Supplemental Material

Table of Contents

Table S1. Characteristics of primary GBMs used in this study.

A. GBMs profiled by MeDIP-seq and MRE-seq

	sex	age	% tumor cells	TCGA#	Verhaak expression subtype	copy number amplif.	copy number homoz. del.
GBM1	M	76	N.D.	-	N.D.	CDK4, MDM2	PTEN
GBM2	F	55	80	08-0354	classical	EGFR	CDKN2A/B
GBM3	F	59	90	08-0359	proneural	none	CDKN2A/B
GBM4	F	46	96	-	mesenchymal	none	CDKN2A/B
GBM5	M	58	95	-	proneural	PDGFRA	CDKN2A/B

N.D. -- not determined
GBMs 2 and 3 were profiled in The Cancer Genome Atlas (TCGA) project; sample ID numbers are given
copy number amplif. – high level copy number amplification detected by Agilent 244K arrays
copy number homoz. del. – homozygous copy number deletion detected by Agilent 244K arrays

B. Other GBMs in this study

	sex	age
GBM6	M	54
GBM7	F	45
GBM8	M	47
GBM9	M	61
GBM10	M	51

Table S2. Complete GREAT enrichments for recurring hyper- and hypomethylated DMRs (Q<1e-13)
-- see accompanying Excel file Table_S2.xlsx

Table S3. Recurring and individual hypomethylated DMRs consistently enrich for a 5p15 amplified region in breast cancer.

Ontology: MSigDB Perturbation.
GSEA ID:NIKOLSKY_BREAST_CANCER_5P15_AMPLICON

		GREAT RESULTS			
	DMR significance threshold	Binom Rank	Binom FDR Q-Val	Hyper Rank	Hyper FDR Q-Val
COMMON HYPO	Q<1E-13	1	2.03E-33	1	1.42E-10
COMMON HYPO	Q<1E-5	45	1.80E-21	64	0.0054369
GBM1 HYPO	Q<1E-13	1	1.07E-59	1	1.29E-11
GBM2 HYPO	Q<1E-13	1	2.05E-83	1	1.79E-12
GBM3 HYPO	Q<1E-13	1	3.64E-33	5	1.7521E-10
GBM4 HYPO	Q<1E-13	1	7.77E-31	4	0.00169554

Table S4. Disease Ontology functional enrichments of DNA hypo/K4me3 loci by GREAT. The 20 top enrichments significant by both binomial and hypergeometric tests are shown.

GBM1

Term Name	Binom Rank	Binom FDR Q-Val	Hyper Rank	Hyper FDR Q-Val
myeloma	3	5.54E-04	2	2.51E-02
hemorrhagic disease	4	6.12E-04	9	4.35E-02
bone marrow cancer	5	5.12E-04	4	2.10E-02
hematologic cancer	8	6.61E-04	3	1.98E-02
bone marrow disease	10	1.59E-03	7	5.27E-02
lymphoma	12	2.61E-03	8	4.65E-02
anthracosis	14	4.63E-03	6	4.75E-02

GBM2

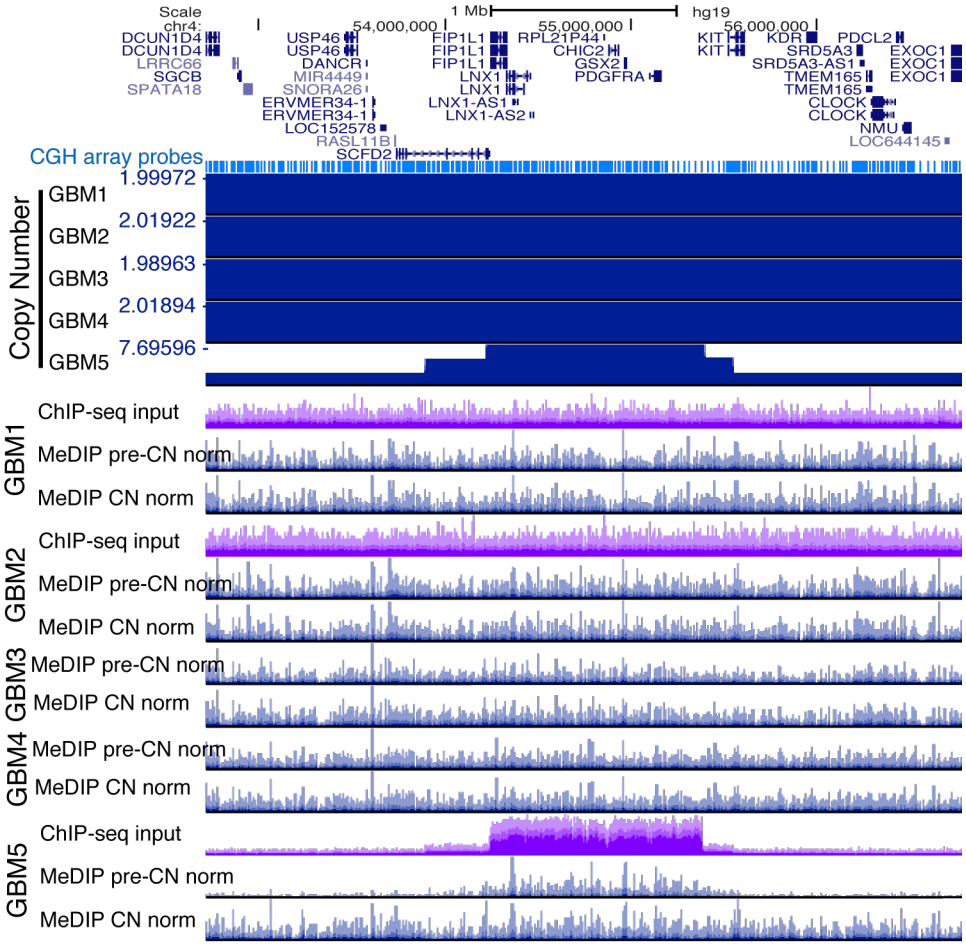
Term Name	Binom Rank	Binom FDR Q-Val	Hyper Rank	Hyper FDR Q-Val
breast carcinoma	5	2.46E-04	9	4.31E-04
glandular and epithelial neoplasm	6	3.65E-04	1	2.71E-04
hemorrhagic disease	7	5.86E-04	28	7.48E-04
ductal, lobular, and medullary neoplasm	8	6.38E-04	8	3.50E-04
blood coagulation disease	9	6.07E-04	31	8.94E-04
lymphoma	10	6.87E-04	2	2.37E-04
multiple myeloma	11	9.40E-04	26	7.23E-04
hematologic cancer	12	8.85E-04	5	2.25E-04
pancreas adenocarcinoma	13	1.06E-03	24	7.46E-04
plasmacytoma	15	1.05E-03	29	7.29E-04
vascular hemostatic disease	16	1.13E-03	40	1.22E-03
mature B-cell lymphocytic neoplasm	17	1.14E-03	32	8.94E-04
fibrous tissue neoplasm	18	1.22E-03	48	2.21E-03
malignant neoplasm of lymphatic and hemopoietic tissue	20	1.13E-03	7	3.13E-04

Table S5. PCR primers used in this study.

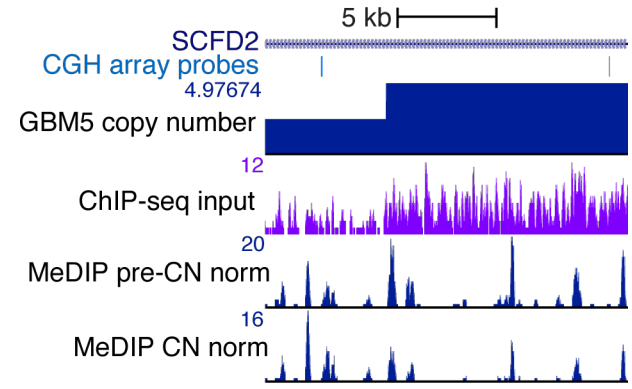
Primer 5'-3' sequence	experiment	amplicon (bp)	reference
GGGTTTGTTGGGTGTTTAGG AATAAATCCCCATACTCCCTACTCT	bisulfite sequencing	329	
GGTTGTAAGGTTGTATAGGAATAGA AAATAAAAAAAACCTTAAATTCCC	bisulfite sequencing	367	
TAGGGTAGGAAGTTTTAGGTATTGA TTTTCCTAACTACCTCACCCAACTA	bisulfite sequencing	540	
GGTTTTTTATTTTTAAGTATGAATATT ATTACAAAACTCTTTCAACTTTACC	bisulfite sequencing	527	
GATACGCGTCACGCACTGCCATCTATCTT AGTAAGCTTCAAGCCCTGTTTAGCAGCTT	cloning for prom assay	1388	
CTAGCTAGCGCTGGCAAAAACCACGTC AGTAAGCTTACTGGGGAAGCCCAGAGTAT	cloning for prom assay	586	
GATACGCGTCCAGGTGGTACCTGTGGAAG TGCTAGATCTTGACAGAAGCCCAGAAGTGA	cloning for prom assay	1199	
GATACGCGTTGTCAGGCTTCCCAGAAGTT TGCTAGATCTATGGGGGAAAATAGGACCTG	cloning for prom assay	695	
CAAACGGCCCGCATGTTCCC TTGAACTGGGCCGTGGCGAG	qRT-PCR	231	Concin et al. 2004 Cancer Research
ACACTCATCAGCCAGTGCAG GTGGCATGTCCTTCTCGTTT	qRT-PCR	102	
ATGGGGAAGGTGAAGGTCG GGGGTCATTGATGGCAACAAT	qRT-PCR	108	
GTAGGACGTGGCTCTTGA	5' RACE 1st strand synthesis	N.A.	
TCCTCACGCAGACGGTGCTC (use with reverse primer from Invitrogen GeneRacer Kit)	5' RACE 1st PCR primer	N.A.	
GACACGGGGACACCGCATA (use with reverse primer from Invitrogen GeneRacer Kit)	5' RACE 2nd PCR primer (GSP3)	N.A.	
GGCGCAAAGTAGCTAGCCAGA TTGCTGAGCCCCTGGACATTC	qRT-PCR (GLI3)	122	
CGGCCTTACACCCTATTTGA TTTCCCGTTTTGTTCTCTGG	qRT-PCR (SIGLEC11)	177	
GTGGCATGTCCTTCTCGTTT GCTGCTGGTGTCTGCTCTC	RT-PCR (TERT)	1884	

S1A,B,C. PDGFRA amplification in GBM5

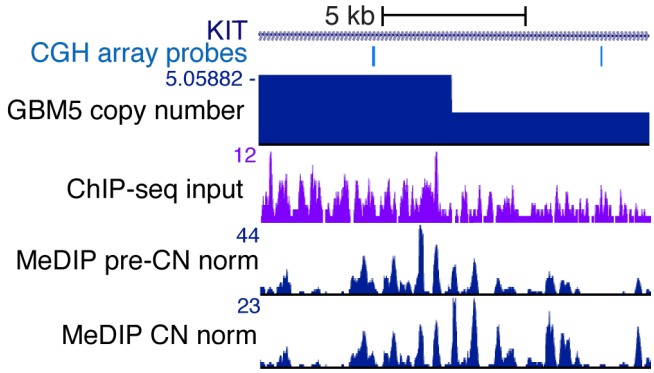
A



B

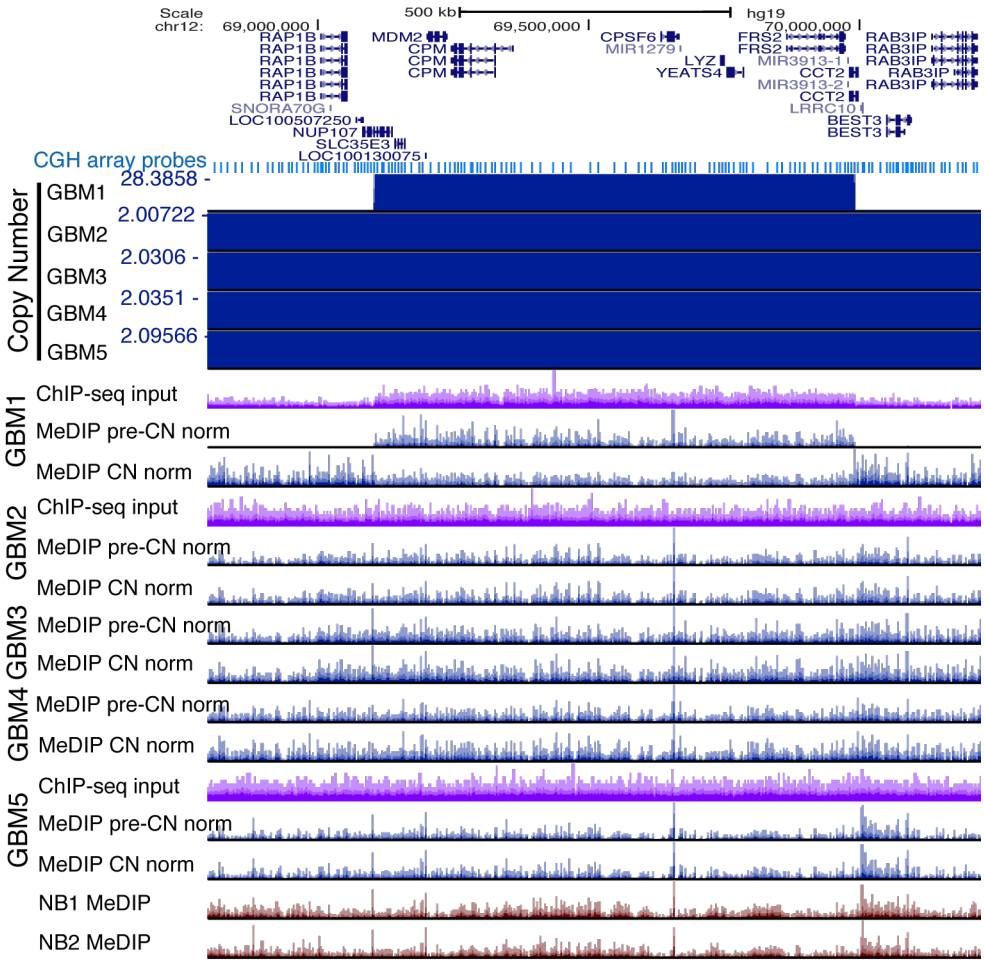


C

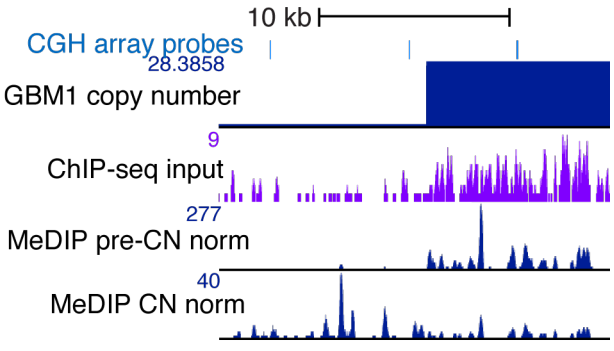


S1D,E,F. MDM2 amplification in GBM1

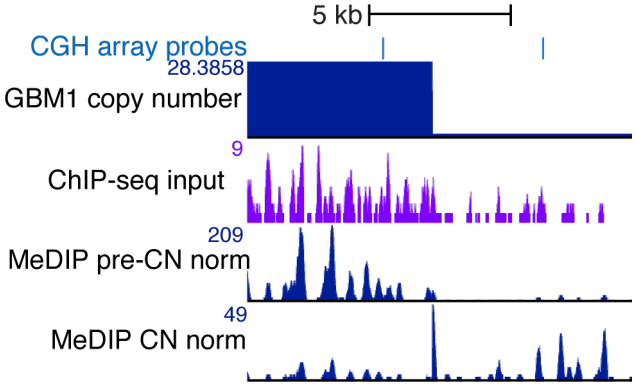
D



E

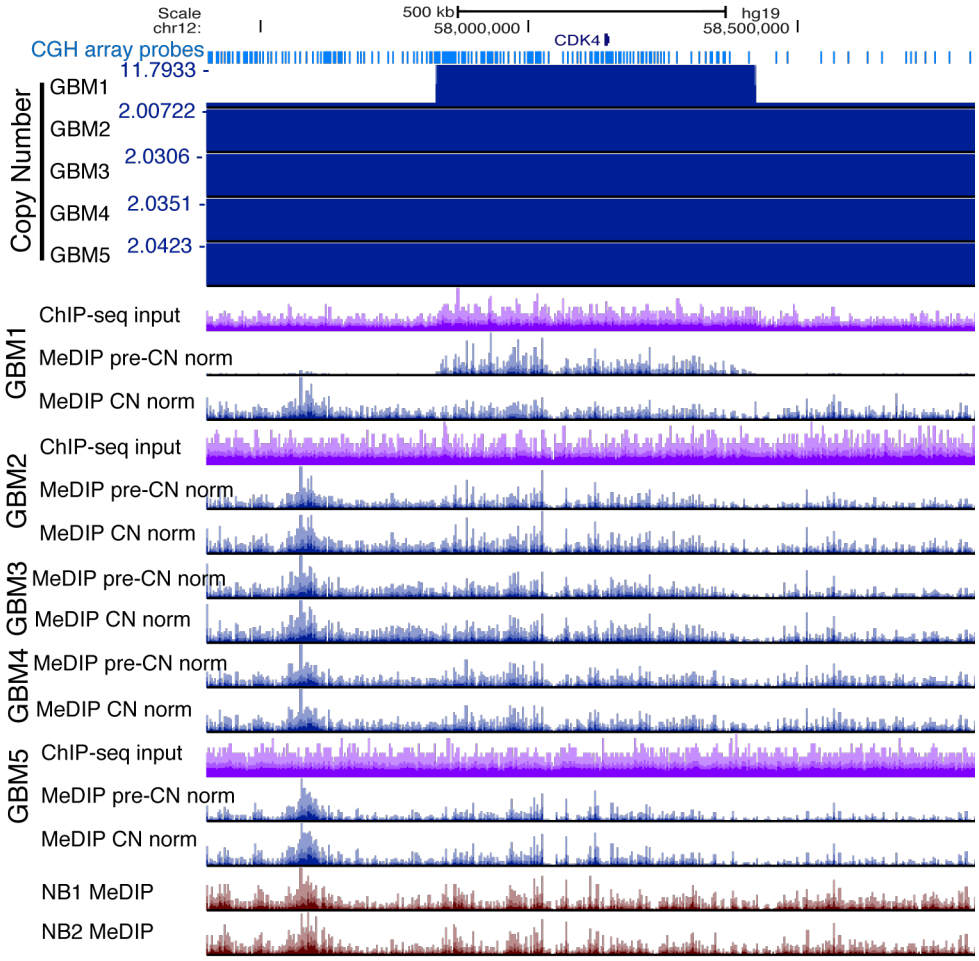


F

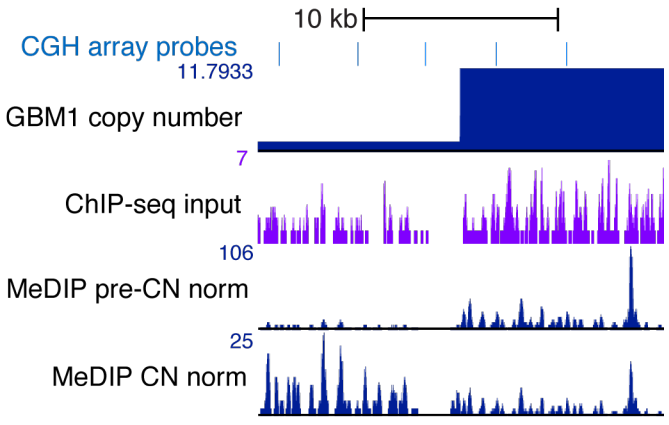


S1G,H,I. CDK4 amplification in GBM1

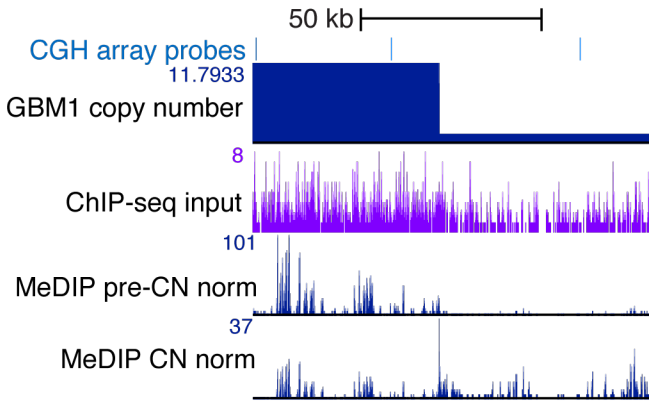
G



H

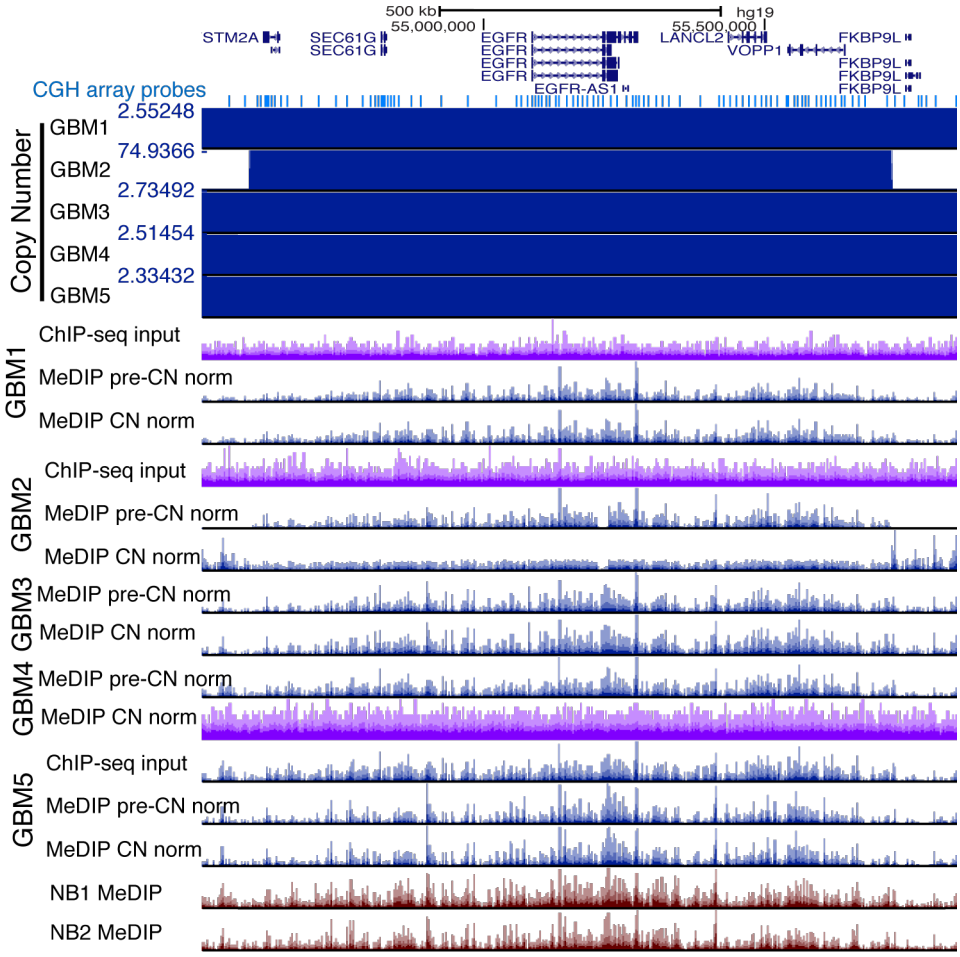


I

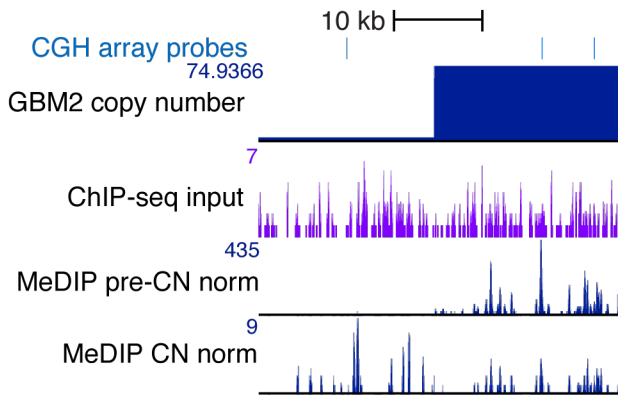


S1J,K,L. EGFR amplification in GBM2

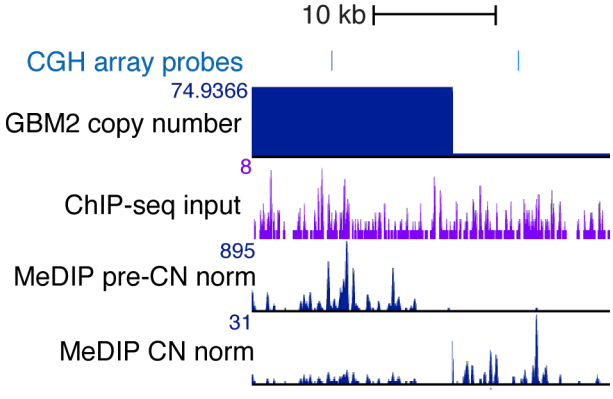
J



K

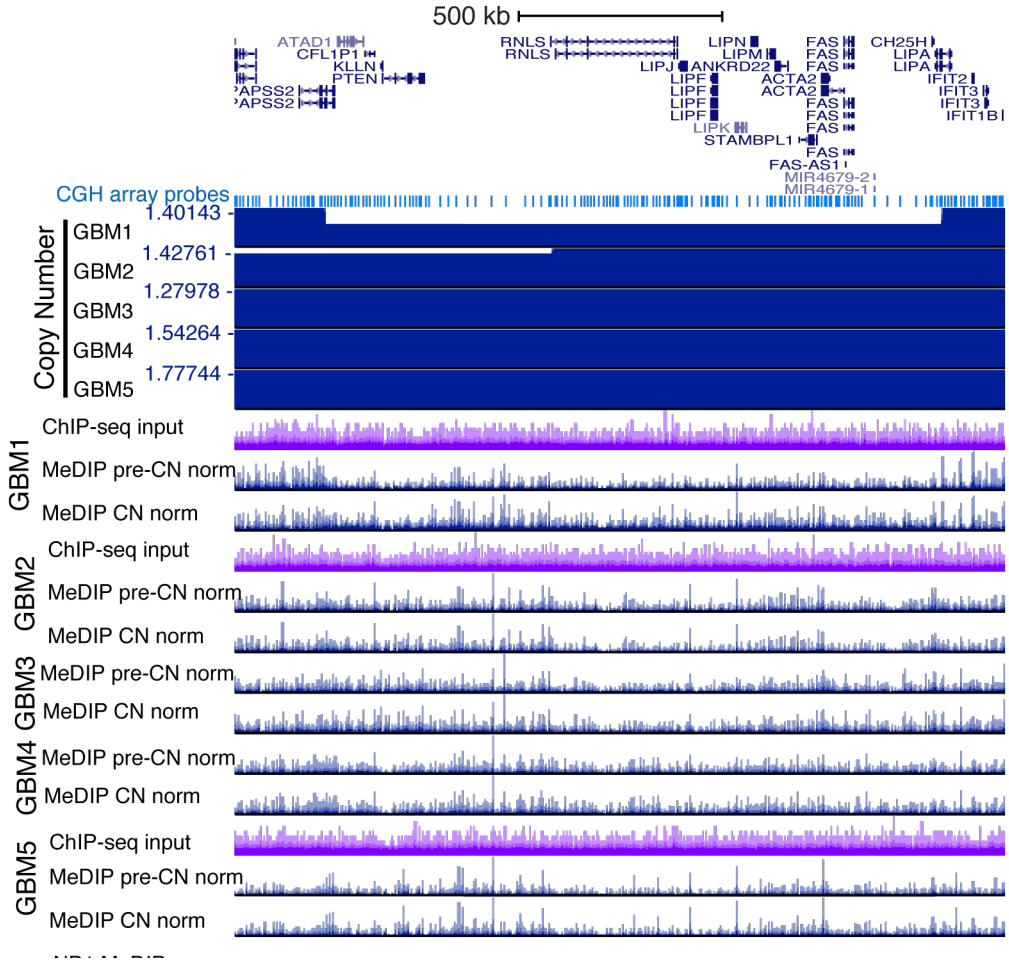


L

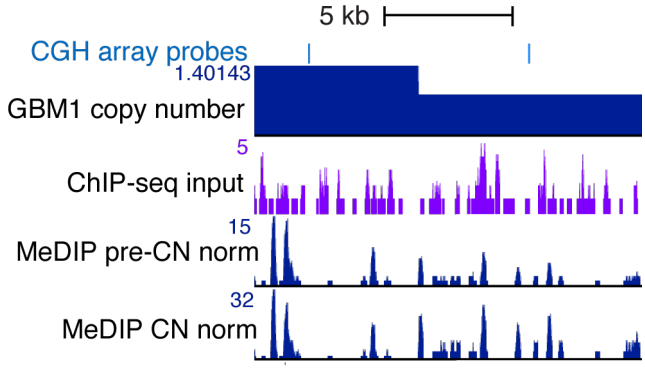


S1M,N,O. PTEN deletion

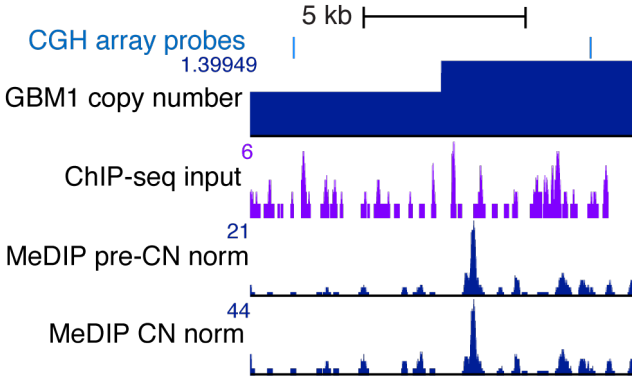
M



N

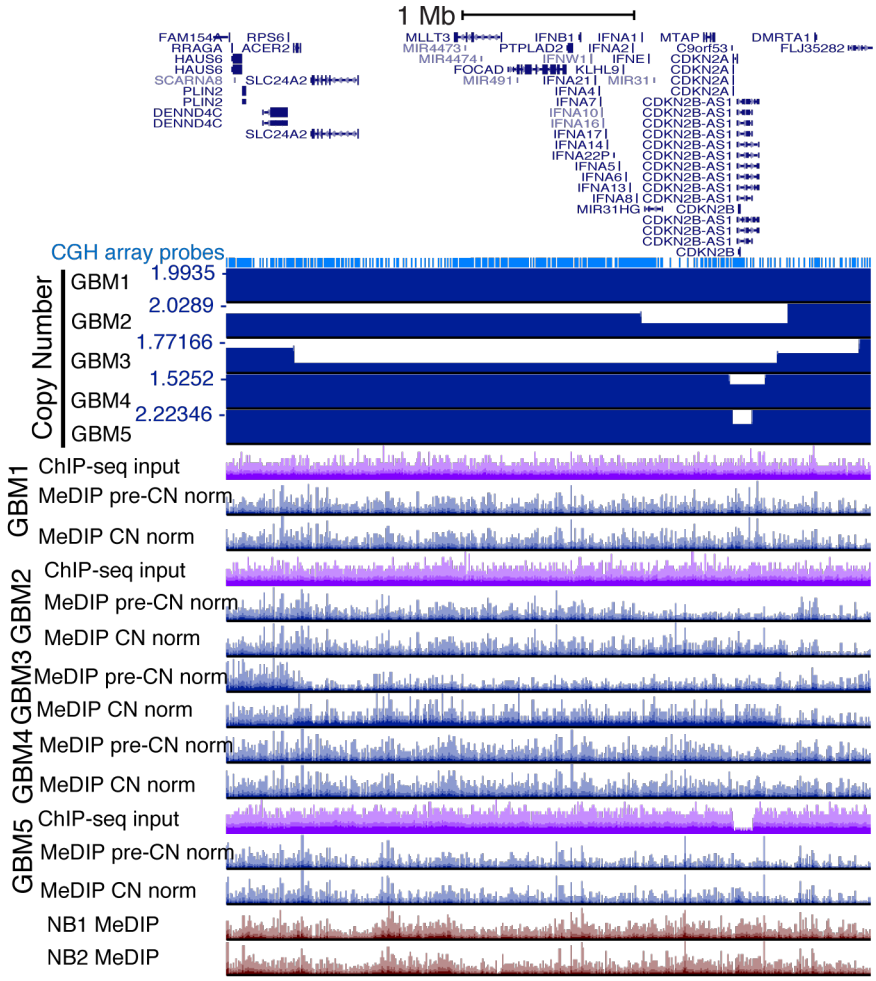


O

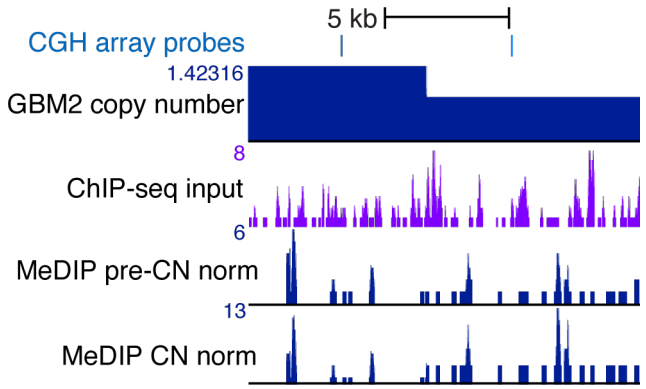


S1P,Q,R. CDKN2A/B deletion

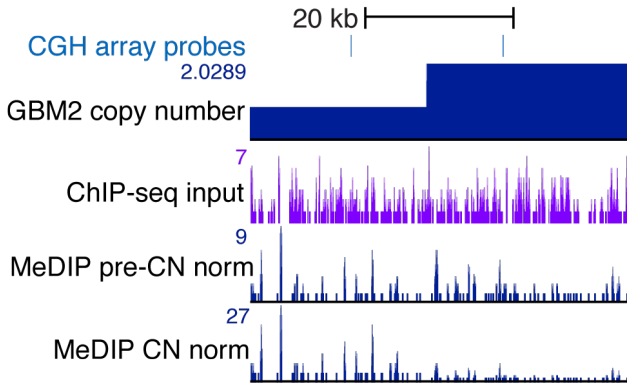
P



Q



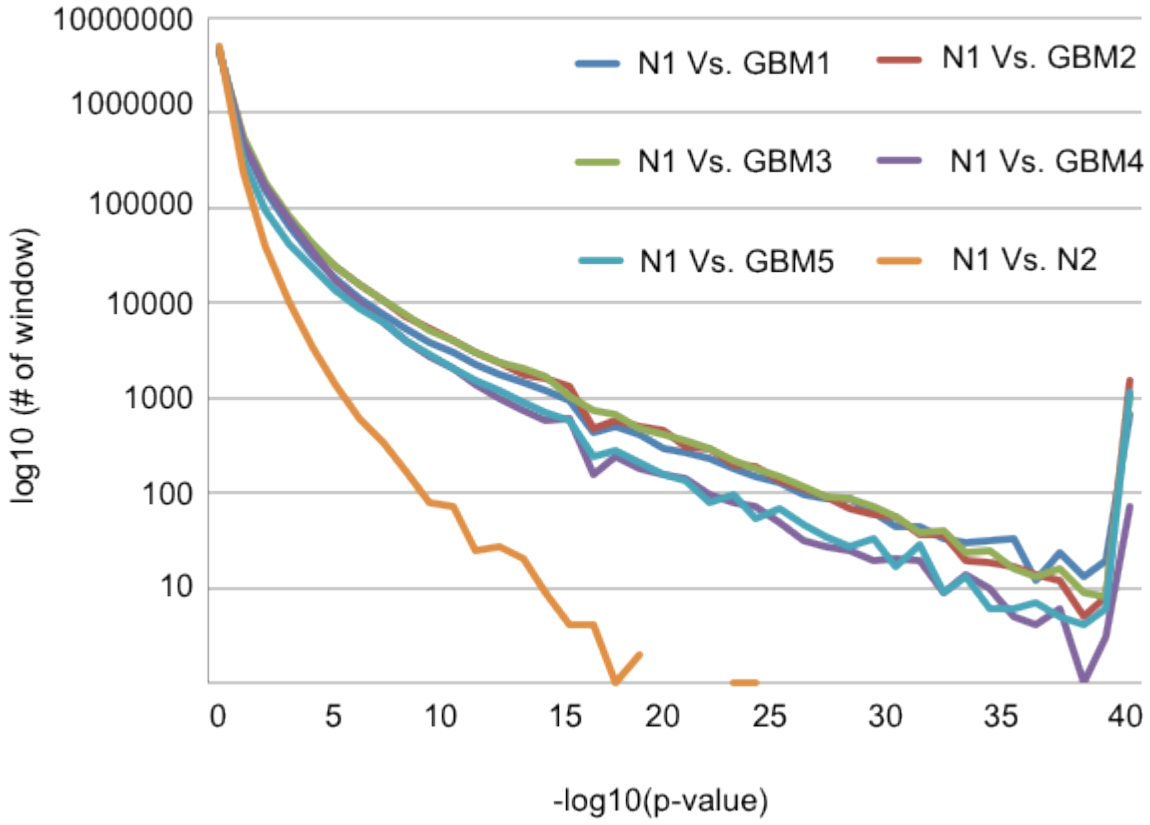
R



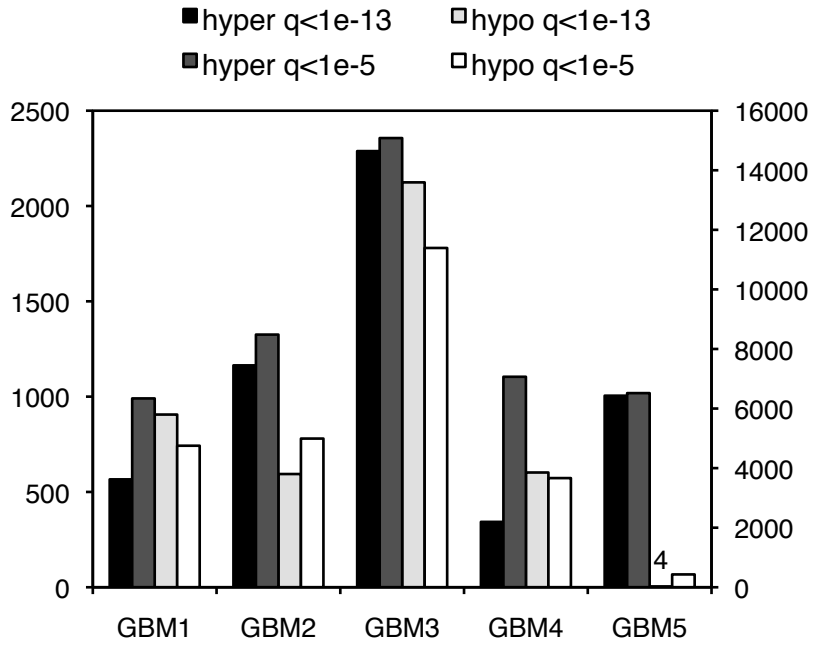
S1. Normalization of methylation sequencing data by array-CGH-derived copy number estimates. A, PDGFRA amplification in GBM5. Shown here are **1)** Agilent CGH array probes locations; **2)** estimated copy number by array-CGH, with the locations of changepoints between segments refined by seqCBS on MeDIP-seq (see methods); **3)** MeDIP-seq read density with no CNA normalization; **4)** MeDIP-seq read density with CNV normalization (CBS) and refinement of segment ends (seqCBS) and **5)** ChIP-seq input read density for GBMs

1,2 and 5. The amplified region is reflected by much higher read density in the non-CNA normalized MeDIP-seq (2nd track from the bottom). The amplification is also evident in the ChIP-seq input read density (3rd track from the bottom). At the left and right boundaries of this amplicon, seqCBS on MeDIP-seq identified change points at the exact location where ChIP-seq input read density changes abruptly (Fig. S1B,C). **D-R,** The same data are shown for CDK4, MDM2, EGFR, PTEN, and CDKN2A/B.

S2A

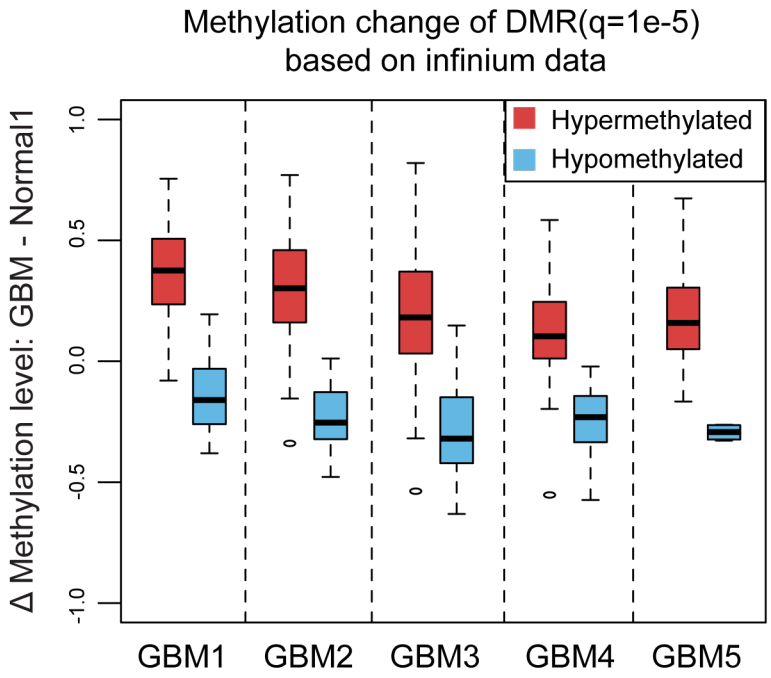


S2B



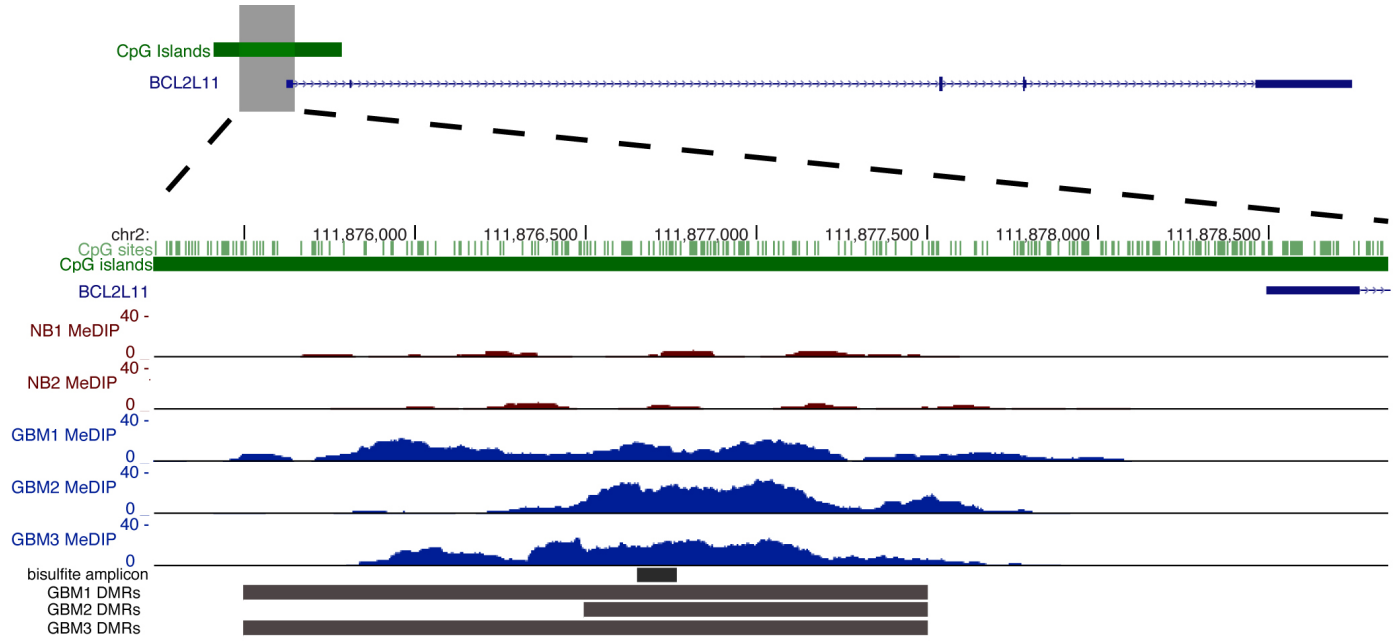
S2. A, DMRs identified by M&M. A, The number of DMRs found in normal brain1 (N1) vs. normal brain2 (N2) is lower at all significance thresholds compared to GBM-normal brain1 comparisons. **B,** Number of hyper- and hypomethylation differentially methylated regions (DMRs) at Q<1e-13 (left axis) and Q<1e-5 (right axis).

S3

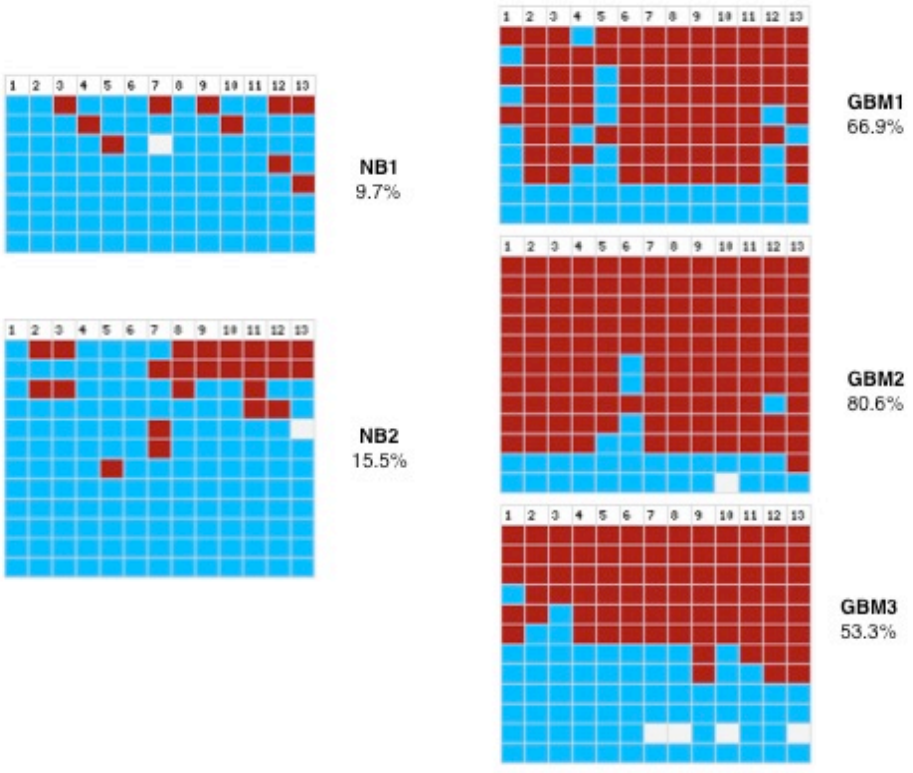


S3. Delta beta values (GBM-normal brain) from Infinium 27K arrays, plotted for CpG sites within individual hyper- and hypomethylated DMRs ($q<10^{-5}$) for each of five GBMs.

S4A

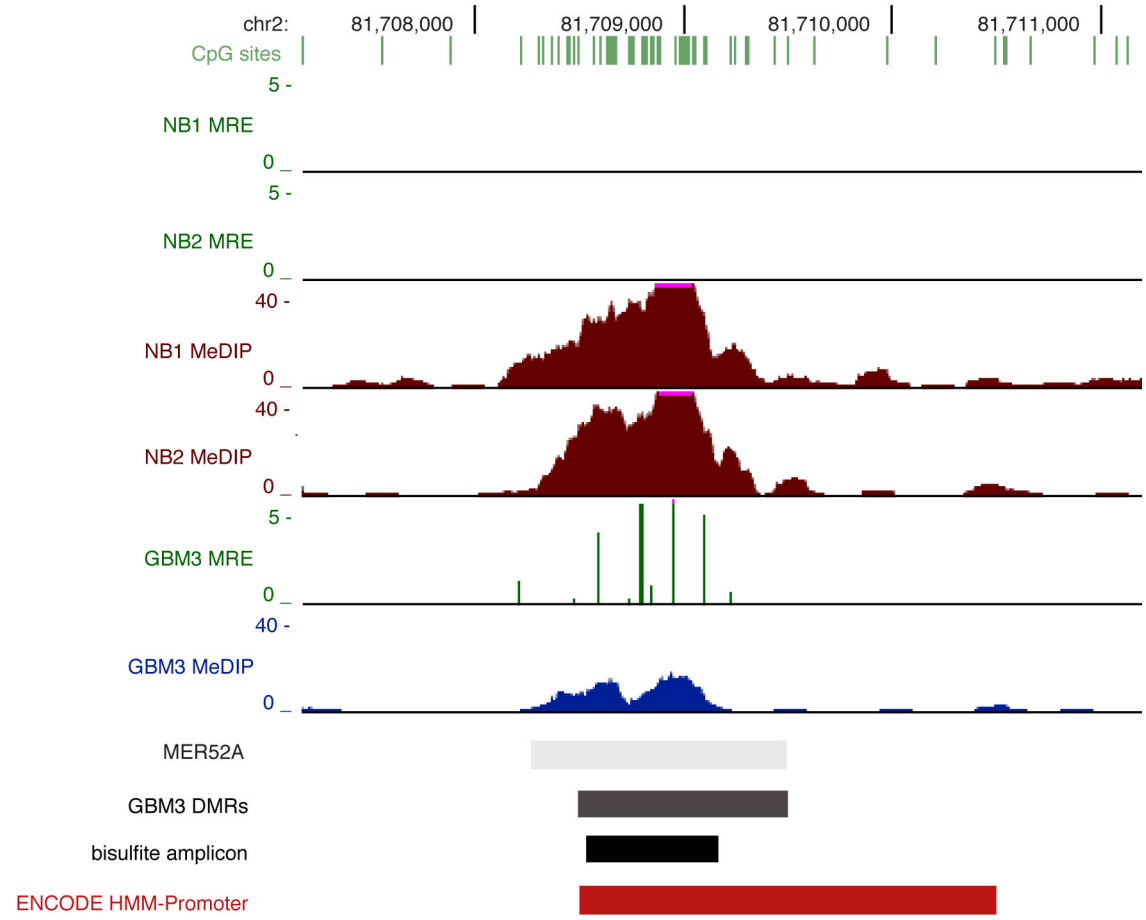


S4B

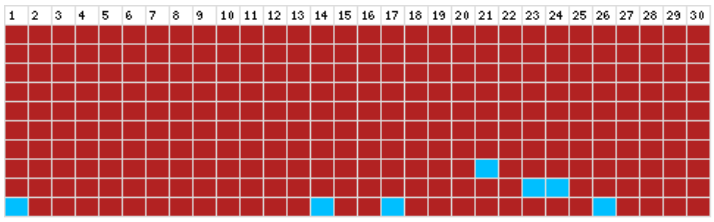


S4. Validation of BCL2L11 hypermethylation by bisulfite PCR, cloning and sequencing. A. Browser shot of hypermethylated region within BCL2L11 5' CpG island. B. Bisulfite sequencing results. Each row is a clone, each column is a CpG site. Blue indicates unmethylated, red indicates methylated. White, no data.

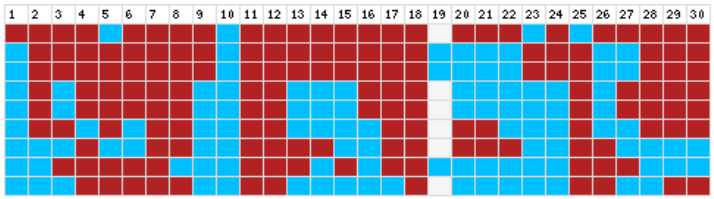
S5A



S5B

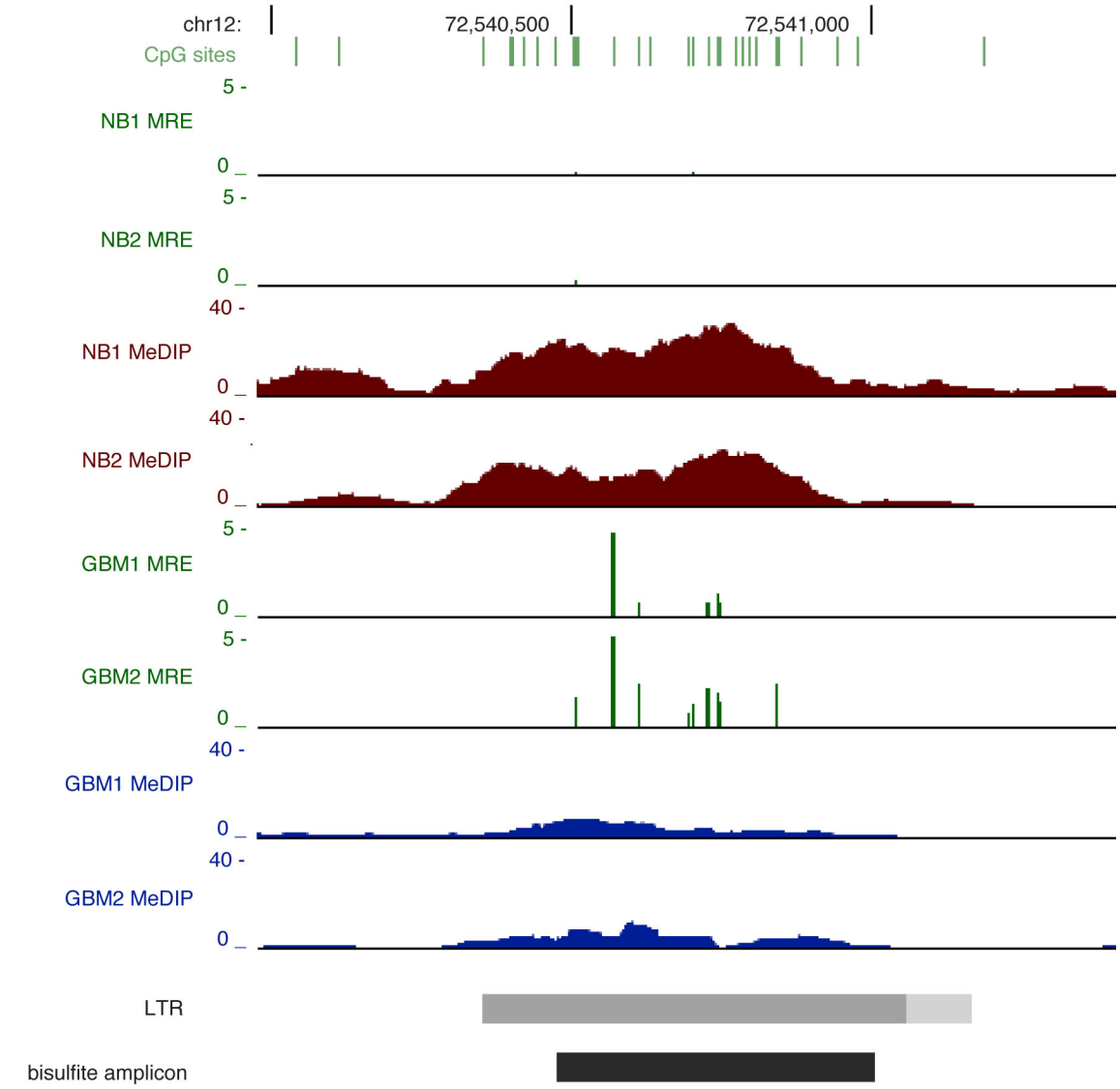


NB1 97.7% methylation

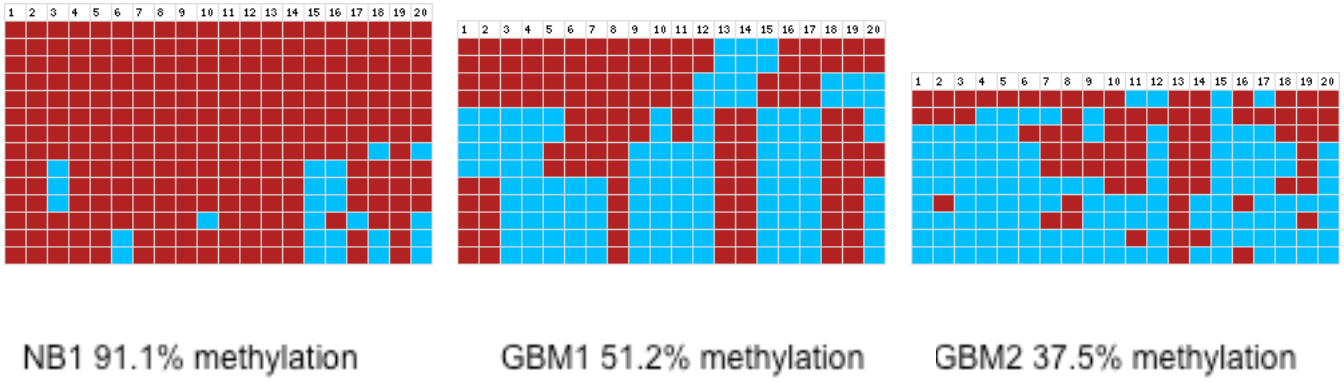


GBM3 59.1% methylation

S5C

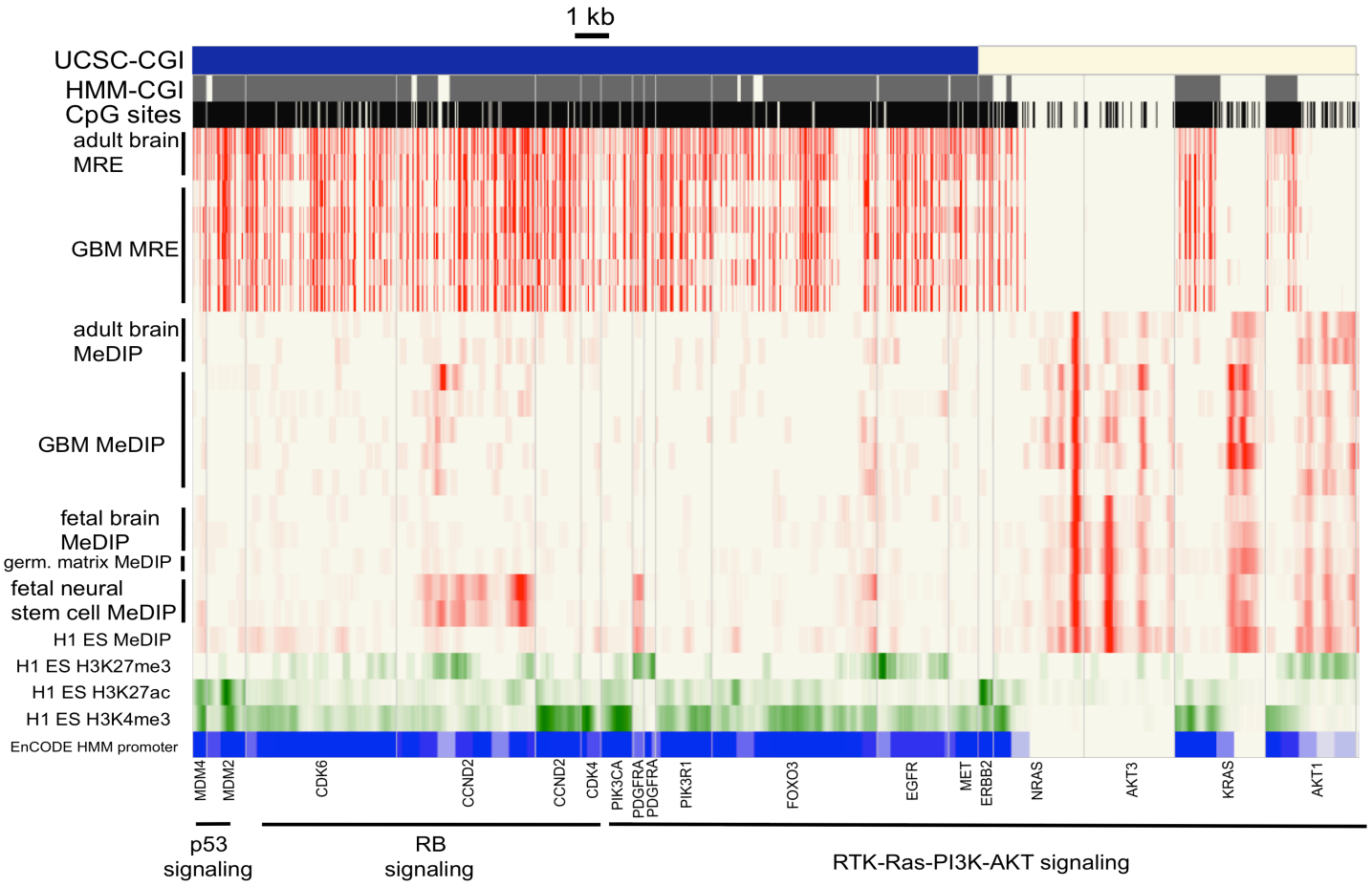


S5D

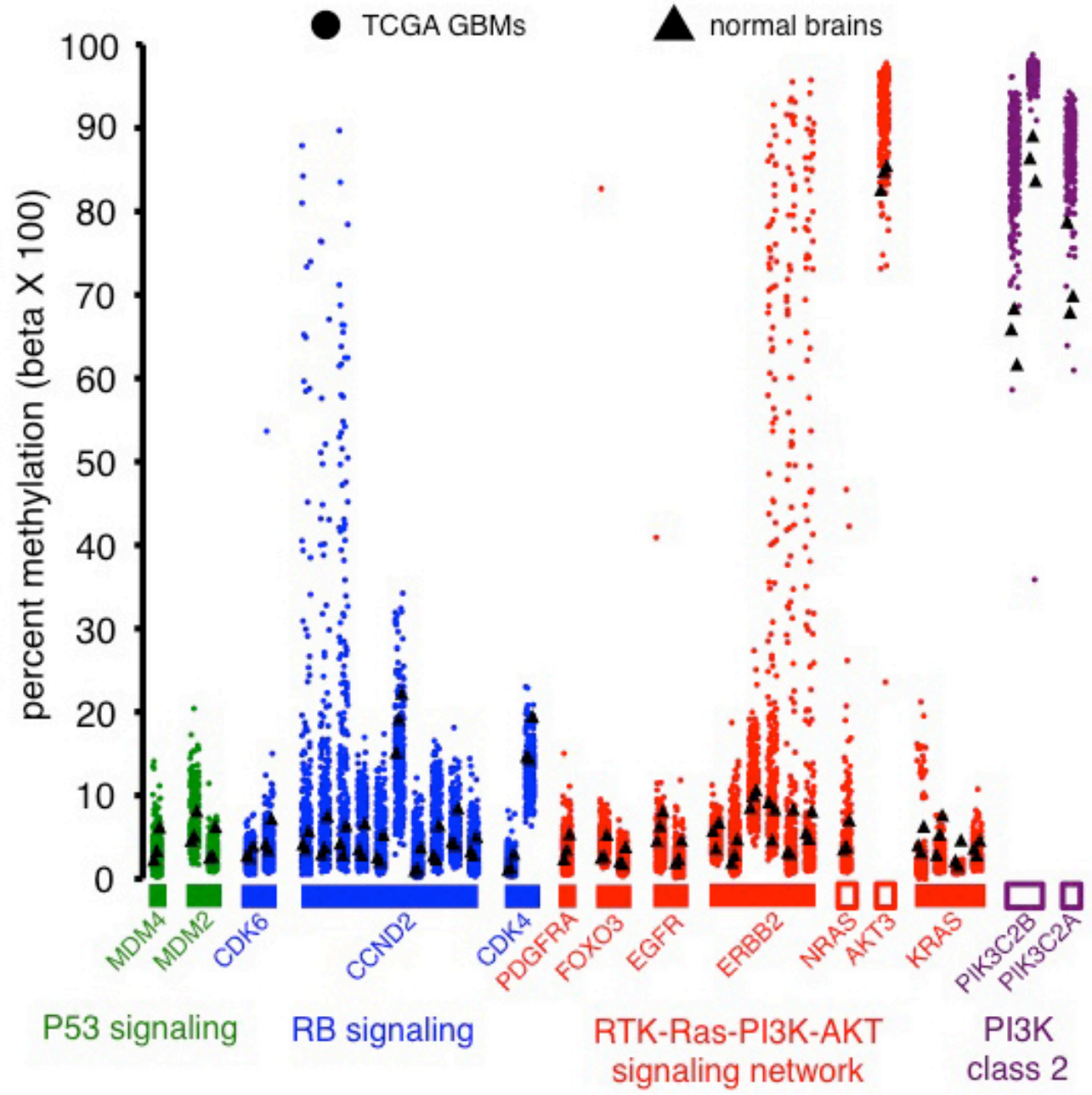


S5. Validation of hypomethylation at repeats. A,C Browser views of hypomethylated MER52A and LTR1B repeats. B,D, Bisulfite sequencing results.

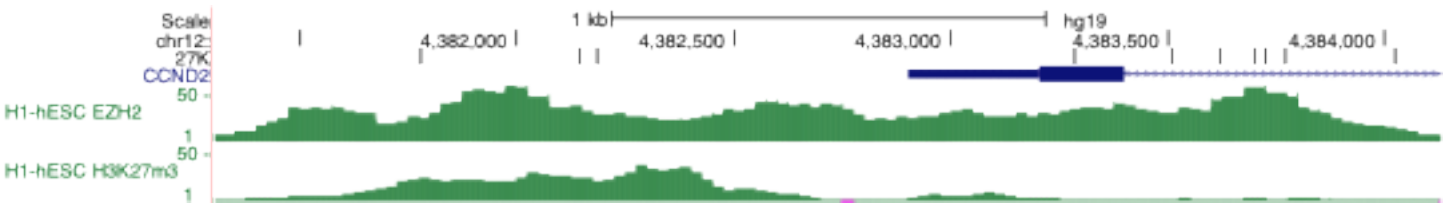
S6A


1 kb
UCSC-CGI
HMM-CGI
CpG sites
adult brain MRE
GBM MRE
adult brain MeDIP
GBM MeDIP
fetal brain MeDIP
germ. matrix MeDIP
fetal neural stem cell MeDIP
H1 ES MeDIP
H1 ES H3K27me3
H1 ES H3K27ac
H1 ES H3K4me3
EnCODE HMM promoter
MDM4
MDM2
CDK6
CCND2
CCND2
CDK4
PIK3CA
PDGFRA
PDGFRA
PIK3R1
FOXO3
EGFR
MET
ERBB2
NRAS
AKT3
KRAS
AKT1
p53 signaling
RB signaling
RTK-Ras-PI3K-AKT signaling

S6B

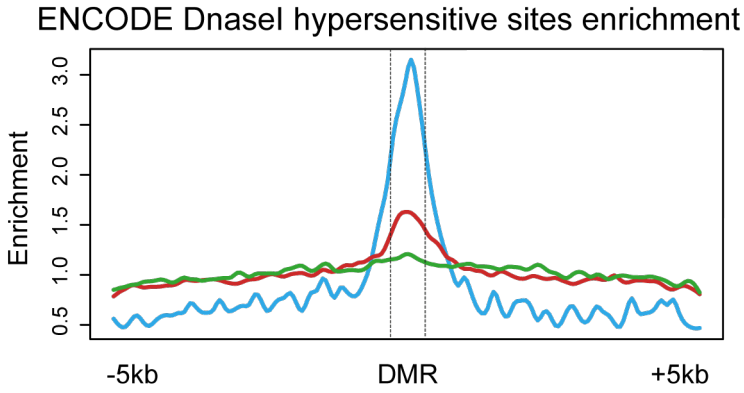

TCGA GBMs
normal brains
percent methylation (beta X 100)
100
90
80
70
60
50
40
30
20
10
0
MDM4
MDM2
CDK6
CCND2
CDK4
PDGFRA
FOXO3
EGFR
ERBB2
NRAS
AKT3
KRAS
PIK3C2B
PIK3C2A
P53 signaling
RB signaling
RTK-Ras-PI3K-AKT signaling network
PI3K class 2

S6C


Scale
chr12:
27K
CCND2
1 kb
hg19
4,382,000
4,382,500
4,383,000
4,383,500
4,384,000
H1-hESC EZH2
50
1
H1-hESC H3K27m3
50
1

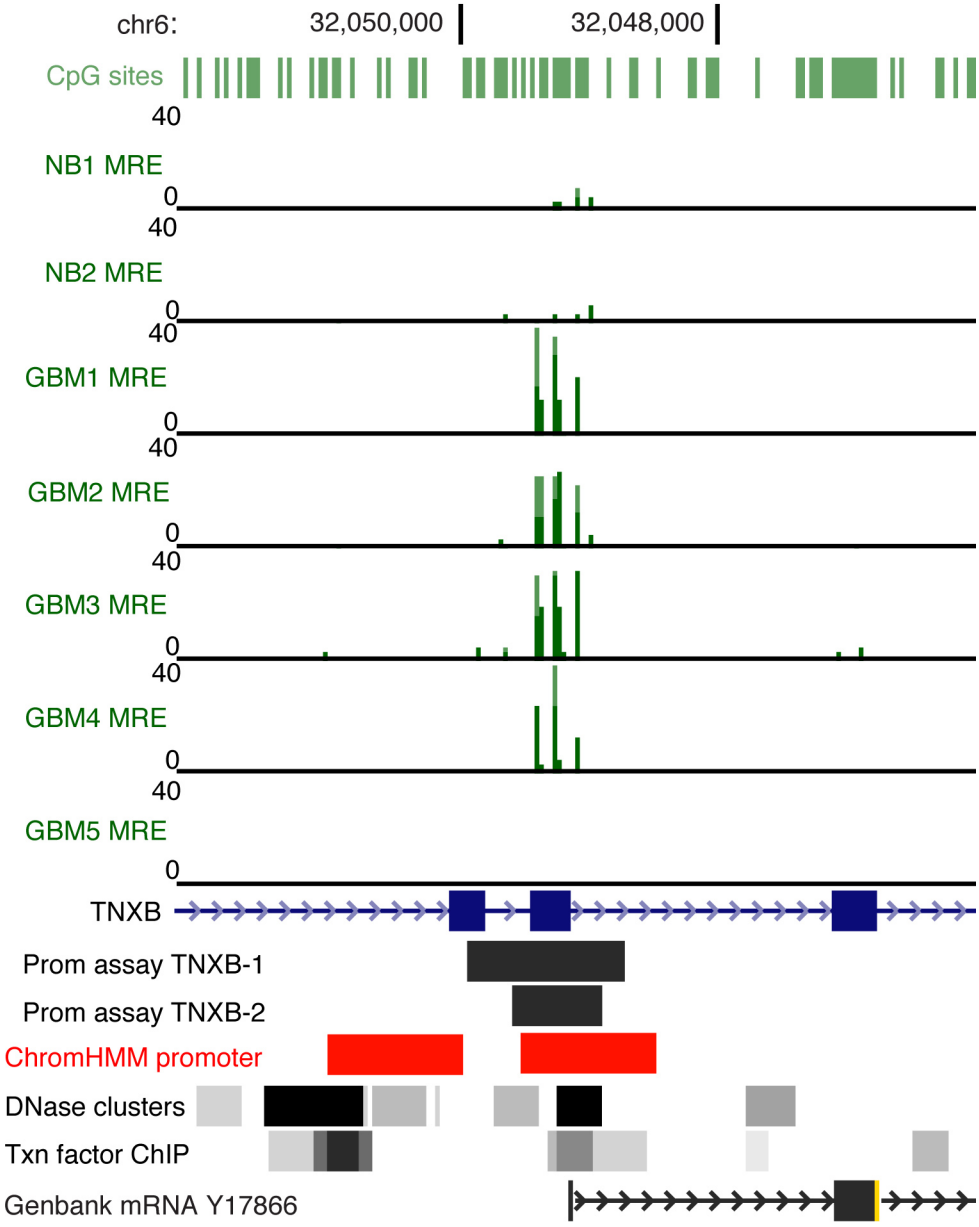
S6. Methylation patterns at common genetically altered GBM oncogene promoters. A, Human Epigenome Browser geneset view of oncogene promoters (McLendon et al. 2008). Different gene promoters are juxtaposed adjacent to one another in this geneset view. For CGI promoters, the entire CGI is shown (promoters on the left). For non-CGI promoters, -2000 to +1000 relative to the TSS is shown. **B**, Infinium 27K methylation data from 292 TCGA GBMs, showing percent methylation for individual CpG sites located in GBM oncogene promoters. GBM Methylation data was downloaded from the TCGA website. Beta values for oncogene promoter CpG sites were extracted and plotted at small colored points. Normal brain data were plotted as black triangles. Colored boxes at bottom group CpG sites at a particular gene promoter; filled boxes indicate CpG island promoters and unfilled non-CGI promoters.

S7

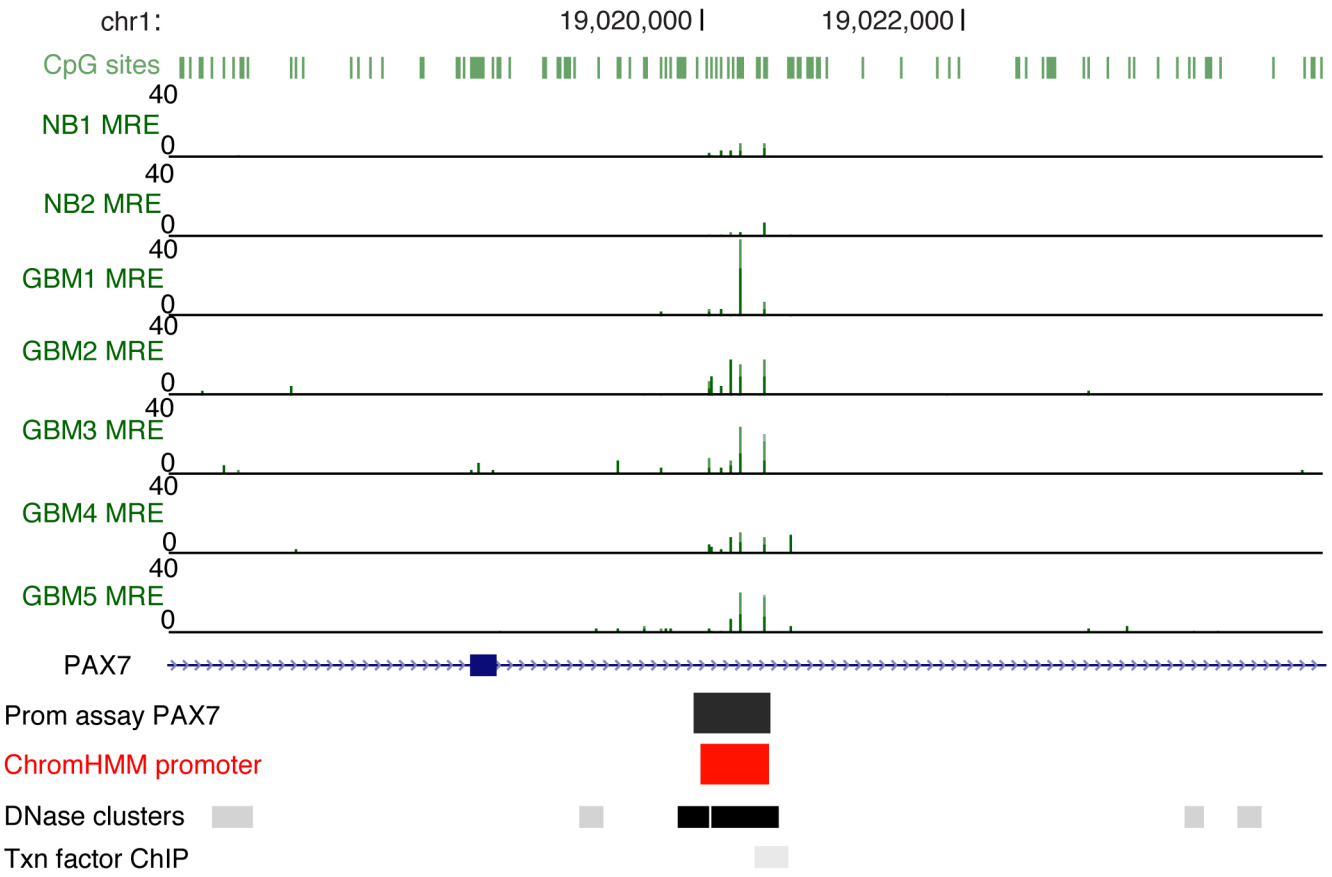


S7. **Spatial enrichments of recurring hypomethylated DMRs for DNase hypersensitivity.** The genomic regions for all recurring hypomethylated DMRs (q<10-13) were expanded to create 10.5 kb windows centered on the middle of each DMR. Digital DNaseI hypersensitivity clusters in 125 cell types from ENCODE were downloaded from the UCSC Genome Browser. The averaged signal from all ENCODE cell lines for DNase hypersensitivity was calculated for each 10.5 kb window. Enrichment in 10 bp bins was then plotted across each 10.5 kb window (red line). As a positive control, enrichment for the set of brain-specific hypomethylated DMRs (hypomethylated in brain relative to breast, blood and ES cells) (Zhang et al. 2013) is plotted in blue. Signal for randomly chosen windows was plotted as a negative control (green line). Using the criterion that DMRs have over 50% length overlapping to ENCODE DHS, in 558 hypomethylated DMRs (q<1e-13), 136 DMRs (24.37%) contain DHS and this is a statistically significant enrichment using a hypergeometric test, p-value = 1.2X10-8.

S8A

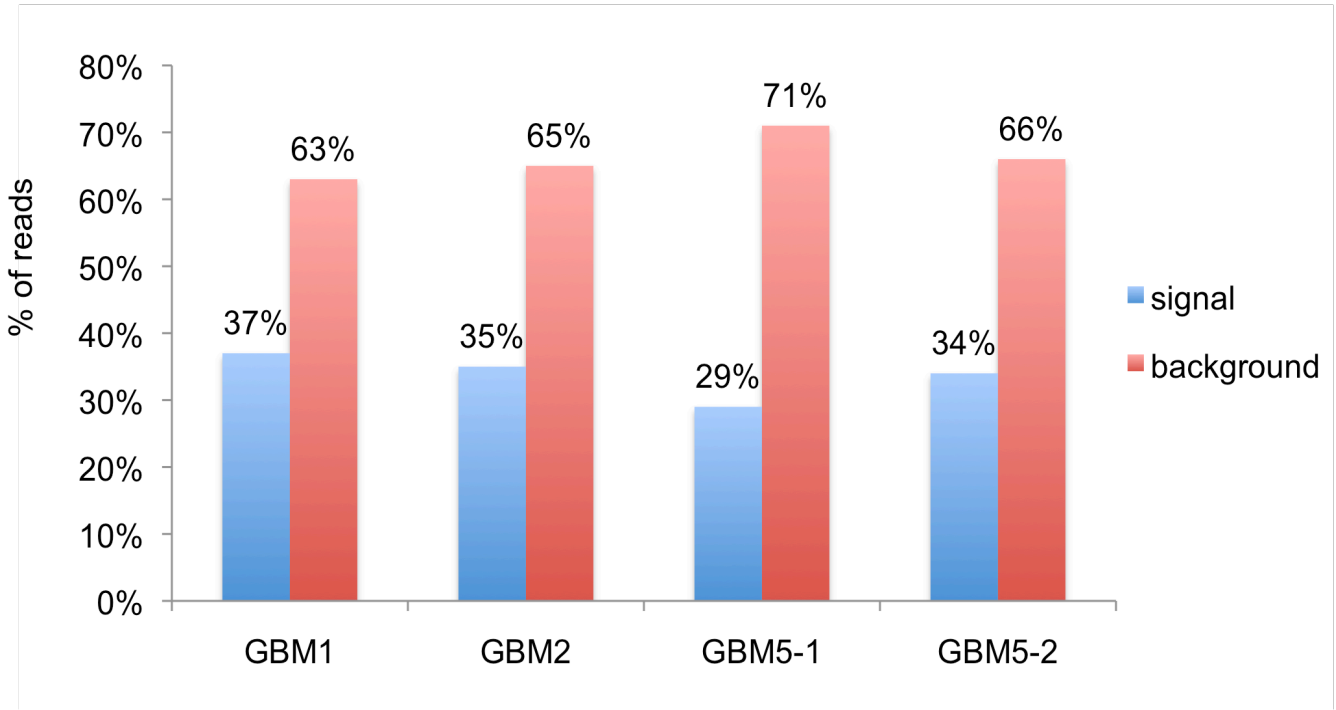

chr6:
32,050,000
32,048,000
CpG sites
40
NB1 MRE
0
40
NB2 MRE
0
40
GBM1 MRE
0
40
GBM2 MRE
0
40
GBM3 MRE
0
40
GBM4 MRE
0
40
GBM5 MRE
0
TNXB
Prom assay TNXB-1
Prom assay TNXB-2
ChromHMM promoter
DNase clusters
Txn factor ChIP
Genbank mRNA Y17866

S8B

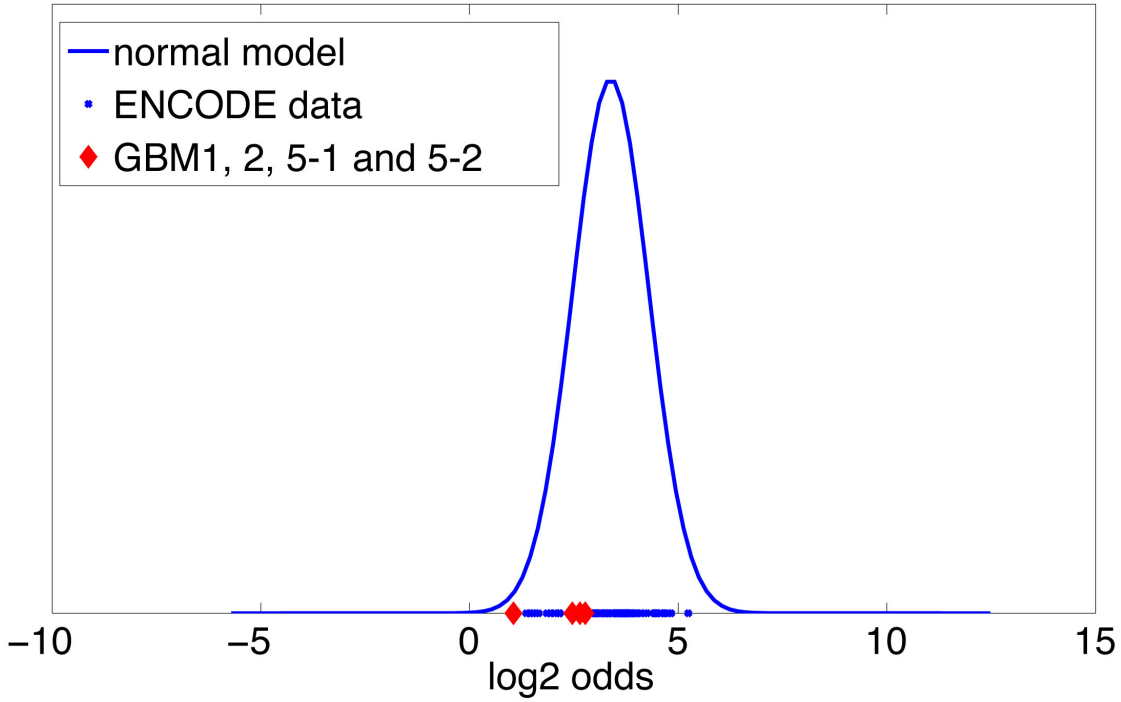


S8. Chrom-HMM chromatin-state defined gene body promoters with confirmed promoter activity by reporter assay. Browser images for the two gene body loci tested by luciferase and shown in Figure 3B-D (other than SIGLEC11 which is shown in Figure 3E). **A,** Confirmed gene body promoter within *TNXB*, showing GBM hypomethylation by MeDIP-/MRE-seq and overlapping promoter state by ChromHMM. The 5' end of a human mRNA (T17866) is shown at the bottom. **B**, Confirmed gene body promoter within *PAX7*.

S9A



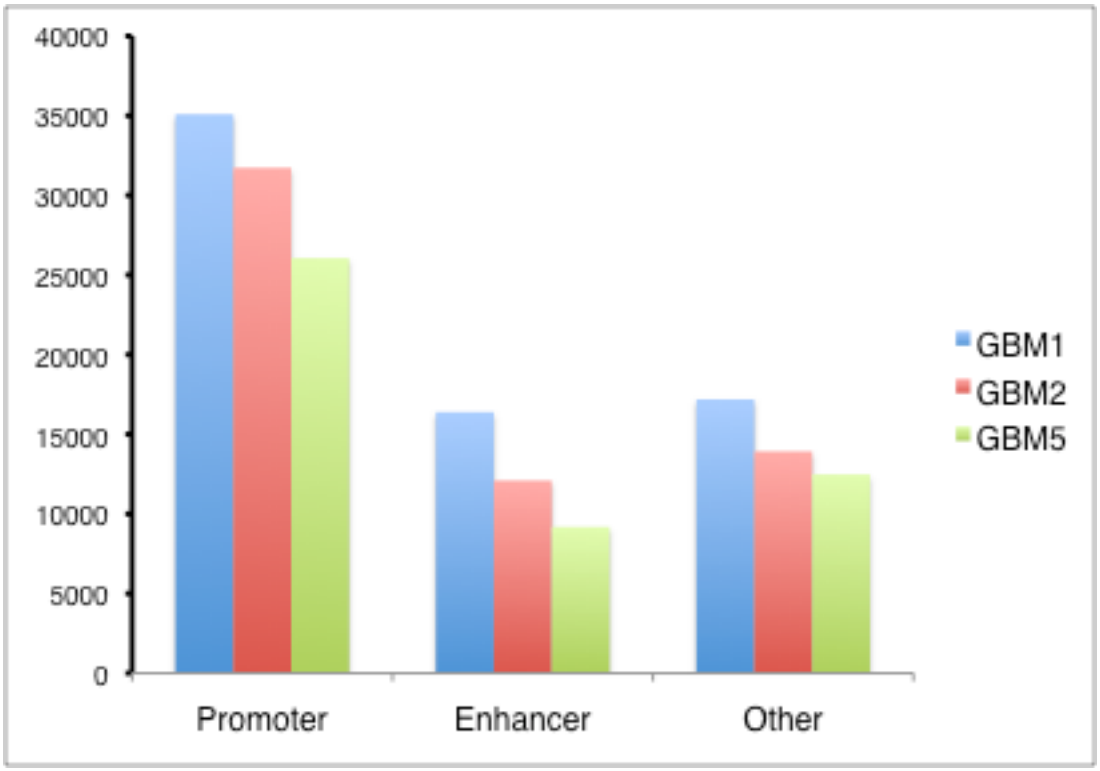
S9B



S9. Quality control analysis of H3K4me3 ChIP-seq by CHANCE. A, The IP strengths of ChIP-seq experiments for GBMs 1, 2, and two replicates of GBM5 (5-1 and 5-2) were assessed by Signal Extraction Scaling (SES) (Diaz et al. 2012b) using the software CHANCE (Diaz et al. 2012a), which estimates the background and signal portions of the IP by determining the largest subset of the genome where the distribution of cumulative tag density in the IP sample matches that of the Input sample. All samples were comprised of 29-37% signal reads. These percentages represent a statistically significant IP/Input differential enrichment as measured by a divergence test with a q-value (positive false discovery rate) of 0.000712. **B,** Using CHANCE,

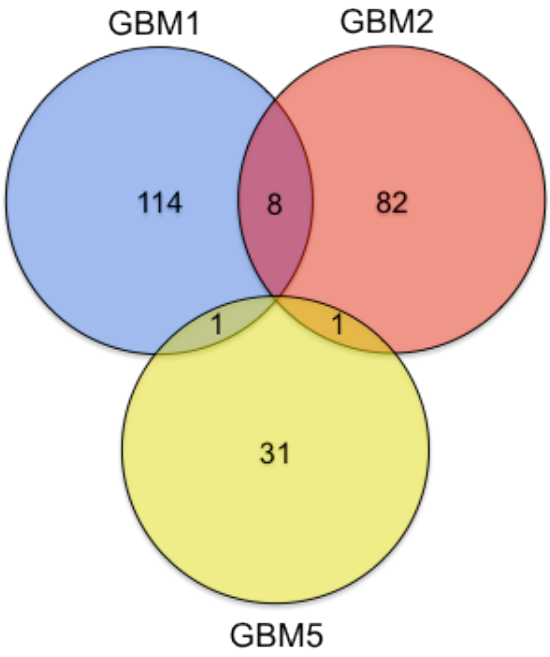
GBM data were compared to called peaks for all ENCODE (Meyer et al. 2013) peak data for H3K4me3 in human, see [2] supplemental file 2 for a full list of datasets used. CHANCE constructs the union of all ENCODE H3K4me3 called peaks and then measures the percentage of reads in the IP which map to a union peak versus background and compares that to Input. Expressed as an odds ratio, this comparison estimates the enrichment in loci of known H3K4me3 modification as documented by ENCODE. On the x-axis the log2 odds ratios of samples SF4948 (log2 odds = 2.4649), SF4060 (log2 odds = 2.7728), SF7996-1 (log2 odds = 1.054) and SF7996-2 (log2 odds = 2.6461), are marked in red while the log2 odds ratios of the ENCODE data are marked in blue. All samples show considerable enrichment in known peaks as demonstrated by positive log2 odds ratio.

S10



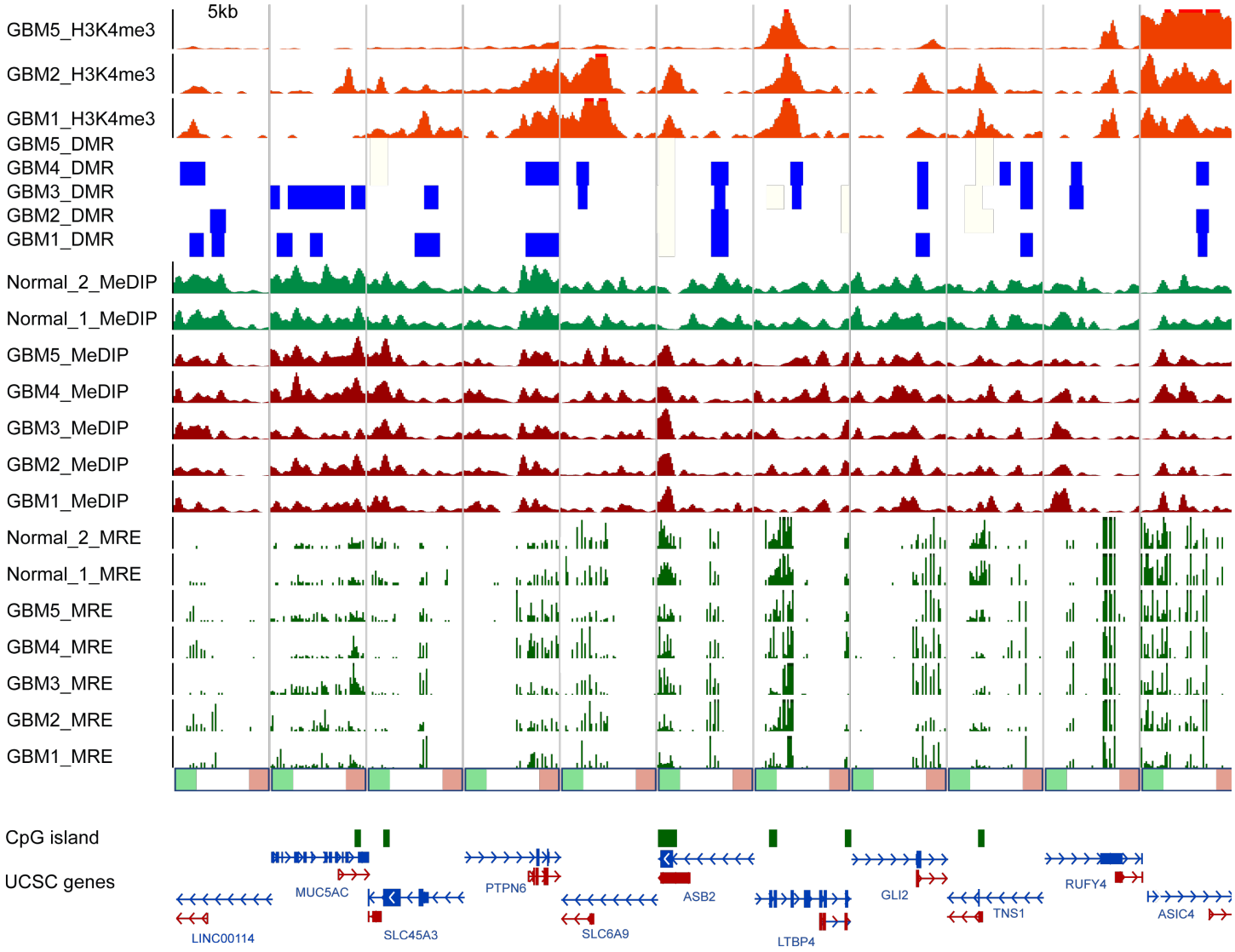
S10. Number of ChromHMM elements that intersect H3K4me3 peaks. For GBMs 1,2 and 5, the number of peaks overlapping with ChromHMM promoters, enhancers, or other elements defined in ENCODE cell lines are shown.

S11



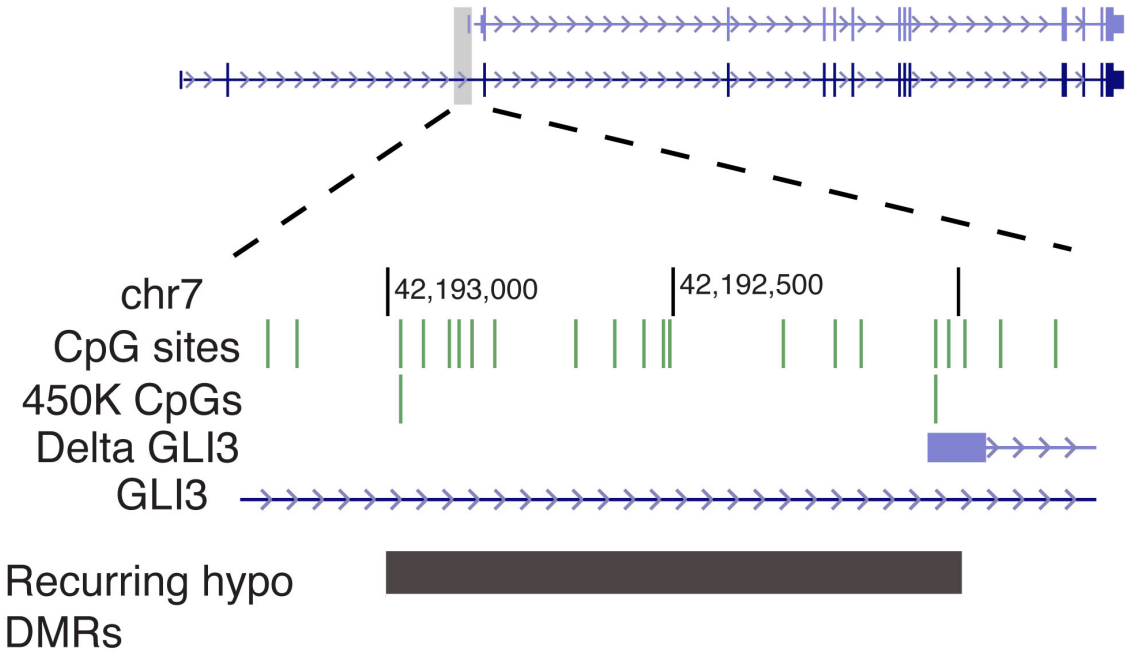
S11. DNA hypo/K4me3 loci in GBM. A, Number and overlap of DNA hypo/K4me3 loci in GBMs 1,2 and 5. B, Gene expression change (GBM-NB) for the gene nearest each DNA hypo/K4me3 locus, which, taking into account the direction of transcription for each gene, can be located upstream (5'), within the gene body, or downstream (3'). Statistically significant gene expression changes were found for genes with gene body DNA hypo/K4me3 in GBM1 and GBM5 (starred).

S12

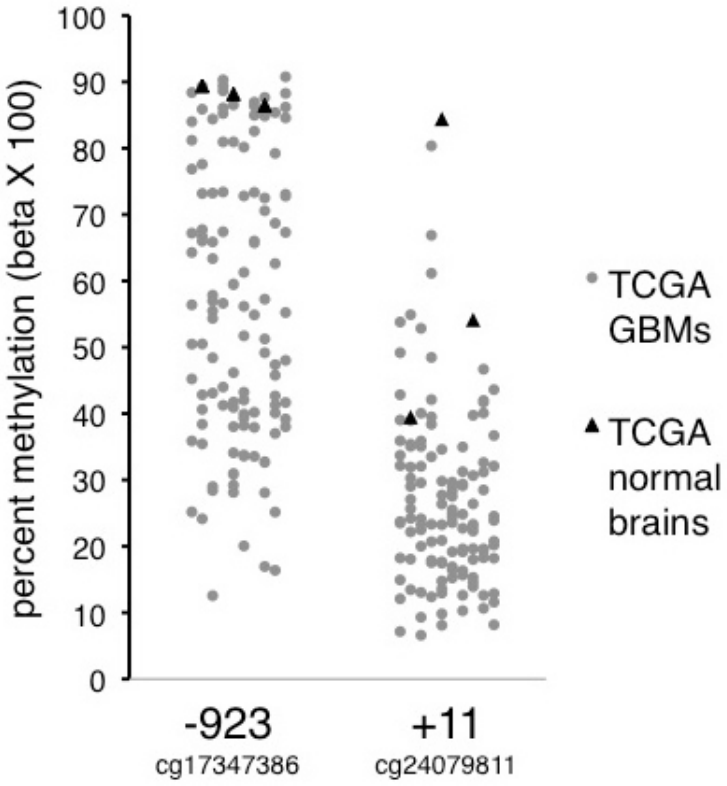


S12. Recurrent hypomethylated gene body promoters with H3K4me3 peaks. WashU Epigenome Browser (Zhou and Wang 2012) shot of 11 of the 22 gene body promoters we identified with both recurrent hypomethylated DMRs and H3K4me3 peaks. Different promoters are juxtaposed together in this image using the genomic juxtaposition function.

S13A

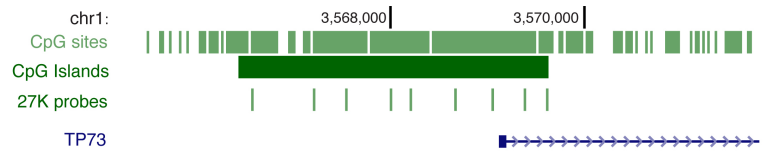


S13B

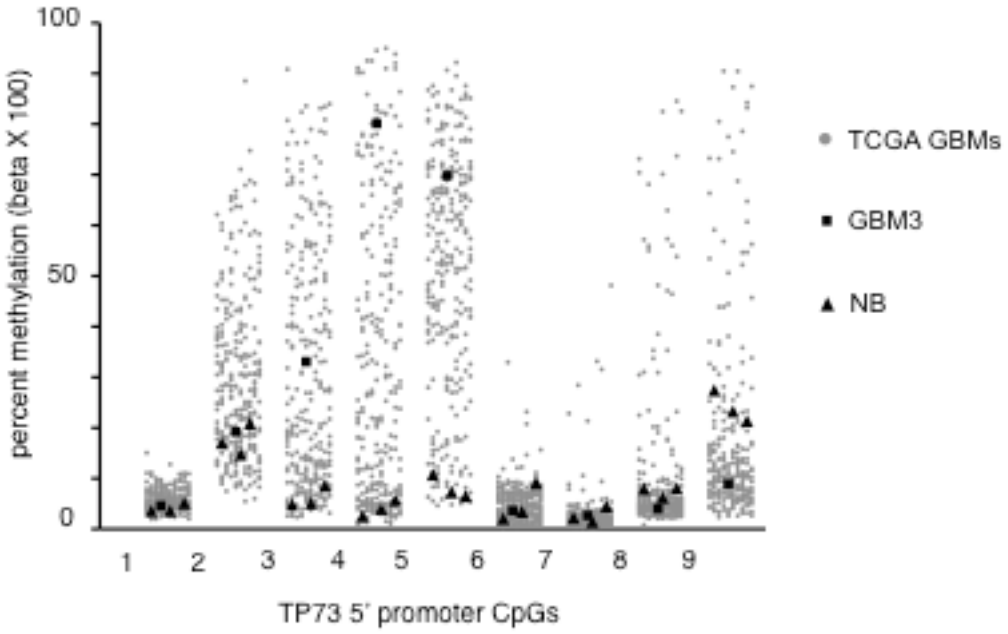


S13. Recurrent GLI3 gene body promoter hypomethylation in TCGA GBMs. A, Browser view of two isoforms of GLI3. Bottom shows zoomed in on Delta GLI3 gene body promoter with Illumina Infinium 450K CpG probes and the hypomethylated DMRs identified by M&M. **B,** Infinium 450K methylation data from 126 TCGA GBMs for two CpG sites (-923 and +11 relative to the gene body transcription start site).

S14A

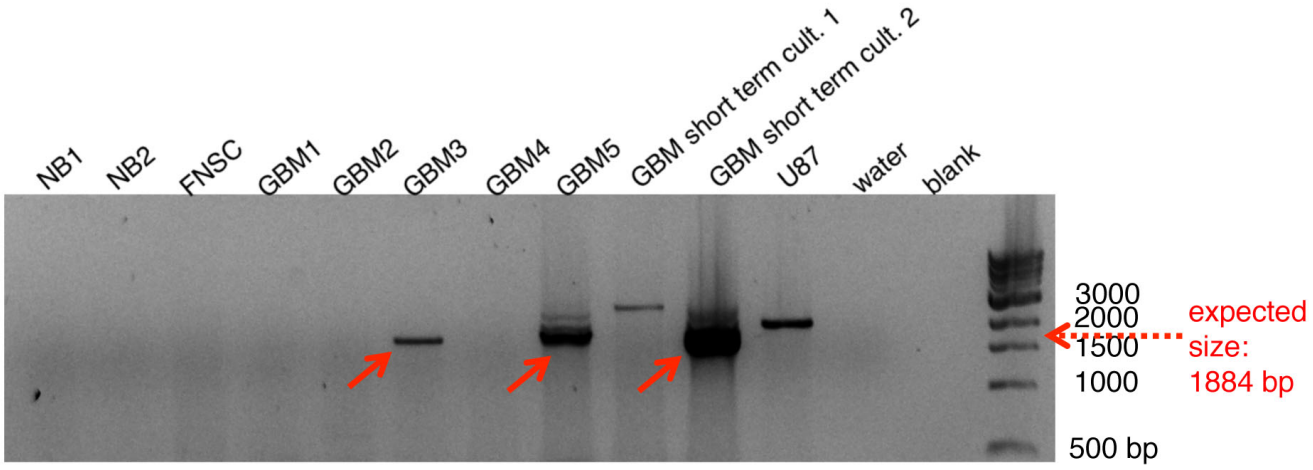


S14B

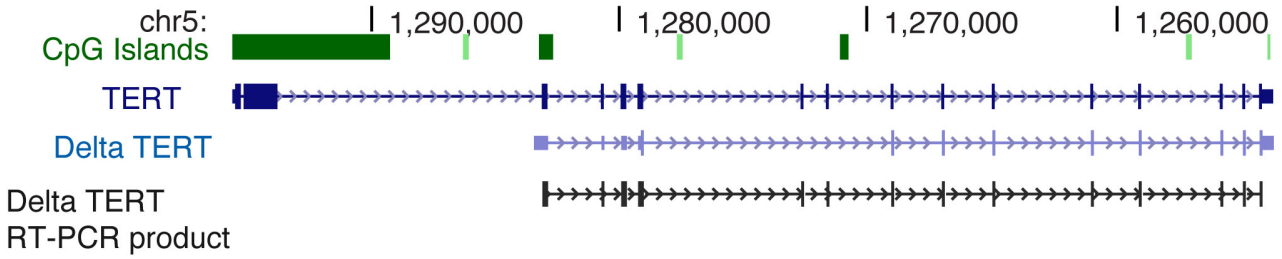


S14. Infinium HumanMethylation27 array probe locations (A) and percent methylation (B) at the 5' CGI of TP73. TCGA GBMs, gray points. GBM3, black squares. Normal brains, black triangles.

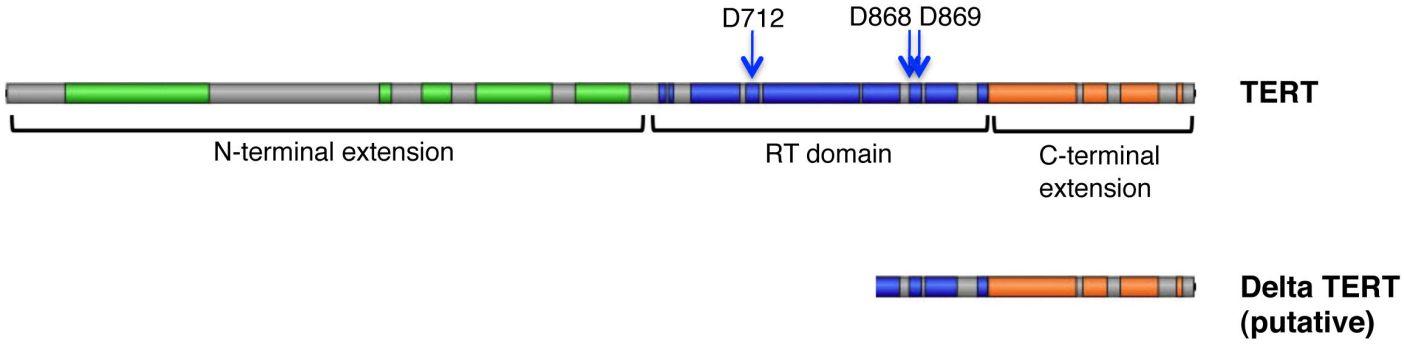
S15A



S15B



S15C



S15. Delta TERT transcript initiating from the gene body A, Exon-joining RT-PCR to detect delta TERT alternate transcript initiating in the hypomethylated gene body. RT-PCR was performed with primers designed to amplify a 1884 bp product specifically from the delta TERT transcript, with the forward primer in the portion of delta TERT exon 1 that is intronic in full-length TERT, and the reverse primer in the shared 3' UTR. A product of approximately 1900 bp was detected in several GBM cDNAs (red arrows), and the bands were gel-purified, TOPO TA cloned, and sequenced in both directions. **B,** Sequencing results shown schematically on the UCSC browser. The sequencing indicated a transcript with an exon splicing pattern exactly matching that of the delta TERT transcript annotated by UCSC genes database, with exons 7 and 8 (based on full-length TERT

numbering) spliced out. **C,** Domain structure of full-length and putative Delta TERT protein. Amino acids shared with full-length TERT are shown. The three invariant aspartic acid residues that are essential for reverse transcriptase catalytic activity in the RT domain are shown at the top. D712 is not present in Delta TERT, along with most the RT domain, suggesting a complete lack of reverse transcriptase activity.

Supplementary References

Diaz A, Nellore A, Song JS. 2012a. CHANCE: comprehensive software for quality control and validation of ChIP-seq data. *Genome Biol* **13**(10): R98.

Diaz A, Park K, Lim DA, Song JS. 2012b. Normalization, bias correction, and peak calling for ChIP-seq. *Statistical applications in genetics and molecular biology* **11**(3).

McLendon R Friedman A Bigner D Van Meir EG Brat DJ M Mastrogianakis G Olson JJ Mikkelsen T Lehman N Aldape K et al. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**(7216): 1061-1068.

Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**(Database issue): D64-69.

Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, Cheng JB, Li D, Stevens M, Lee HJ et al. 2013. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res*.

Zhou X, Wang T. 2012. Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr Protoc Bioinformatics* **Chapter 10**: Unit10 10.