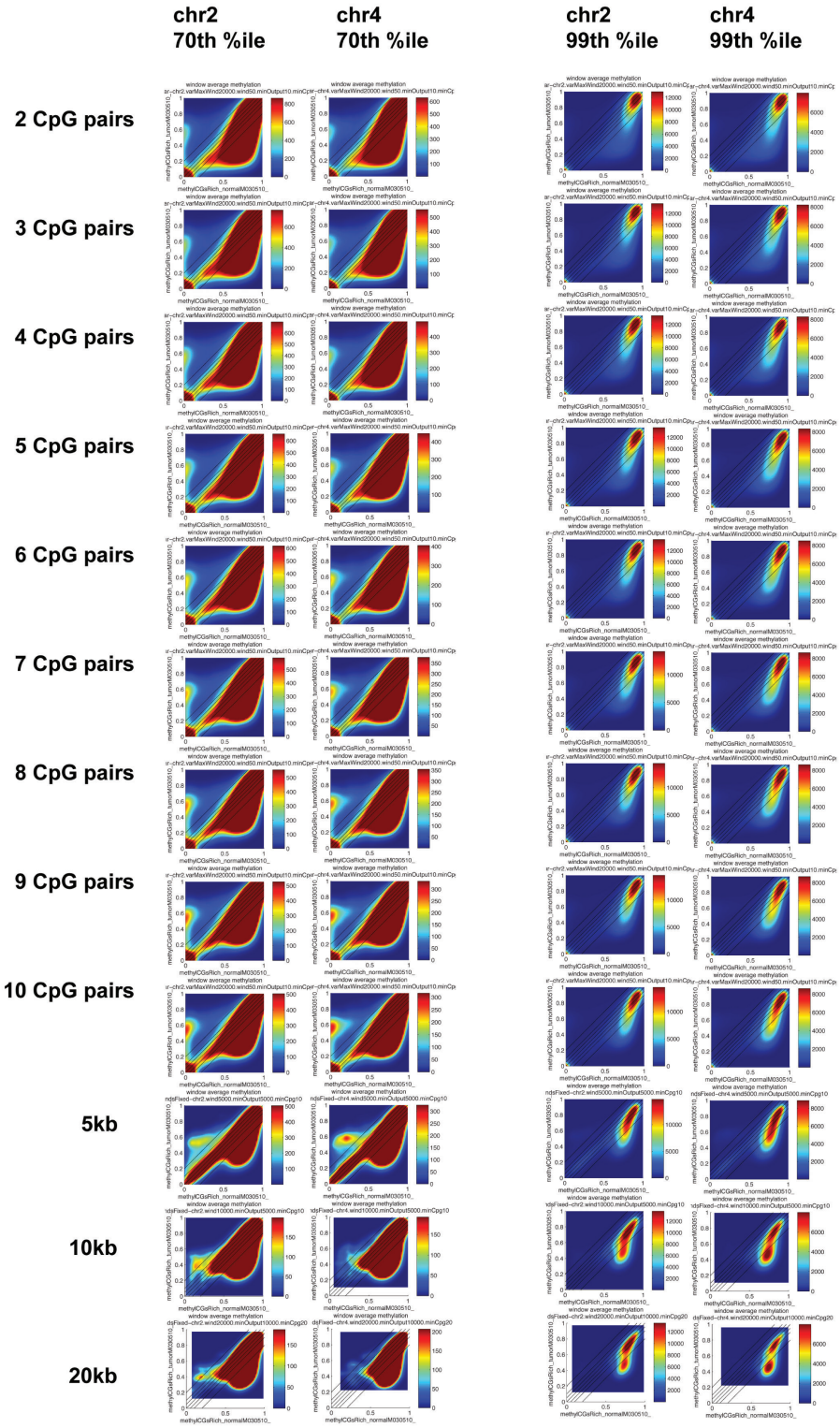
Supplementary Figure 3: Bisulfite-seq and Infinium methylation array give similar methylation levels, even in regions of altered copy number. Values are shown for 21,713 CpGs present on Infinium 27k array and with at least 5 reads on each strand of bisulfite-seq data for the normal colon mucosa (a,c), and 18,923 are shown for CpGs with at least 5 reads on each strand of bisulfite-seq data for the colon tumor (b,d). Top panels (a,b) are colored by feature density, while lower panels (c,d) are color-coded by copy number as determined using Infinium 1M SNP array (CN1, copy number loss, CN2 diploid copy number, CN3+ copy number gain).

Berman et al., 2011 Supplemental Figure 4

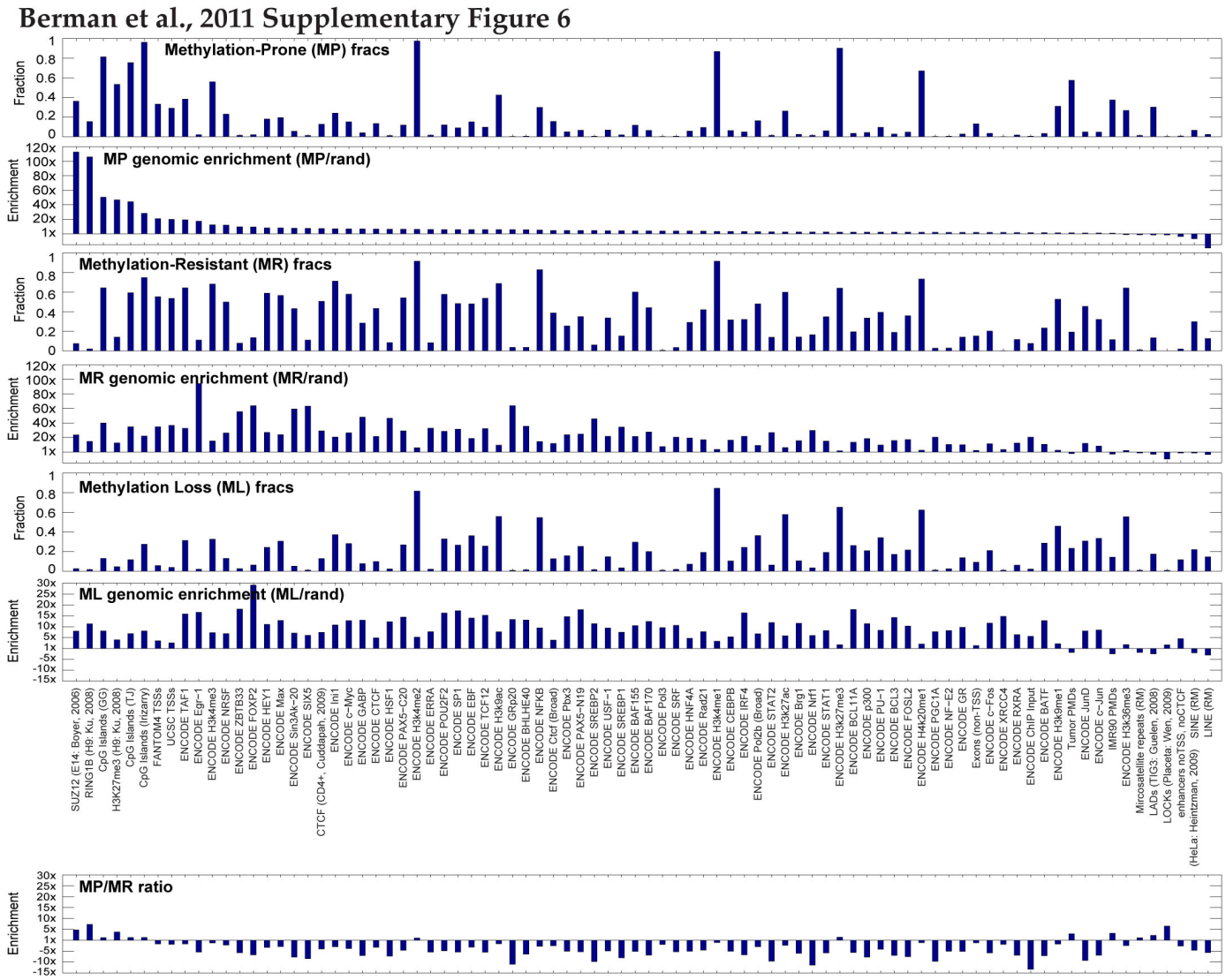


Supplementary Figure 4: Unlike hESCs, colon tumor and normal mucosa have extremely low levels of CpH methylation. Normal and colon methylomes (green) were compared to somatic (blue) and hESC/iPSC (gray) cell lines from ref2. Methylation levels were compared at cytosines with at least 15 uniquely positioned reads. Charts show the percentage of cytosines with at least 25% of reads methylated, for CpH (top) and CpG (bottom). For comparison, tumor and normal colon levels were also compared for the mitochondrial genome, which lacks methylation. While the tumor and normal samples do have about 0.1% of nuclear CpHs methylated, significantly more than the mitochondrial genome, this only represents about 30,000 individual CpHs. These levels are similar to those of somatic cell lines from ref2, and only about 10% of CpH levels in hESC/iPSC cells.

Berman et al., 2011 Supplemental Figure 5

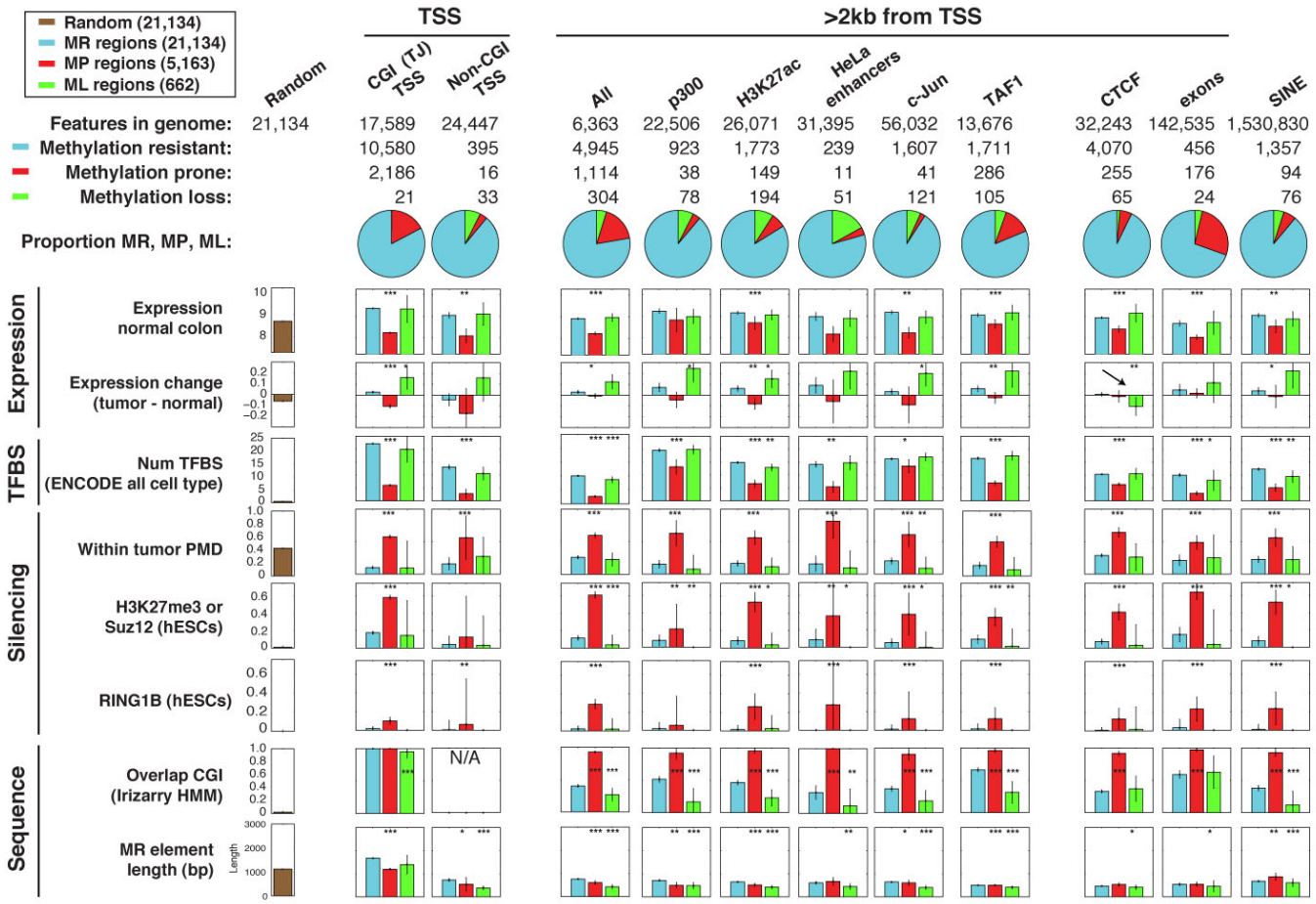


Supplementary Figure 5: The majority of the tumor and normal colon genomes are methylated. Small genomic windows reveal unmethylated regions in the normal colon, while large genomic windows reveal Partially Methylated Domains in the tumor. Moving averages of CpG methylation were taken for windows of varying sizes, and joint densities of tumor and normal averages were plotted as scatterclouds (each axis is split into 50 bins of 0.02 methylation units). For each window size, scatterclouds are shown for two different chromosomes (chr2 and chr4), and two different color scales – for the columns labeled "70th %ile", the darkest red corresponds to the 70th percentile of all 2,500 bin values, and likewise for the 99th percentile. In rows labeled 2-10 CpG pairs, variable window sizes with a minimum number of CpG dinucleotides were used. For rows labeled 5-20 kb, fixed window size were used.

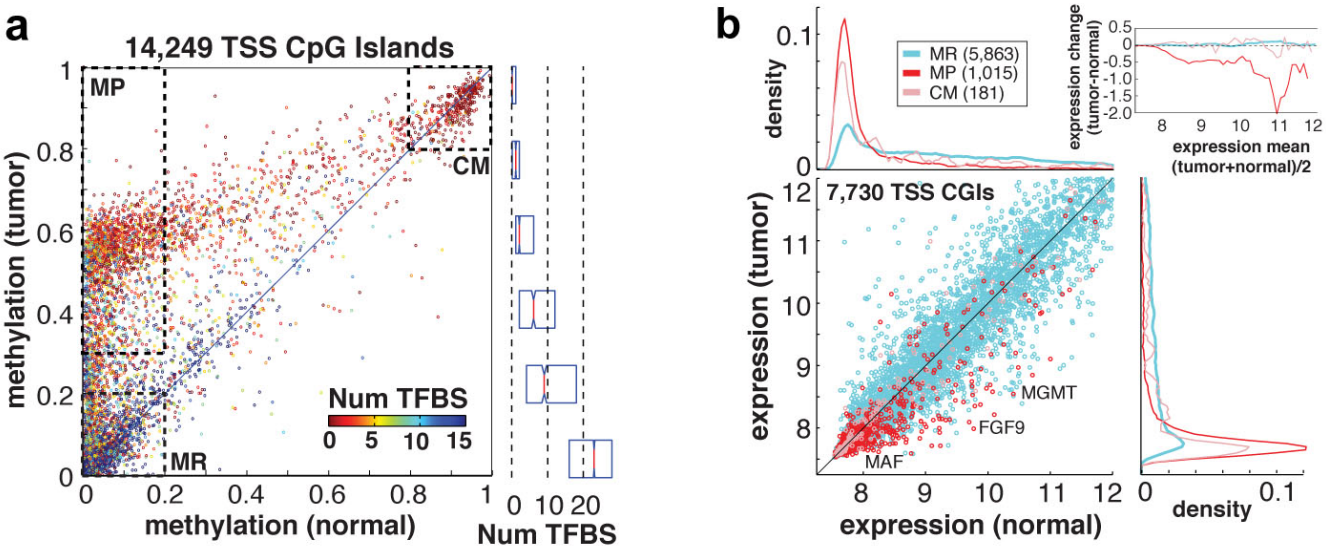


Supplementary Figure 6: Methylation-Resistant (MR) and Methylation Loss (ML) focal elements are enriched in similar sets of genomic features, distinct from those of Methylation-Prone (MP) elements. Comparison of each methylation class to ENCODE protein/DNA binding (ChIP-seq) data10 and other genomic features. “MP fracs” plot shows the absolute fraction of MP elements overlapping each feature, while “MP genomic enrichment” was determined by dividing this fraction by the average fraction within ten sets of size-matched, randomly generated genomic locations (shown as fold change). To reduce random variation for features with small numbers of items, a pseudocount of 5 was used to calculate each mean in the ratio. At bottom, the ratio between MP and MR elements is shown, also using a pseudocount of 5.

Berman et al., 2011 Supplementary Fig 7

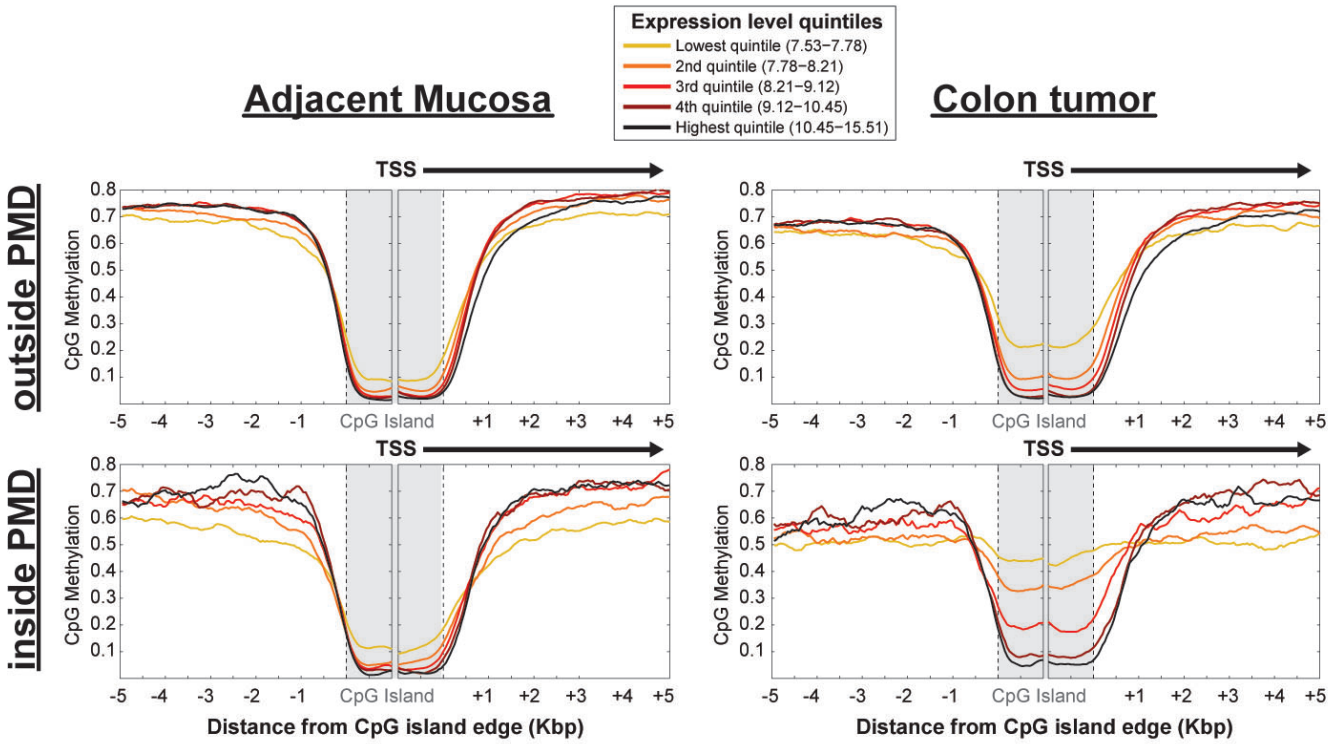


Supplementary Figure 7: Hypermethylated focal elements are more abundant at promoters than non-coding regulatory sequences, but all have low expression in normal colon, Polycomb Repressive Complex activity, and relative enrichment within PMDs. Sequence and expression properties are shown for selected classes of MP, MR, and ML elements that either overlap (left) or do not overlap (right) an annotated TSS (from either the UCSC knownGenes or FANTOM411 databases). Each plot shows mean values for the three methylation classes, along with 95% confidence intervals shown as error bars. MP and ML values were each compared to MR values using a two-sided Mann-Whitney U-test, and the significant outcomes were indicated by *(p<0.05), **(p<0.01), or ***(p<0.0001).



Supplementary Figure 8: Hypermethylated CpG Island promoters lack transcription factor binding across a number of cell types, and are predominantly unexpressed in normal colon, and are down-regulated in colon tumor. Mean DNA methylation values for 14,249 Takai-Jones CpG Islands overlapping a transcription start site (TSS) are shown for the colon tumor (y-axis) and the normal-adjacent sample (x-axis). CpG islands were categorized based on their DNA methylation levels - MR - DNA Methylation-Resistant, with methylation<0.2 for both normal and tumor tissues (n=10,631 loci), MP - DNA Methylation-Prone, with methylation levels <0.2 in the normal-adjacent tissue and >0.3 in the tumor sample (n=1,826 loci), CM - constitutively methylated, with methylation >0.8 in both samples (n=397 loci), and the remainder (n=1,395, labeled “misc”) with intermediate methylation levels. Point color corresponds to the number of ENCODE ChIP-seq transcription factor binding sites (TFBS) from any cell type that overlap the CpG island promoter. b Mean log2-transformed microarray expression levels for two technical replicates of normal-adjacent colon and colon tumor RNA, showing all genes corresponding to a CpG Island TSSs from panel (a). Each axis shows the expression density plots for each of the three DNA methylation classes. The majority of MP and CM genes having low expression in both tissues. Inset (MA-plot) shows expression change in each methylation class as a function of mean expression in both tissues (using a moving average of 0.1 expression units).

Berman et al., 2011 Supplementary Figure 9

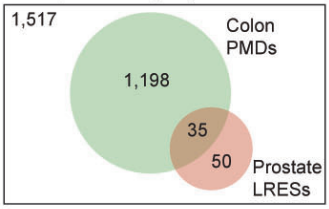


Supplementary Figure 9: Promoters for most genes with low expression become hypermethylated within tumor PMDs, but not the rest of the genome, and focal hypermethylation corresponds to regional hypomethylation both upstream and downstream of these PMD promoters. 7,153 CpG islands overlapping a promoter with valid microarray expression data were analyzed with respect to DNA methylation and expression. The top two plots show the 5,809 CGI-TSS located outside of tumor PMDs (Partially Methylated Domains from this study), and the bottom two plots show the 1,344 located inside of PMDs. In each plot, genes are stratified by expression level, and methylation levels are shown relative to the edges of the CpG island, with the left edge representing the 5' upstream region and the right edge representing the transcribed region. Methylation was averaged within sliding 150 bp windows in increments of 50 bp.

Berman et al., 2011 Supplemental Figure 10

a

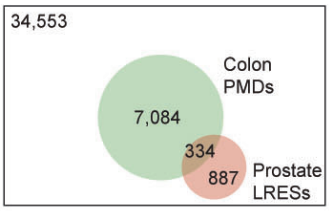
Base pair level (Mb)



	Not in LRES	In LRES	Total
In PMD	1,198	35	1,233
Not in PMD	1,517	50	1,567
Total	2,715	85	2,800

p=0.66 (two-tailed X-square test w/ Yates correction)

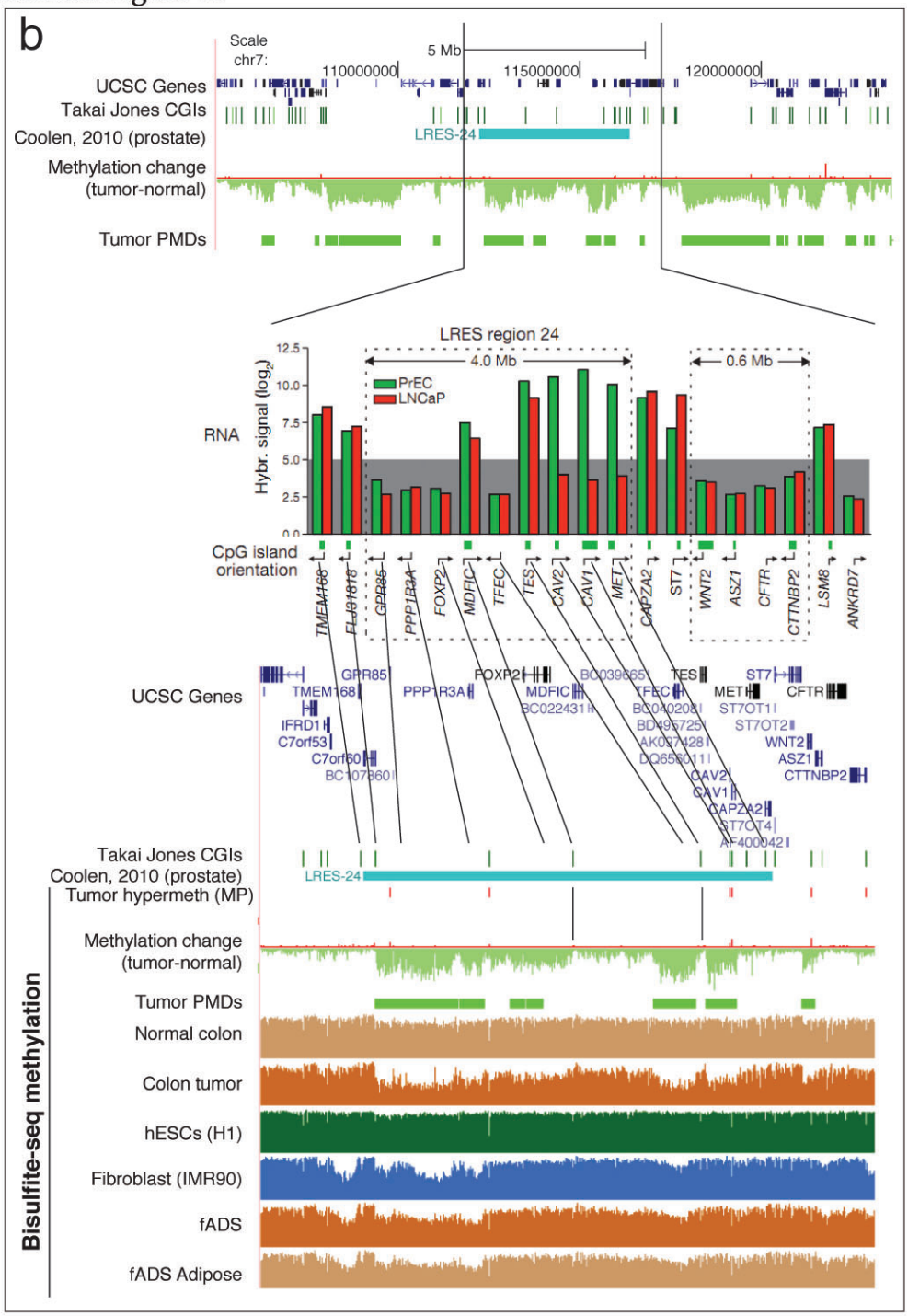
Gene level



	Not in LRES	In LRES	Total
In PMD	7,084	334	7,418
Not in PMD	34,553	887	35,440
Total	41,637	1,221	42,858

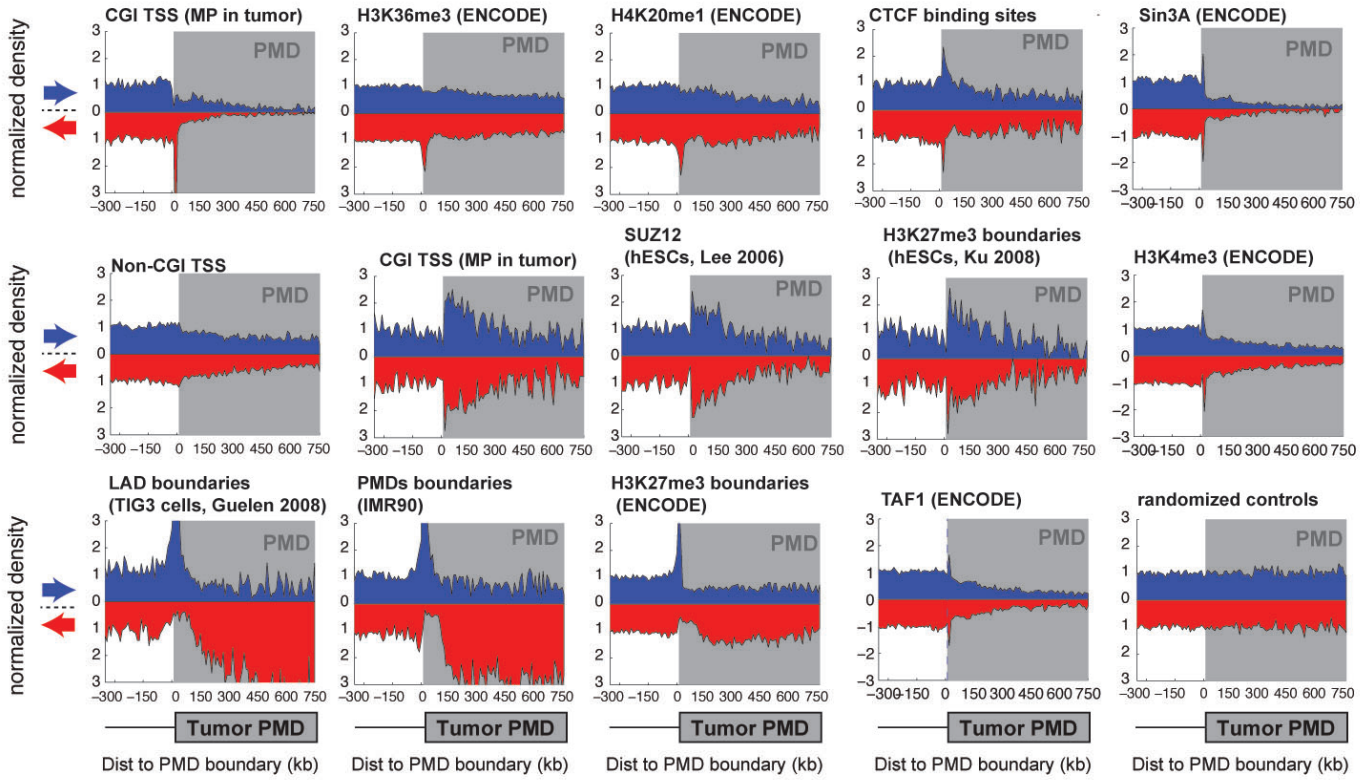
p<0.0001 (two-tailed X-square test w/ Yates correction)

b



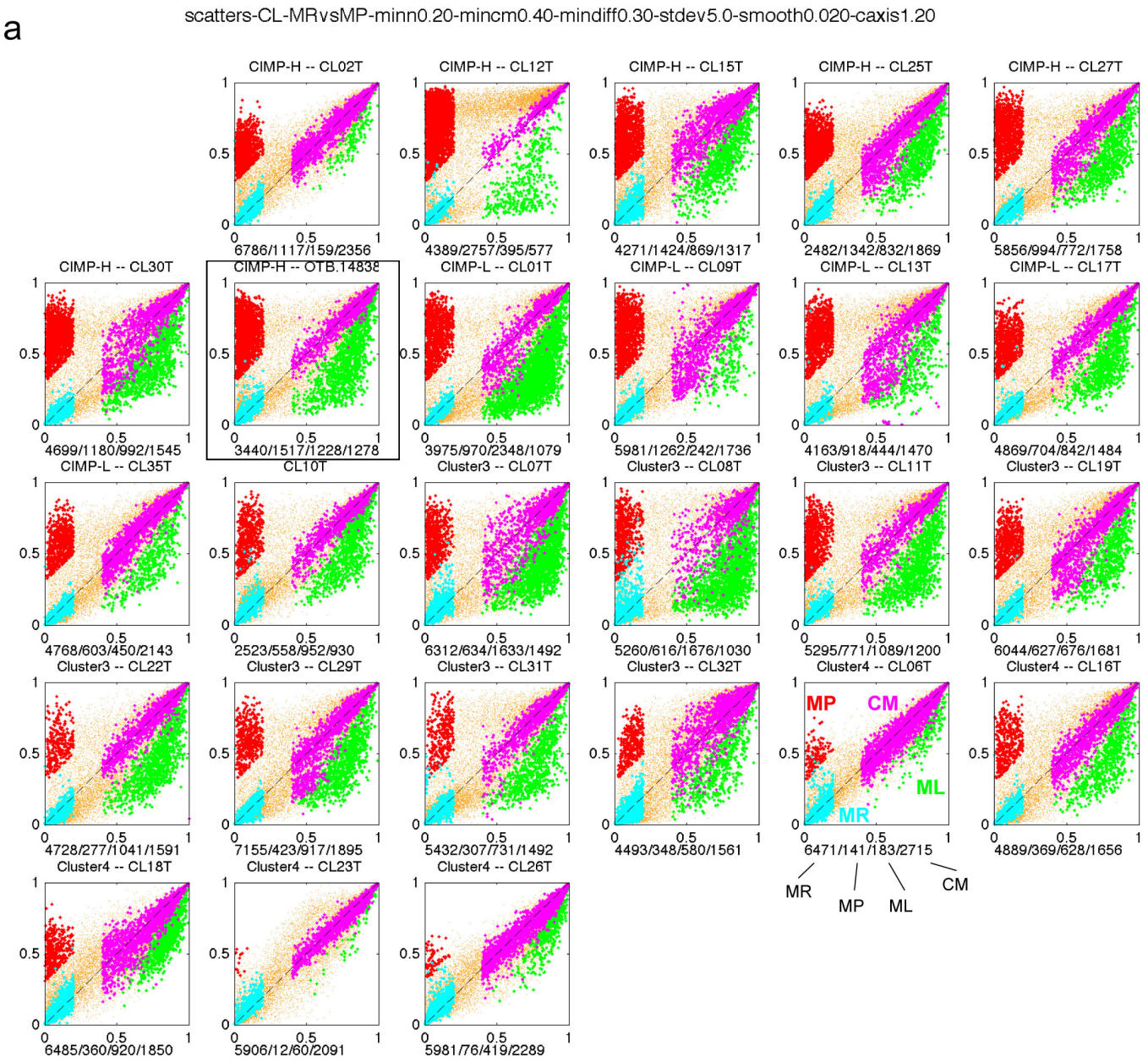
Supplementary Figure 10: Comparison to domains of Long Range Epigenetic Silencing from prostate cancer12 shows similarities and differences. 48 genomic regions from a previous study12 were compared to 2,850 tumor PMDs identified in the current study. a Overlap at the base pair level is close to random, while genes within Prostate LRESs are significantly enriched for tumor PMD genes (p<0.0001). b The region shown in Figure 3 of ref12 (LRES-24), with methylation tracks from the current study and prostate expression levels from Figure 3 of ref12. The top shows that LRES-24 is only one of several colon tumors within the larger genomic region. The bottom shows LRES-24 in greater detail, showing that a large segment in the center of it (from the MFDIC gene to the BC0396651 transcript) is not hypomethylated in the colon tumor. Because MFDIC does not appear to be silenced in LNCaP prostate cancer cells, this region is most likely not part of a true LRES, and is only included in LRES-24 due to the low resolution of the expression array used. Plots for the 7 examples included in ref12 show this to be the case for a number of LRESs, and are available as a Supplemental Document.

Berman et al., 2011 Supplemental Fig 11



Supplementary Figure 11: Partially Methylated Domain boundaries suggest a close relationship with nuclear lamina association and Polycomb silencing programs. Average genomic density of a number of annotation features is plotted for 10 kb bins relative to colon tumor PMD boundaries. Plots are oriented with regions outside the PMD to the left of the midpoint and regions inside the PMD to the right of the midpoint (shaded gray), as shown in the diagrams below each plot. Genomic density was normalized by dividing the value within each bin by the average density within bins lying outside of PMDs. Densities in blue indicate features oriented on the forward strand relative to the PMD boundary (i.e. pointing the same direction as the PMD boundary), and red indicate those oriented on the reverse strand (i.e. pointing the opposite direction).

Berman et al., 2011 Supplemental Figure 12



Supplementary Figure 12: An independent validation set consisting of 25 colon and rectal tumors and matched normal mucosa shows that the extent of hypermethylation and hypomethylation are correlated within a tumor. Infinium HumanMethylation27k scatter plots are shown here for colon and rectal tumor/normal pairs used in Figure 6, along with the case used for bisulfite-seq (OTB.14838). They are ordered by methylation subtype from their original study13, “CIMP-H”, “CIMP-L”, “Cluster 3”, and “Cluster 4”. For each pair, four methylation categories (MR, MP, ML, CM) were identified as described in Methods, and array features within each category are shown in a particular color. The number of features in each category is shown below the scatter, as illustrated for “Cluster 4 – CL06T”. OTB.14838 is outlined

Supplementary Figure 13

HOMER Results:

Methylation-Prone CGI promoters (MP-CGIs)

vs.

Methylation-Resistant CGI promoters (MR-CGIs)

Nature Genetics: doi:10.1038/ng.969

Homer *de novo* Motif Results

Supplementary Figure 13: HOMER results for MP CGI promoters vs. MR CGI promoters

Known Motif Enrichment Results
Gene Ontology Enrichment Results
If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into STAMP
More information on motif finding results: HOMER | Description of Results | Tips
* - possible false positive

Rank	Motif	P-value	log P-pvalue	Best Match/Details	Motif File
1		1.267e-113	-2.600e+02	SeqBias: CA-repeat More Information | Similar Motifs Found	motif file (matrix)
2		7.799e-80	-1.822e+02	SeqBias: GA-repeat More Information | Similar Motifs Found	motif file (matrix)
3		6.427e-58	-1.317e+02	PB0164.1_Smad3_2 More Information | Similar Motifs Found	motif file (matrix)
4		9.425e-55	-1.244e+02	PB0052.1_Plagl1_1 More Information | Similar Motifs Found	motif file (matrix)
5		6.197e-49	-1.110e+02	PF0024.1_GGGAGGRR More Information | Similar Motifs Found	motif file (matrix)
6		6.154e-48	-1.087e+02	MA0079.2_SP1 More Information | Similar Motifs Found	motif file (matrix)
7		4.624e-47	-1.067e+02	PF0099.1_ATCMNTCCGY More Information | Similar Motifs Found	motif file (matrix)
8		9.644e-44	-9.905e+01	MA0399.1_SUT1 More Information | Similar Motifs Found	motif file (matrix)
9		4.409e-43	-9.753e+01	SeqBias: GCW-triplet More Information | Similar Motifs Found	motif file (matrix)
10		7.393e-42	-9.471e+01	YY1/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
11		2.503e-36	-8.198e+01	PB0206.1_Zic2_2 More Information | Similar	motif file

				Motifs Found	(matrix)
12		1.151e-33	-7.584e+01	Sp1/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
13		3.673e-30	-6.778e+01	CN0218.1_LM218 More Information | Similar Motifs Found	motif file (matrix)
14		3.422e-29	-6.554e+01	PB0009.1_E2F3_1 More Information | Similar Motifs Found	motif file (matrix)
15		3.243e-28	-6.330e+01	PPARE(DR1)/3T3L1-Pparg-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
16		6.186e-27	-6.035e+01	MA0320.1_IME1 More Information | Similar Motifs Found	motif file (matrix)
17		2.267e-20	-4.523e+01	PB0203.1_Zfp691_2 More Information | Similar Motifs Found	motif file (matrix)
18		8.133e-17	-3.705e+01	MA0205.1_Trl More Information | Similar Motifs Found	motif file (matrix)
19		1.338e-16	-3.655e+01	MA0073.1_RREB1 More Information | Similar Motifs Found	motif file (matrix)
20		4.627e-16	-3.531e+01	GATA:SCL/Ter119-SCL-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
21		5.860e-14	-3.047e+01	CN0146.1_LM146 More Information | Similar Motifs Found	motif file (matrix)
22 *		4.551e-12	-2.612e+01	CN0177.1_LM177 More Information | Similar Motifs Found	motif file (matrix)

Supplementary Figure 14

HOMER Results:

Methylation-Resistant CGI promoters (MR-CGIs)

vs.

Methylation-Prone CGI promoters (MP-CGIs)

Nature Genetics: doi:10.1038/ng.969

Homer *de novo* Motif Results

Supplementary Figure 14: HOMER results for MR CGI promoters vs. MP CGI promoters

Known Motif Enrichment Results
Gene Ontology Enrichment Results
If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into STAMP
More information on motif finding results: HOMER | Description of Results | Tips
* - possible false positive

Rank	Motif	P-value	log P-pvalue	Best Match/Details	Motif File
1		3.504e-285	-6.550e+02	Sp1/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
2		4.433e-197	-4.521e+02	NRF1/MCF7-NRF1-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
3		7.761e-193	-4.423e+02	PB0007.1_Bhlhb2_1 More Information | Similar Motifs Found	motif file (matrix)
4		9.236e-189	-4.330e+02	GABPA/Jurkat-GABPa-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
5		2.773e-175	-4.019e+02	PB0020.1_Gabpa_1 More Information | Similar Motifs Found	motif file (matrix)
6		1.472e-135	-3.105e+02	YY1/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
7		2.107e-122	-2.802e+02	MA0043.1_HLF More Information | Similar Motifs Found	motif file (matrix)
8		8.647e-117	-2.672e+02	PH0154.1_Prrx1 More Information | Similar Motifs Found	motif file (matrix)
9		1.690e-113	-2.597e+02	NFY/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
10		5.878e-105	-2.400e+02	PH0071.1_Hoxc6 More Information | Similar Motifs Found	motif file (matrix)
11		2.200e-99	-2.272e+02	GFY/Promoter/Homer More Information | Similar Motifs Found	motif file (matrix)
12		3.242e-91	-2.084e+02	PB0161.1_Rxra_2 More Information | Similar Motifs Found	motif file (matrix)

Nature Genetics: doi:10.1038/ng.969

13		3.380e-85	-1.945e+02	PB0016.1_Foxj1_1 More Information | Similar Motifs Found	motif file (matrix)
14		1.810e-82	-1.882e+02	PH0174.1_Vax1 More Information | Similar Motifs Found	motif file (matrix)
15		9.761e-82	-1.865e+02	PH0117.1_Nkx3-1 More Information | Similar Motifs Found	motif file (matrix)
16		1.397e-81	-1.862e+02	PB0182.1_Srf_2 More Information | Similar Motifs Found	motif file (matrix)
17		7.416e-77	-1.753e+02	NFAT:AP1/Jurkat-NFATC1-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
18		8.895e-70	-1.590e+02	PB0096.1_Zfp187_1 More Information | Similar Motifs Found	motif file (matrix)
19		7.653e-69	-1.568e+02	PB0109.1_Bbx_2 More Information | Similar Motifs Found	motif file (matrix)
20		6.485e-65	-1.478e+02	PB0109.1_Bbx_2 More Information | Similar Motifs Found	motif file (matrix)
21		7.747e-64	-1.453e+02	PH0137.1_Pitx1 More Information | Similar Motifs Found	motif file (matrix)
22		1.127e-52	-1.196e+02	CN0230.1_LM230 More Information | Similar Motifs Found	motif file (matrix)
23		4.178e-49	-1.114e+02	PB0091.1_Zbtb3_1 More Information | Similar Motifs Found	motif file (matrix)
24		3.107e-45	-1.025e+02	PF0151.1_RYAAAKNNNNNNTTGW More Information | Similar Motifs Found	motif file (matrix)

Supplementary Figure 15

HOMER Results:

Methylation Loss elements (ML)

vs.

Methylation-Resistant elements (MR)

Nature Genetics: doi:10.1038/ng.969

Homer *de novo* Motif Results

Supplementary Figure 15: HOMER results for ML CGI promoters vs. MR CGI promoters

Known Motif Enrichment Results
Gene Ontology Enrichment Results
If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into STAMP
More information on motif finding results: HOMER | Description of Results | Tips
* - possible false positive

Rank	Motif	P-value	log P-pvalue	Best Match/Details	Motif File
1		2.025e-86	-1.973e+02	NFAT:AP1/Jurkat-NFATC1-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
2		5.799e-86	-1.963e+02	MA0303.1_GCN4 More Information | Similar Motifs Found	motif file (matrix)
3		1.575e-47	-1.078e+02	MA0345.1_NHP6A More Information | Similar Motifs Found	motif file (matrix)
4		2.177e-41	-9.363e+01	PB0116.1_Elf3_2 More Information | Similar Motifs Found	motif file (matrix)
5		6.933e-38	-8.556e+01	MA0386.1_TBP More Information | Similar Motifs Found	motif file (matrix)
6		5.176e-20	-4.441e+01	PB0120.1_Foxj1_2 More Information | Similar Motifs Found	motif file (matrix)
7		3.888e-19	-4.239e+01	MA0346.1_NHP6B More Information | Similar Motifs Found	motif file (matrix)
8		5.040e-19	-4.213e+01	PH0148.1_Pou3f3 More Information | Similar Motifs Found	motif file (matrix)
9		5.822e-17	-3.738e+01	PH0139.1_Pitx3 More Information | Similar Motifs Found	motif file (matrix)
10		9.258e-	-	PB0095.1_Zfp161_1 More Information |	motif file

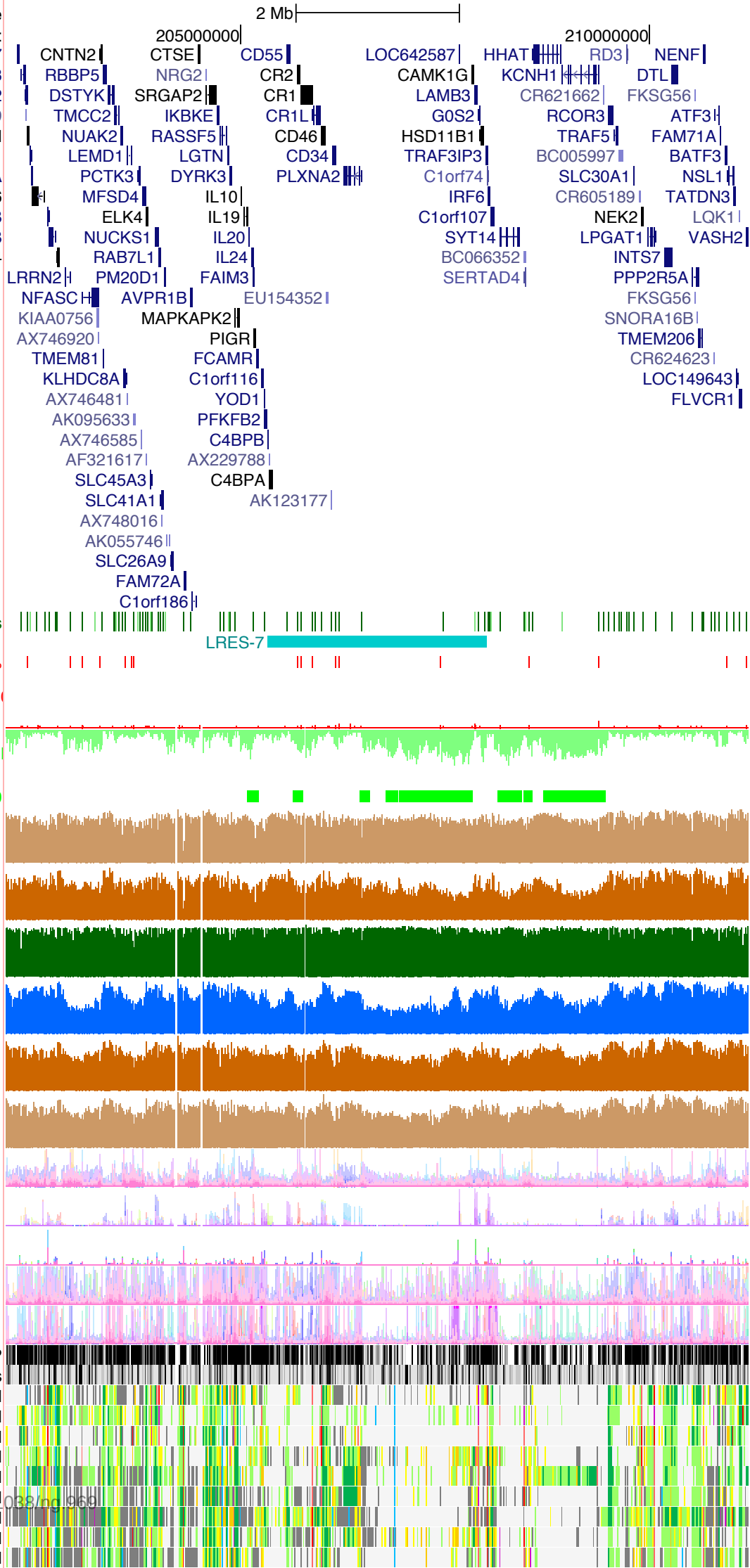
		16	3.462e+01	Similar Motifs Found	(matrix)
11		1.702e-15	3.401e+01	CN0106.1_LM106 More Information | Similar Motifs Found	motif file (matrix)
12		1.098e-14	3.214e+01	PB0091.1_Zbtb3_1 More Information | Similar Motifs Found	motif file (matrix)
13		1.343e-14	3.194e+01	MA0088.1_znf143 More Information | Similar Motifs Found	motif file (matrix)
14		1.655e-14	3.173e+01	PB0028.1_Hbp1_1 More Information | Similar Motifs Found	motif file (matrix)
15		1.064e-13	2.987e+01	FOXA1:AR/LNCAP-AR-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)
16 *		4.097e-10	2.162e+01	PH0148.1_Pou3f3 More Information | Similar Motifs Found	motif file (matrix)
17 *		4.762e-10	2.147e+01	RUNX(AML)/CD4+-PolII-ChIP-Seq/Homer More Information | Similar Motifs Found	motif file (matrix)

Supplementary Figure 16

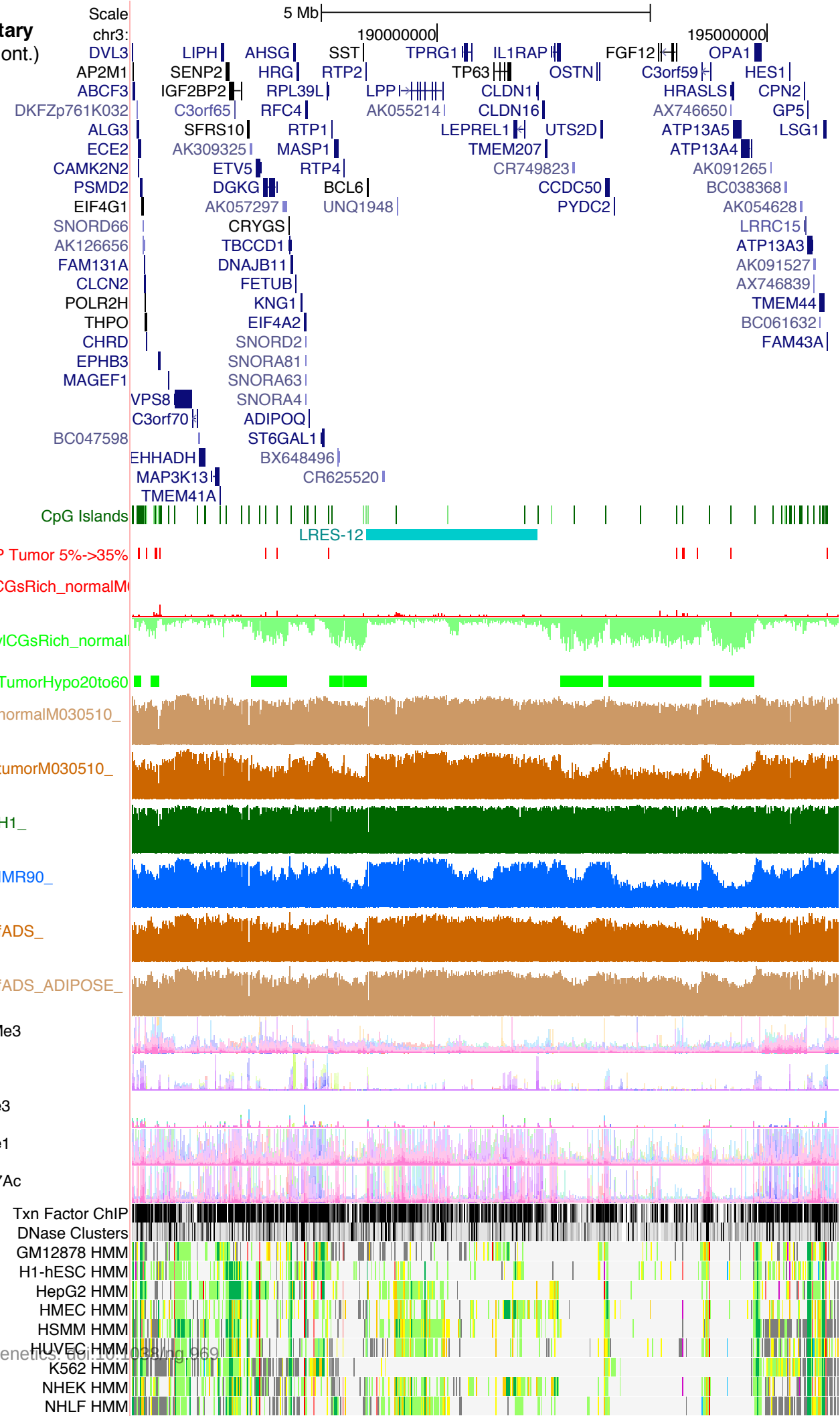
Methylation patterns around prostate cancer Long-Range Epigenetic Silencing (LRES) domains from Coolen, et. al. 2010[12]

Nature Genetics: doi:10.1038/ng.969

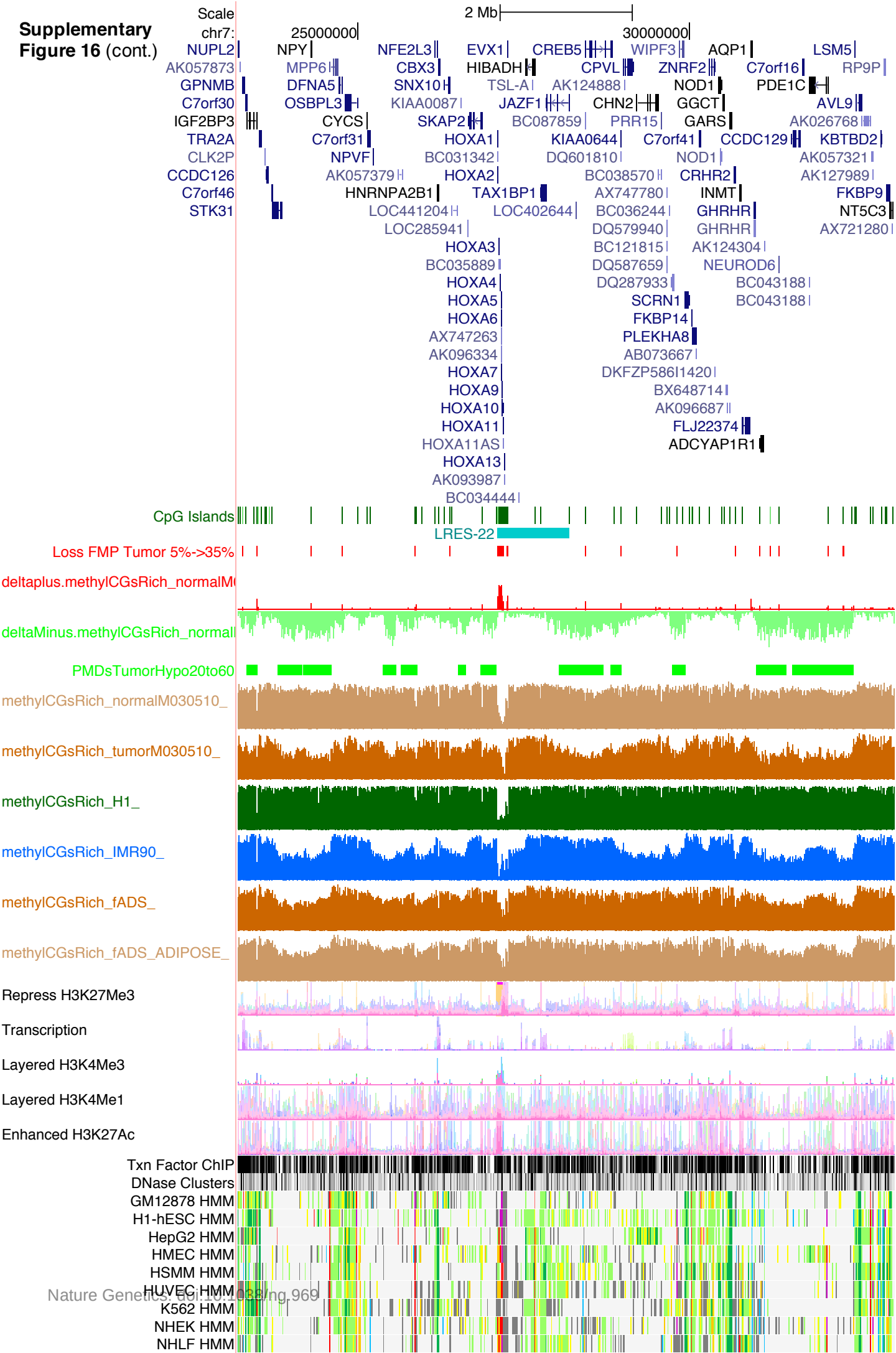
Supplementary Figure 16 (cont.)



Supplementary Figure 16 (cont.)

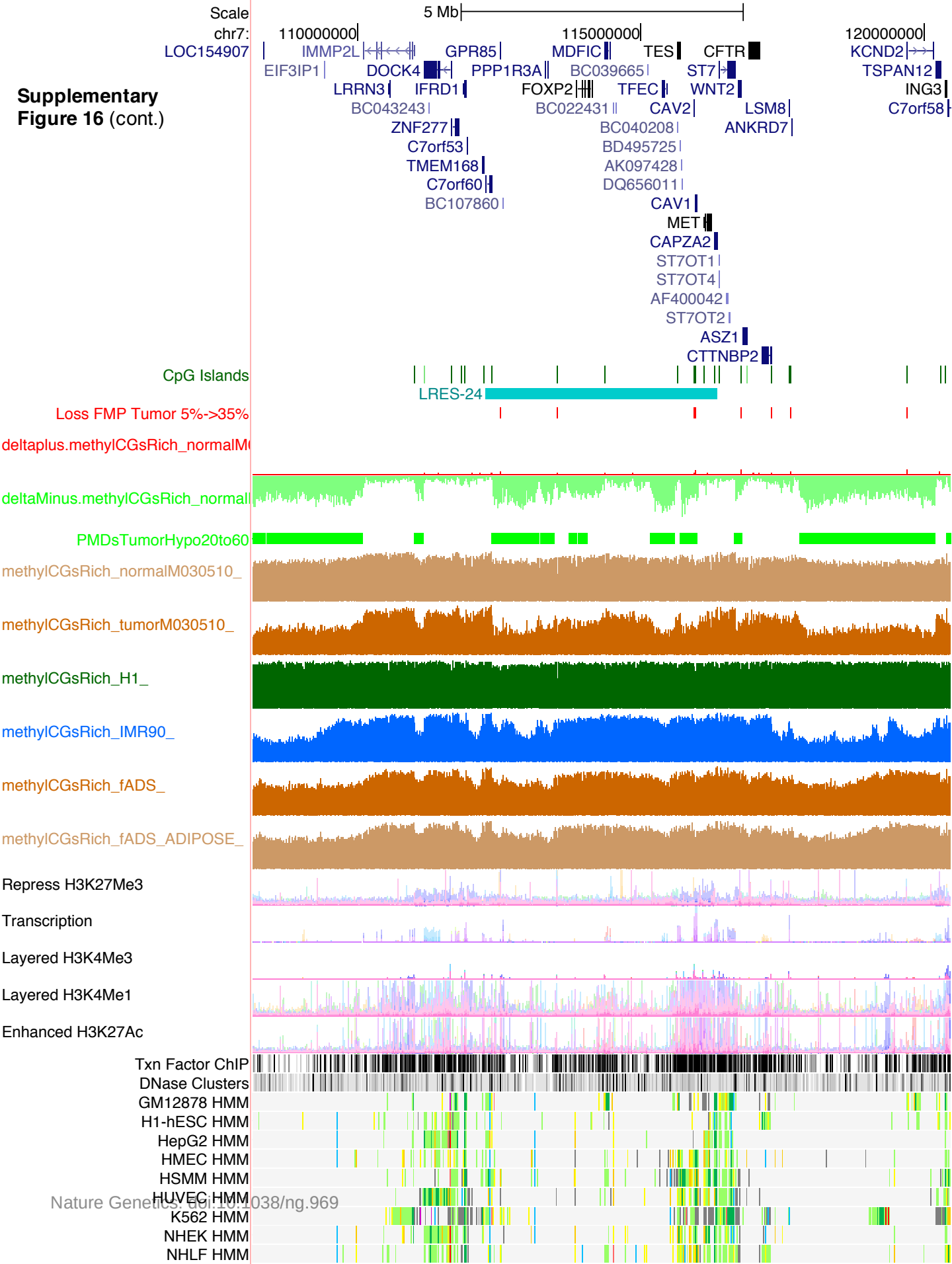


Supplementary Figure 16 (cont.)



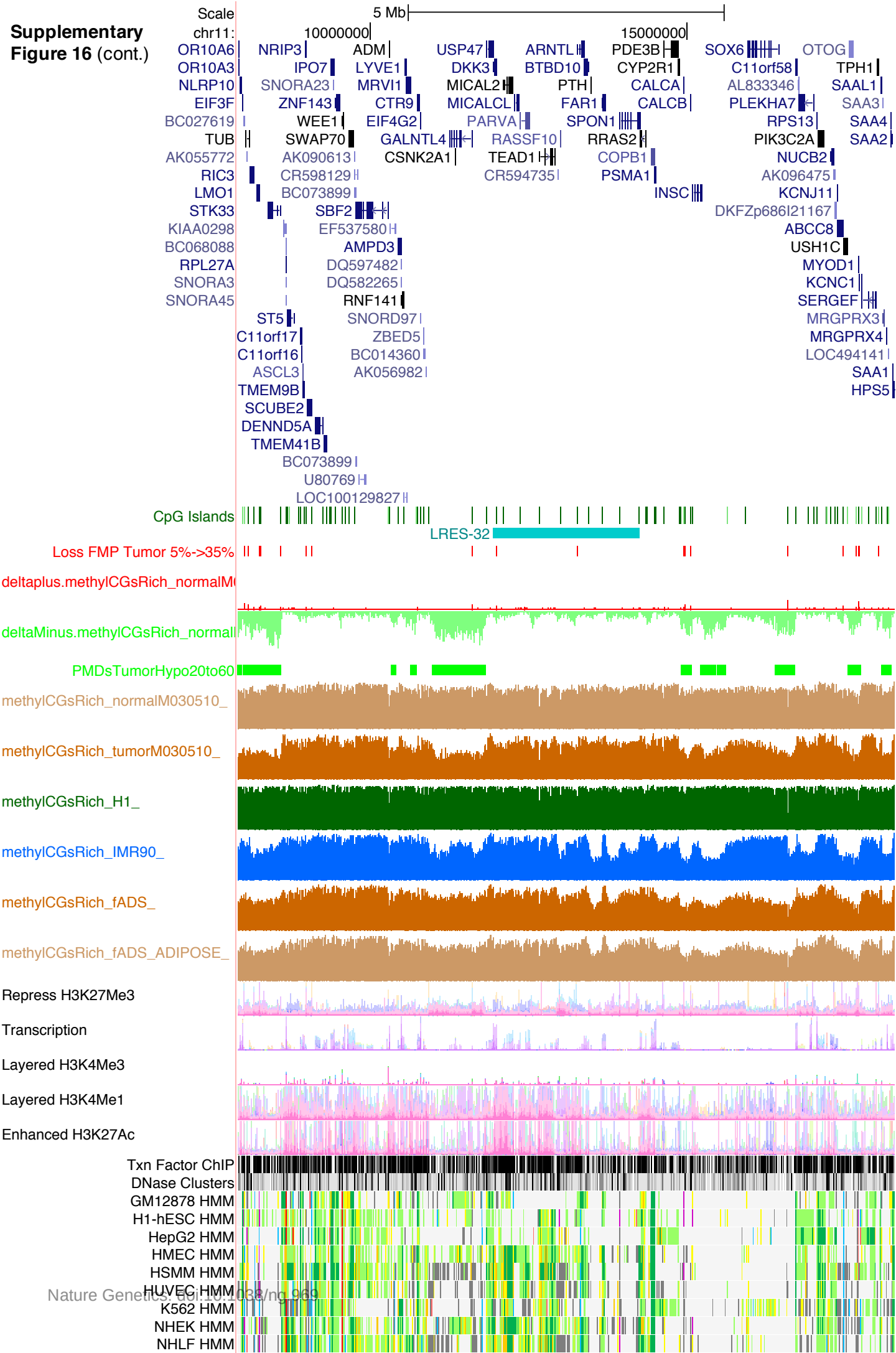
Nature Genetics: doi:10.1038/ng.969

Supplementary Figure 16 (cont.)

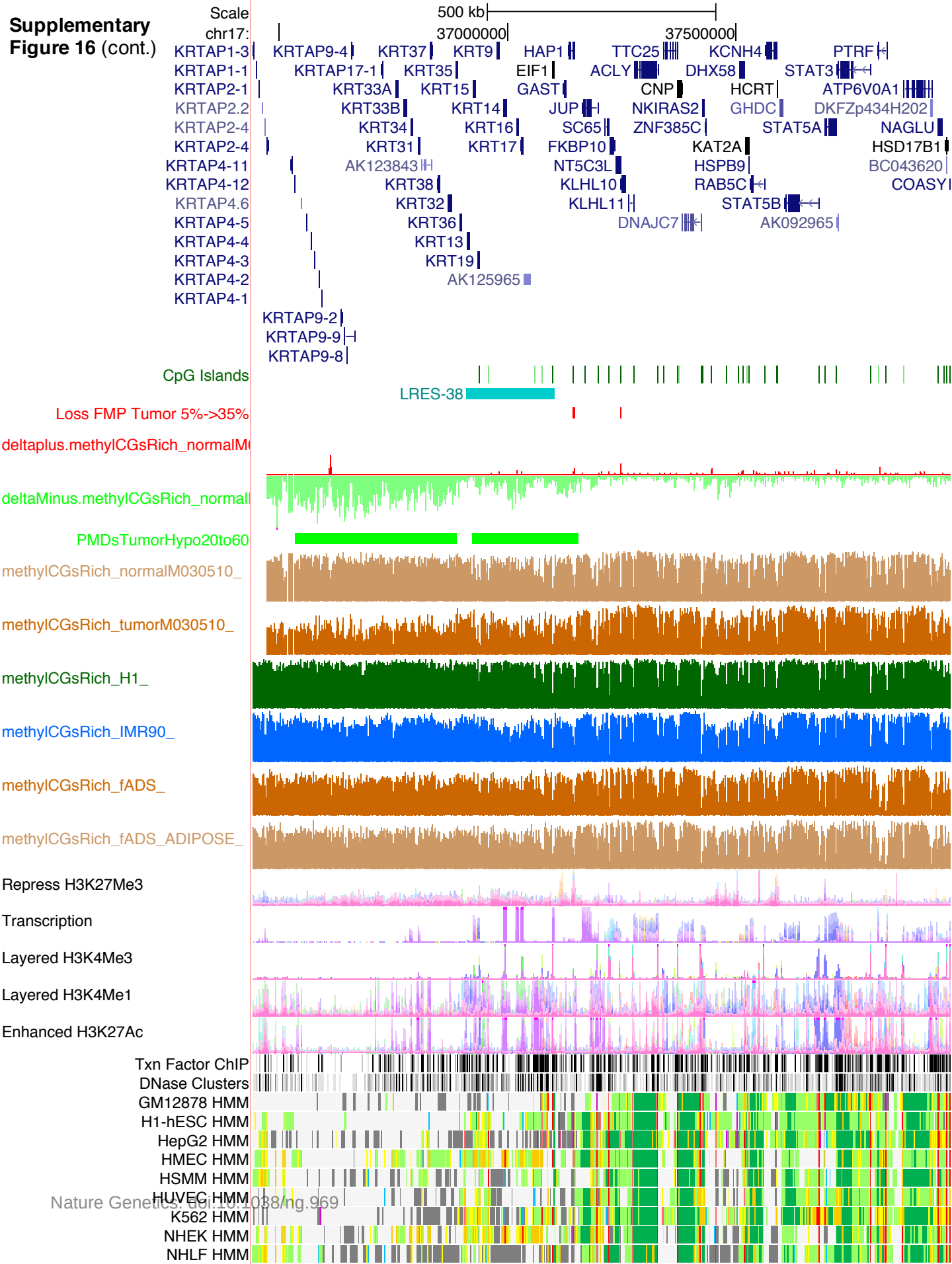


Nature Genetics. doi:10.1038/ng.969

Supplementary Figure 16 (cont.)

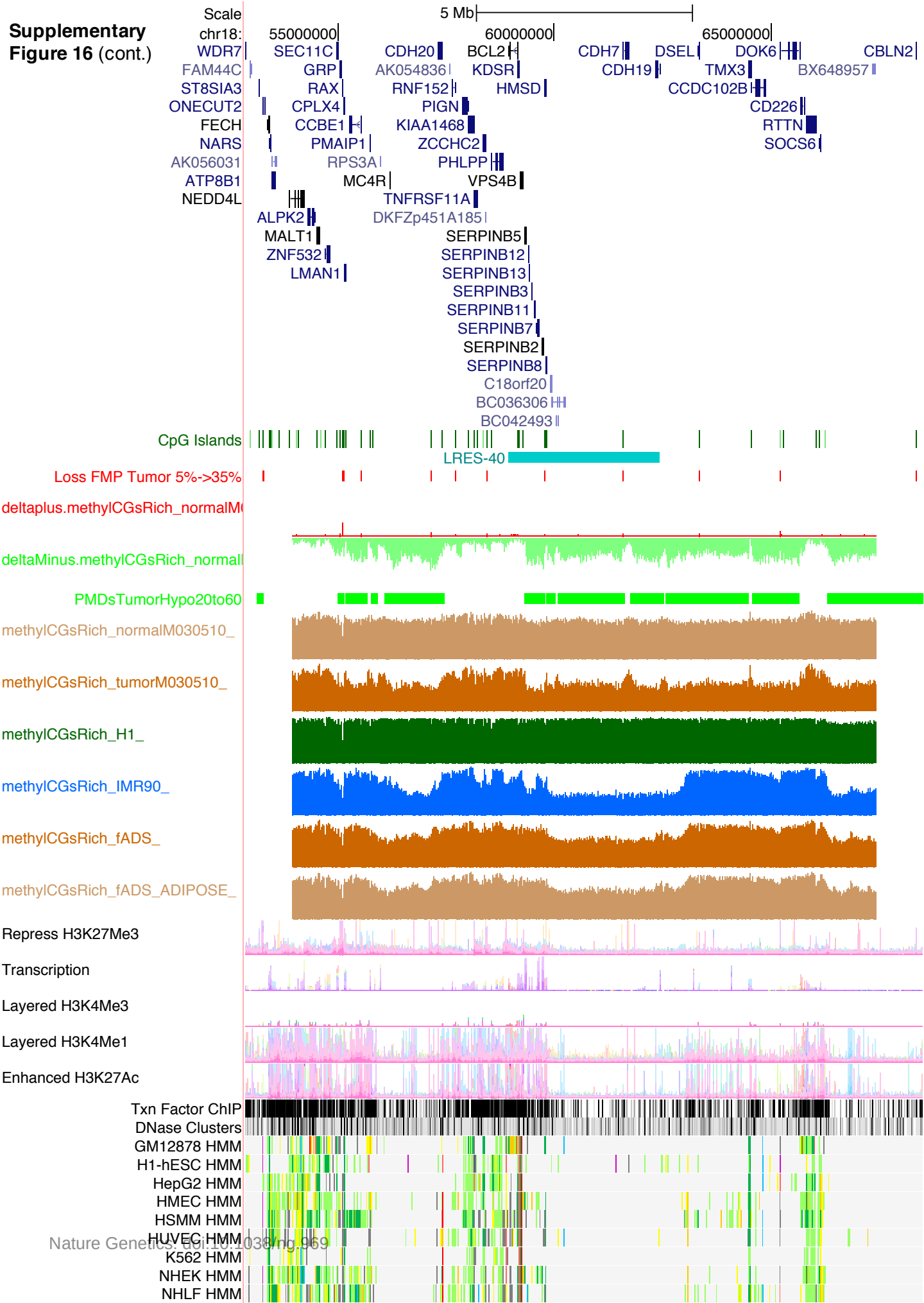


Supplementary Figure 16 (cont.)



Nature Genetics. doi:10.1038/ng.969

Supplementary Figure 16 (cont.)



References:

1. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
2. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, 1-8 (2011).
3. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res* **20**, 320-31 (2010).
4. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**, 576-89 (2010).
5. Goecks, J., Nekrutenko, A., Taylor, J. & Team, G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**, - (2010).
6. H., W. *ggplot2: elegant graphics for data analysis*, (Springer, New York, NY, 2009).
7. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
8. Hansen, R.S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences* **107**, 139-144 (2010).
9. Sun, W. *et al.* Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic acids research* **37**, 5365-77 (2009).

10. Consortium, T.E.P. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol* **9**, e1001046 (2011).
11. Balwierz, P.J. *et al.* Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data. *Genome Biol* **10**, R79 (2009).
12. Coolen, M.W. *et al.* Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nat Cell Biol* **12**, 235-46 (2010).
13. Hinoue, T. *et al.* Genome-Scale Analysis of Aberrant DNA Methylation in Colorectal Cancer. *Genome research* (2011).