Supplemental Figures

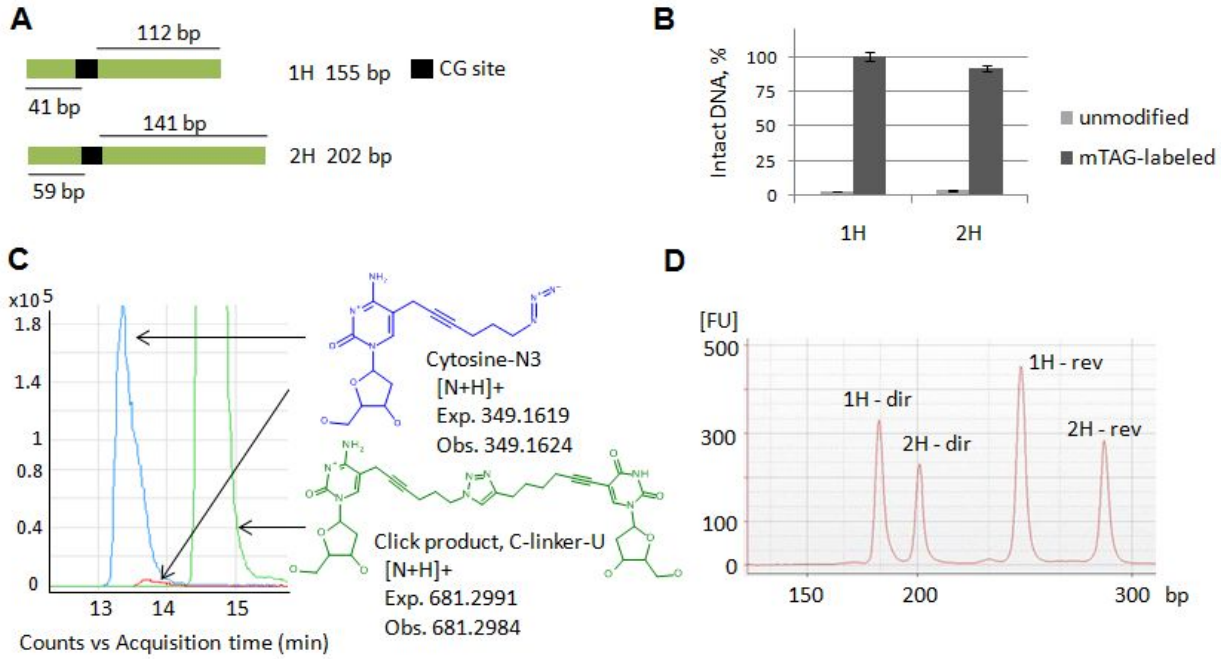

Figure S1. Step-wise assessment of the TOP-seq procedure in a model system, related to Figure 1. A, Schematic representation of model DNA fragments (1H, 155 bp and 2H, 202 bp) which lead to four asymmetric DNA amplicons (shown as black lines) after CG-specific tagging and TOP amplification. **B,** qPCR analysis of the eM.SssI-directed modification of uCG sites in model DNA. A nearly complete alkylation of uCG sites was observed judging from the fraction of DNA that is protected from cleavage by R.Hin6I restriction endonuclease. **C,** Chemical tethering of a priming oligonucleotide to azide-tagged CG sites. Efficiency of the azide-alkyne cycloaddition reaction was determined by HPLC-MS analysis of reaction products. Mass chromatograms of nucleosides show efficient conversion of the azide-tagged cytosine (cytosine-N3, blue) into the Click product (C-linker-U conjugate, green). Only a trace amount of unconverted azide-modified C (red) is observed after the reaction. Theoretical (Exp.) and observed (Obs.) masses of respective protonated nucleoside (N) products are as indicated. **D,** Agilent Bioanalyzer profiles of TOP-seq priming products obtained from 1H and 2H model DNA fragments. Since either cytosine of the double-stranded CG site can be tagged by M.SssI in the mTAG reaction (Fig. S1A), the primed DNA synthesis results in two DNA products comprising the upper and bottom DNA strands of template DNA. Theoretical (and experimentally derived) sizes of the TOP amplicons (including 135 bp adapters) are as follows: 1H-dir 176 (183) bp, 2H-dir 194 (202) bp, 1H-rev 247 (248) bp, 2H-rev 276 (284) bp.

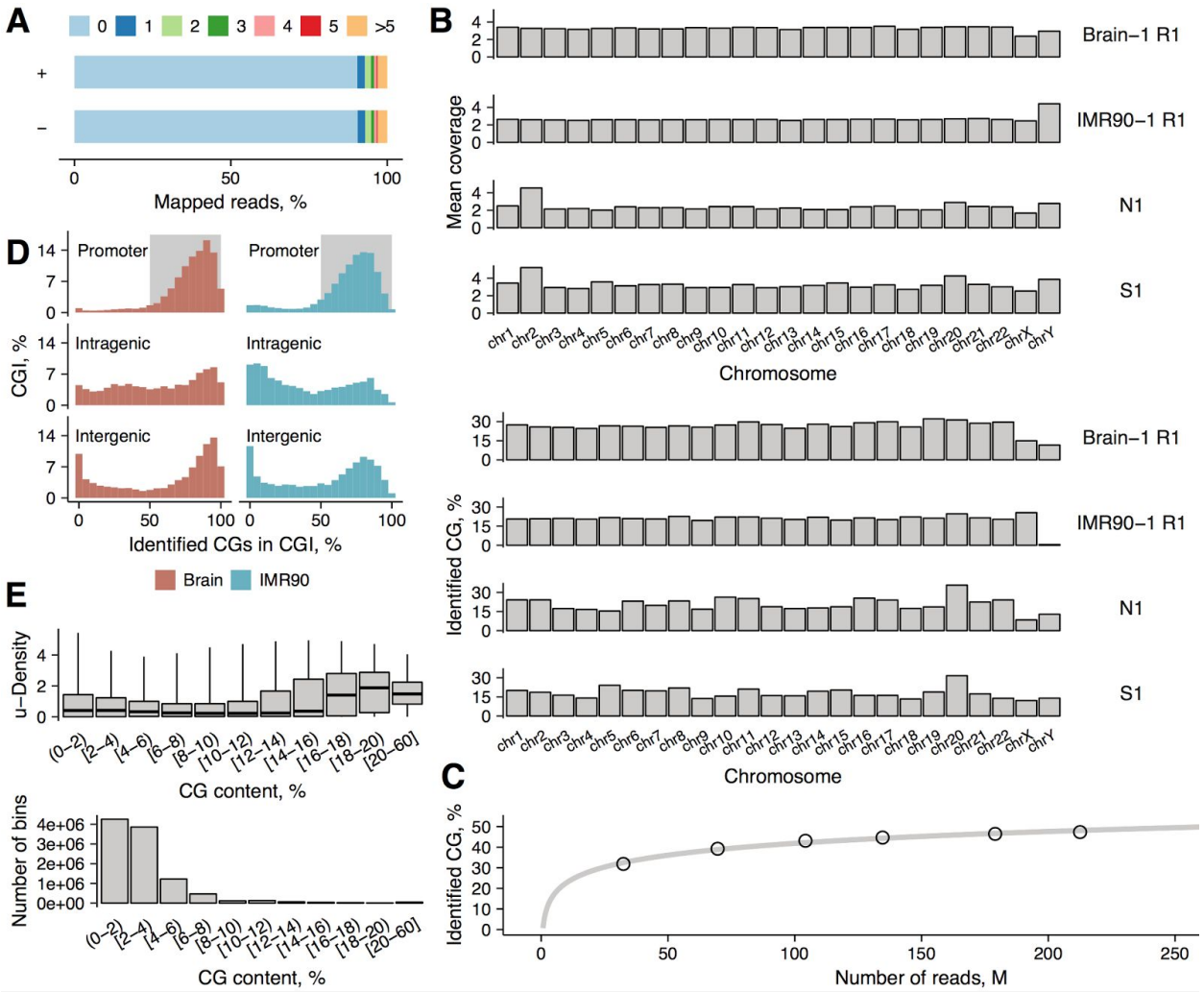

Figure S2. TOP-seq analysis of the human genome, related to Figure 3. A, Distribution of read distances to the nearest CG site in the Brain-1 R1 library. Reads from the upper and lower DNA strands are shown as "+" and "-", respectively. **B,** Mean sequencing coverage of identified uCGs and the amount of identified uCGs per chromosome (for the Brain-1 R1, IMR90-1 R1, N1 and S1 libraries). **C,** The relationship of the fraction of identified CG sites on the number of useful TOP-seq reads using six IMR90 PI chips. Altogether, six PI chips result in ~220 M of reads and cover ~47% of all CG sites in the human genome (16x coverage of identified uCGs). The grey line shows a logarithmic fit extrapolating the growth of identified uCGs. "Low coverage" TOP-seq library with 30 million useful (86 million raw reads, derived from one PI chip) identifies 33.5% of CGs (the most unmethylated fraction of the human genome) with 4.1x mean coverage. **D,** Histograms showing the percentage of promoter, intragenic and intergenic CG islands stratified according to percentage of TOP-seq identified CGs in the CGIs. TOP-seq identifies 50-100% CGs in 93-96% of promoter CGIs (grey box area) indicating their highly unmethylated state. **E,** (Top) Distribution of u-density stratified according to CG content. TOP-seq u-density bandwidth 180 bp, CG density bandwidth 80 bp were used; (Bottom) Number of bins with different CG content.

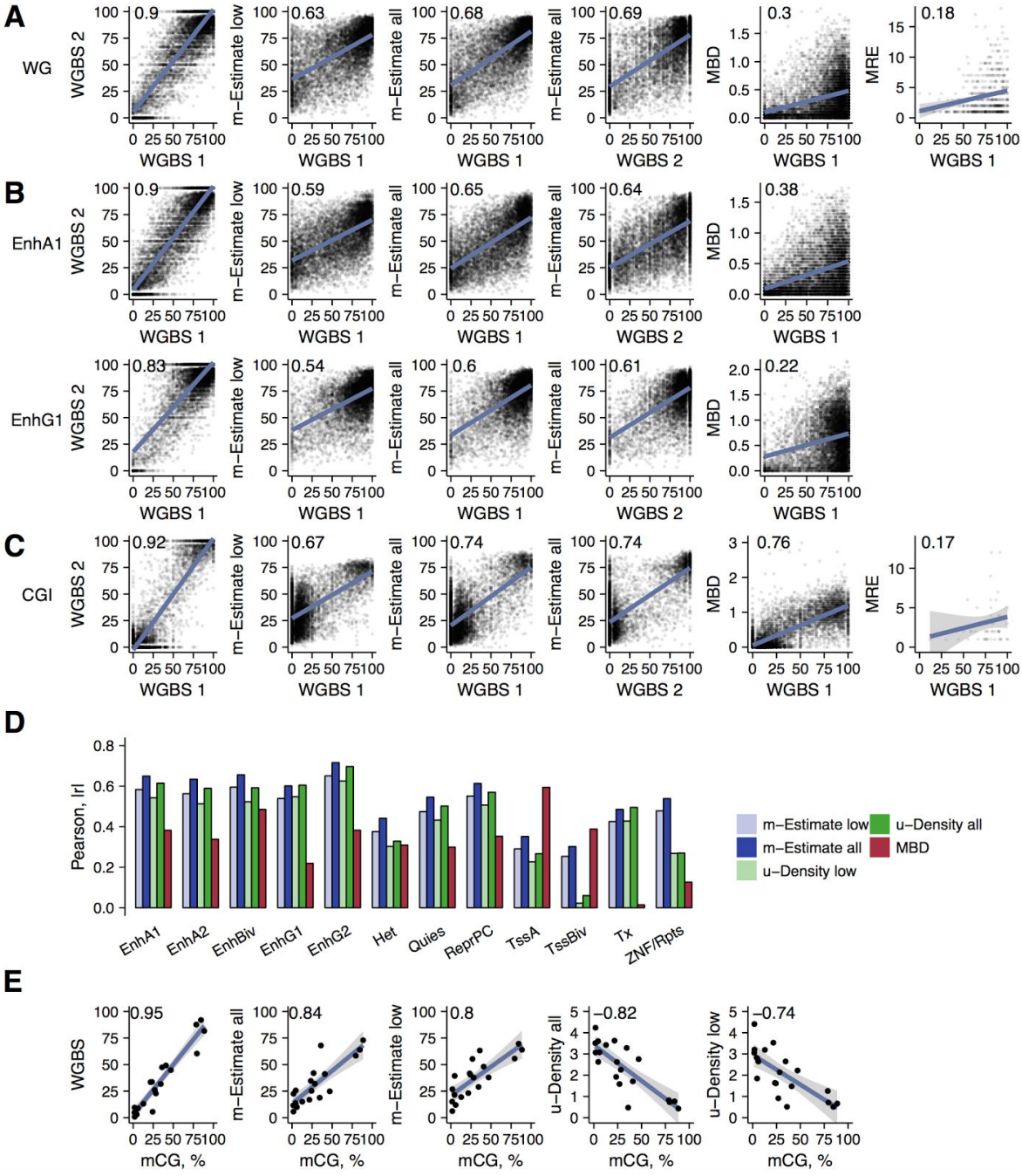

Figure S3. Technology comparison for IMR90 methylome profiling at single CG resolution, related to Figure 3. **A,** Scatterplots illustrating the correlation between whole-genome datasets obtained using respective methylome profiling methods: WGBS 1 (Lister et al., 2009); WGBS 2 (Ziller et al., 2013); TOP-seq m-estimates of low (IMR90-1 R2) or high coverage TOP-seq ('all' means the combined IMR90 dataset, Table S1); MBD-seq (data from Bert et al., (2013); each CG in an analysis window received the same methylation score); MRE-seq. Pearson r values are shown above the graphs. **B,** Scatterplots demonstrating correlation between the methylome profiling methods across active enhancers A1 and genic enhancers G1 of IMR90 (chromatin data from Kundaje et al., 2015). Comparison of MRE-seq with WGBS is not shown due to insufficient data for analysis. **C,** Correlation scatterplots across CG islands. **D,** Correlation between TOP-seq data (u-density and m-estimates), MBD-seq and WGBS across various chromatin states. **E,** Comparison of TOP-seq with pyrosequencing in 20 selected regions in CGIs and putative enhancers of IMR90. Distribution of IMR90 WGBS 1 and TOP-seq signals ("low" data of IMR90-1 R2 and "all" data of the combined IMR90 dataset) compared to pyrosequencing CG methylation values. Average methylation values of the respective method across the regions were used for comparison.

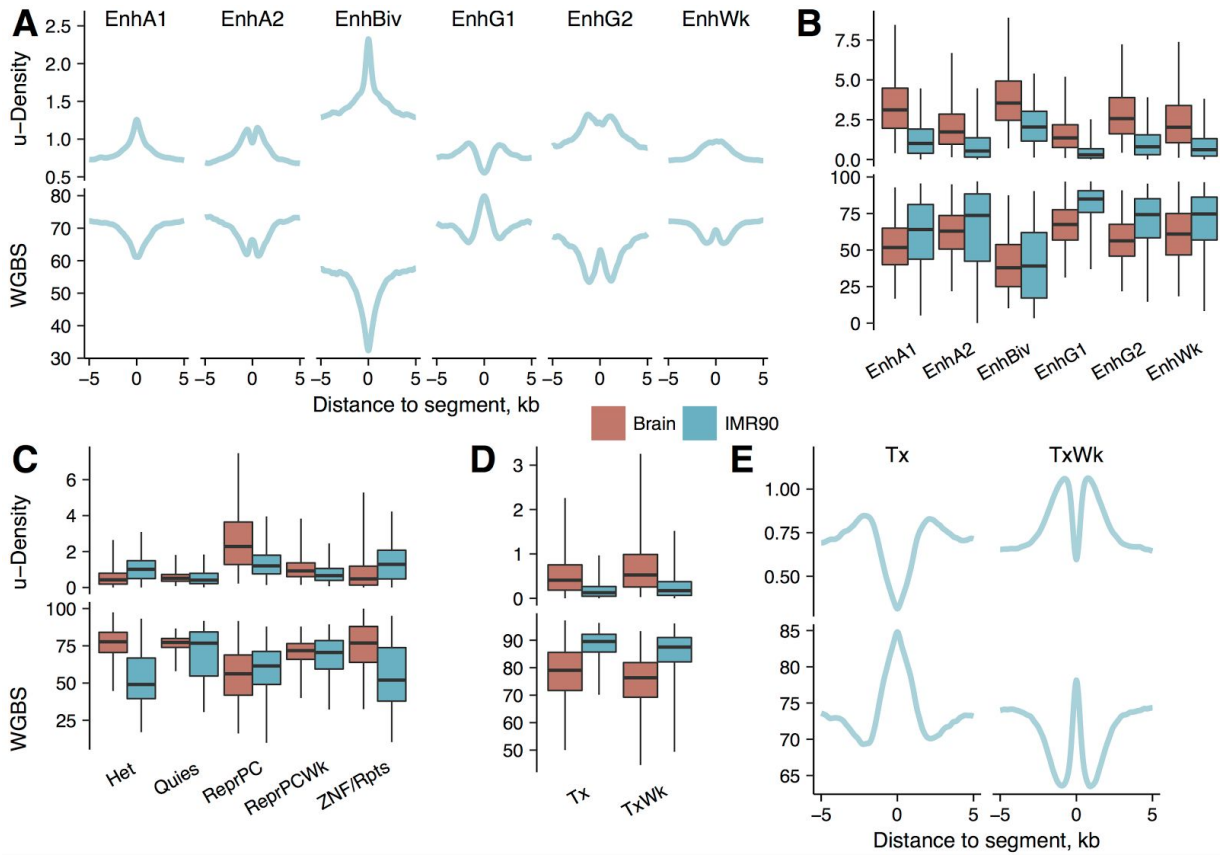

Figure S4. DNA methylation and TOP-seq u-density profiles at various chromatin states, related to Figure 4. A, Smoothed profiles of TOP-seq u-density and WGBS (IMR90; Lister et al., 2009) 5 kb upstream and downstream around enhancer segments. Enhancer states were defined as specific chromatin states (Kundaje et al., (2015)) comprising: two groups of active enhancers (containing strong H3K27ac signal - EnhA1/EnhA2); two groups of genic enhancers (EnhG1/EnhG2); bivalent enhancers (EnhBiv) and weak enhancers (EnhWk). Previous reports (Kundaje et al, 2015) and our studies found decreased DNA methylation in bivalent and active enhancers, while genic enhancer elements showed intermediate methylation levels. **B,** Mean TOP-seq u-density and WGBS (Lister et al., 2009; Wen et al., 2014) for different enhancer related chromatin states. **C,** Mean TOP-seq u-density and WGBS in various repressed chromatin segments: heterochromatin (Het); quiescent state (Ques); repressed polycomb (ReprPC); repressed polycomb weak (ReprPCWk) and ZNF repressed state (ZNF/Rpts). Increased methylation was detected in two inactive states, Het and Ques, consistent with previously reported overall hypermethylation of these segments in human tissues and their relative hypomethylation in IMR90 DNA. **D,** Mean TOP-seq u-density and WGBS in two actively transcribed states, strong (Tx) and weak transcription (TxWk) **E,** Profiles of TOP-seq u-density and WGBS (IMR90) 5 kb upstream and downstream of Tx and TxWk segments. Between the two actively transcribed states, the regions of higher expression (Tx) are more extensively methylated than those of lower expression (TxWk).

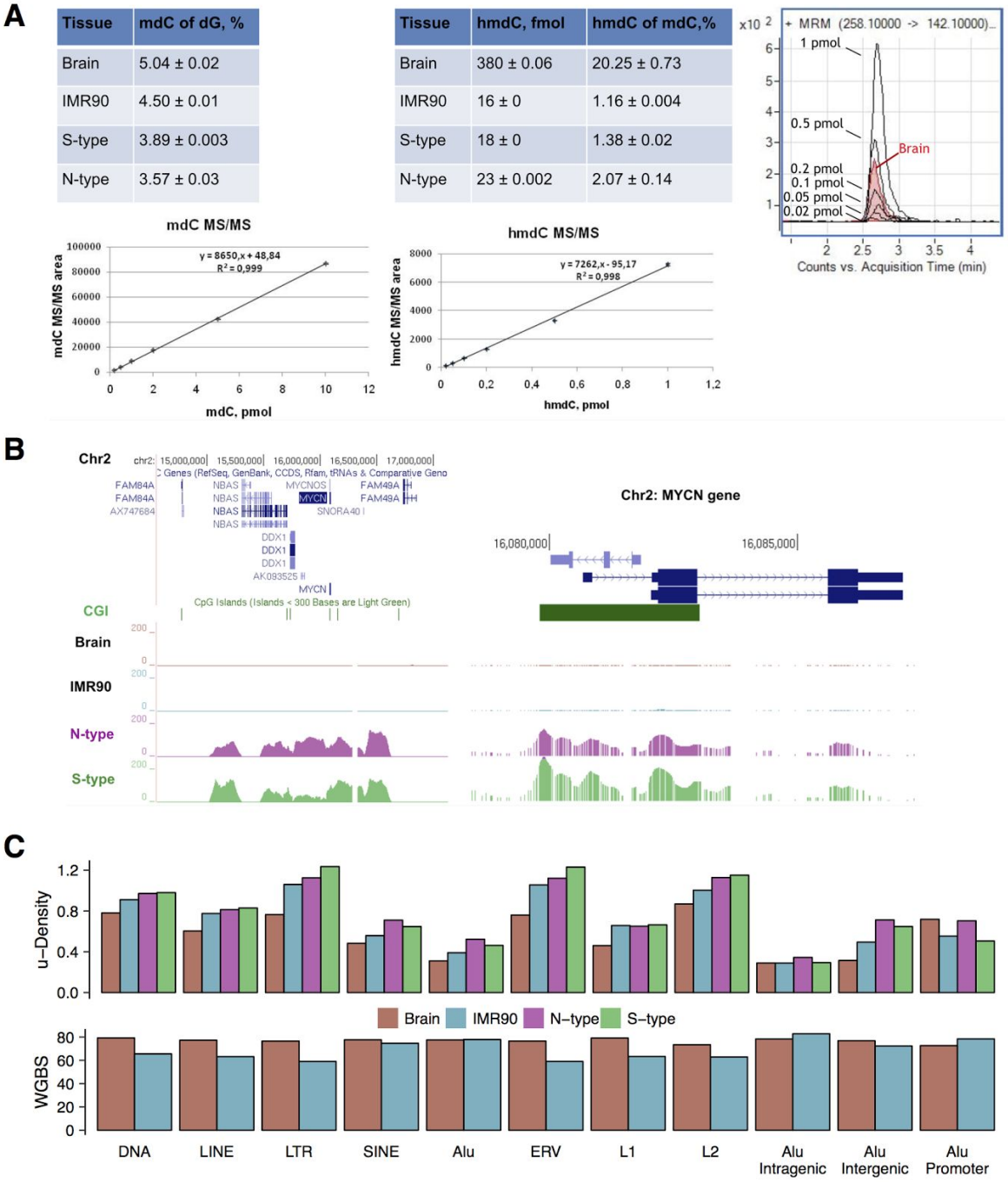

Figure S5. TOP-seq analysis of the neuroblastoma N and S cell types, related to Figure 5. A, HPLC-MS/MS analysis of mdC (Left) and hmdC (Right) in genomic DNA of human tissues. Standard curves in the range of 0.2-10 pmol for mC and 0.02-1 pmol for hmC. MS/MS chromatograms of standard and endogenous 5hmC in the brain cortex sample are shown (Right). Data are represented as mean ±SD. **B,** TOP-seq signal in chr2 shows high u-density around *MYCN* gene region in both N- and S-type cells (chr2:15,026,730-16,640,120; ~1.6 Mb region, covering *NBAS*, *DDX1* and *MYCN* genes). Close-up view of TOP-seq signal around *MYCN* gene is shown (Right). **C,** Distribution of mean TOP-seq u-density (top) and WGBS (bottom) for different repeat classes and families in the neuroblastoma cells, Brain and IMR90. (Left) SINE, LINE, LTR, DNA elements; (Middle) the most abundant families within each class (Alu for SINE, L1/L2 for LINE and ERV for LTR); (Right) promoter, intragenic and intergenic Alu repeats. Distribution analysis of Alu elements that contributed to repeat hypomethylation in the N-cells showed that all Alu elements displayed elevated TOP-seq u-density relative to the other investigated tissues, though the intergenic Alu elements showed the highest hypomethylation among the three Alu groups.

A

CGI type	N/B hyperM	N/B hypoM	S/B hyperM	S/B hypoM	N/S hyperM	N/S hypoM	L/B hyperM	L/B hypoM
Promoter	203 (202)	4285 (4155)	1414 (1407)	2541 (2460)	487 (483)	4156 (4073)	3017 (2988)	538 (523)
Intragenic	827 (567)	5321 (3509)	1153 (940)	4729 (3055)	2752 (1760)	3104 (2456)	2168 (1796)	2790 (1751)
Intergenic	543	2208	593	2037	1042	1469	990	946

B

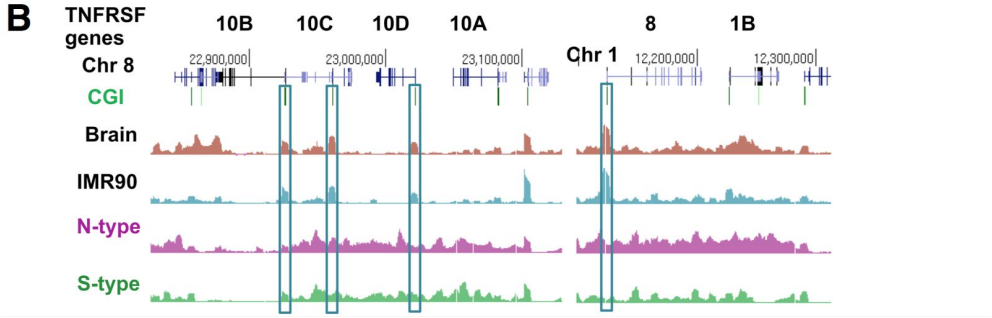

Figure S6. TOP-seq analysis of differential CGI methylation in human tissues, related to Figure 5. A. In pairwise comparisons between cancerous and normal tissues, four types of DMRs were analyzed: N/B- or S/B-specific hypomethylated regions (hypoM, regions that show higher Top-seq u-density in N or S cells than in Brain) and N/B- or S/B-specific hypermethylated regions (hyperM, higher u-density in Brain relative to N or S). Additionally, hypoM regions between the N or S cells were identified. The table shows numbers of hypoM and hyperM CGI-DMRs for N-type, S-type and IMR90 cells (relative to Brain reference) and N/S-hypoM and S/N-hypoM CGI regions (N and S cells compared to each other). Number of genes with assigned tissue-specific CGI-DMRs is shown in parentheses. **B,** Browser view of TOP-seq u-density profiles along two clusters of the tumor necrosis factor (TNF) receptor coding genes. The TNF family proteins are capable of inducing apoptosis, and their expression is largely disturbed in most cancer lines, including NB (Grau et al., 2010). The ~200 kb region in Chr8 covering *TNFRSF10A*, *TNFRSF10B*, *TNFRSF10C* and *TNFRSF10D* genes and the ~86 kb region in Chr1 coding for TNFRSF8, TNFRSF1B receptors are shown. Identified hyperM promoter and intragenic CGIs (relative to Brain and IMR90) are boxed. Strikingly, TOP-seq analysis found strong hypomethylation across the whole TNF-receptor clusters, which was particularly prominent in the N-type cells. This tempts us to speculate that besides the detected CGI methylation across the TNFR-genes, they are subject to an additional layer of regulation, presumably through involvement of the N-specific large hypomethylated domains (Berman et al., 2012)

Supplemental Tables

Table S1. TOP-seq sequencing of the human tissues, related to Figure 3.
Table S2. Regions for pyrosequencing, related to Figure 3.
Table S3. Differential methylation in CGIs, related to Figure 5.
Table S4. Validation of TOP-seq DMRs by pyrosequencing, related to Figure 5.
Table S5. Functional annotation analysis, related to Figure 5.
Table S6. Differential methylation in CGIs overlapping NB prognostic markers, related to Figure 5.

Table S1. TOP-seq sequencing of the human tissues, related to Figure 3. Mapped reads: number of reads aligned to the reference genome and filtered for mapping quality. Unique per CG: number of reads remaining after removal of PCR duplicates and filtered for distance <= 3bp. Called uCGs: number and percentage of identified uCGs from a total of 28 M CGs in human genome. Coverage: averaged TOP-seq read coverage for identified uCGs (first averaged between the chromosomes).

[a] - combined data of 2.5 PI chips.

Chip ID	Chip No	Sample	Raw Reads, M	Mapped Reads, M	Unique per CG	Called uCG, M	Called uCG (%)	Coverage
11	1	Brain-1 R1 (low1)	55.60	31.00	24.70	7.49	26.60	3.27
11	1	Brain-1 R2 (low2)	30.60	17.60	14.40	5.78	20.50	2.48
11	1	Brain-1 R1/R2	86.20				33.50	4.11
9	2	Brain-2 R1 (low1)	49.60	28.80	19.50	5.71	20.20	3.37
9	2	Brain-2 R2 (low2)	45.40	26.60	17.60	5.53	19.60	3.15
9	2	Brain-2 R1/R2	95.00				29.10	4.46
16	3	IMR90-1 R1 (low1)	37.10	18.40	15.50	5.98	21.20	2.68
16	3	IMR90-1 R2 (low2)	44.10	22.60	18.60	6.27	22.20	3.05
16	3	IMR90-1 R1/R2	81.20				31.90	3.93
15	4	IMR90-2 R1 (low1)	45.90	24.10	15.70	5.71	20.20	2.84
20-22-24	#	IMR90-3[a] (high1)	241.00	113.00	92.10	9.98	35.40	8.90
21-23-24	#	IMR90-4[a] (high2)	236.00	131.50	97.10	9.64	34.20	9.72
20-24	#	IMR90-3/4	477.00	244.50	189.20	13.10	46.30	14.10
16, 20-24	#	IMR90-1/3/4 (all)	563.2				47	16
17	#	S-R1	46.70	23.60	17.30	5.13	18.20	3.26
18	#	S-R2	40.70	21.10	15.50	4.93	17.50	3.05
19	#	S-R3	40.50	22.10	15.70	5.06	17.90	3.01
17-18-19		S-R1/2/3	127.90				31.80	5.53
17	#	N-R1	35.90	18.60	14.70	5.86	20.80	2.37
18	#	N-R2	36.70	19.00	15.30	6.18	21.90	2.34
19	#	N-R3	42.10	23.10	17.90	6.67	23.70	2.56
17-18-19		N-R1/2/3	114.70				38.70	4.28

Table S4. Validation of TOP-seq DMRs by pyrosequencing, related to Figure 5. logFC - log fold change. Pyro - pyrosequencing. adjPV - adjusted P value.

		DMRs for S/IMR90				DMRs for N/IMR90			
Region ID	Prognostic NB marker genes	TOP-seq logFC (S/IMR90)	TOP-seq logFC (S/IMR90)	Pyro logFC (S/IMR90)	adjPV (S/IMR90)	TOP-seq logFC (N/IMR90)	TOP-seq logFC (N/IMR90)	Pyro logFC (N/IMR90)	adjPV (N/IMR90)
CGI-10		0.28	0.3	-1.1	0.0001572	-1.6	-1.5	0.4	0.0001572
CGI-12		-4.59	-4.7	1.2	0.0053838	-2.12	-2.3	1.1	0.0053838
CGI-14		-2.07	-2.2	0.9	0.0000118	-0.55	-0.5	-2.6	0.0000118
CGI-20		-0.61	-0.6	0.7	0.0016830	-1.94	-2.1	1	0.0016830
CGI-24		1.06	1.1	-0.5	0.0005622	0.85	0.9	-2.5	0.0005622
CGI-31		0.29	0.3	-0.1	0.0005622	1.51	1.5	-0.7	0.0005622
CGI-32	TNFRSF8	-2.8	-2.7	2.7	0.0000021	-1.38	-1.6	2.5	0.0000021
CGI-33	RASSF1	-4.49	-3.3	2	0.0015126	-4.29	-3.2	1.9	0.0015126
CGI-34	ZMYND10	-3.09	-3.2	3.9	0.0000021	-3.1	-3	3.8	0.0000021
CGI-6		-0.65	-0.7	0.5	0.0006452	-0.92	-1	-0.1	0.0006452
Regions that failed in pyrosequencing validation are highlighted.									

Table S6. Differential methylation in CGIs overlapping NB prognostic markers, related to Figure 5. Log fold change in TOP-seq u-density is shown for N/S, N/B, S/B and IMR90/B pairs. Positive fold change in TOP-seq u-density signal means lower methylation in the first member of the respective pair and *vice versa*.

gene	CGI type	chr	start	end	N/S	N/B	S/B	IMR90/B
ACSS3	promoter	chr12	81330608	81331514	0	0	0	0
CCND1	promoter	chr11	69451136	69458596	-0.27	-0.96	-0.69	0
CCND2	promoter	chr12	4378366	4382222	-0.29	-0.65	-0.36	-0.21
GNAS	promoter	chr20	57414595	57414896	0	0	0	0
HIST1H3C	promoter	chr6	26044203	26044469	0.79	-3.67	-4.46	0
HOXD3	promoter	chr2	177001221	177001783	-0.5	-1.29	-0.78	0
JAK2	promoter	chr9	4984543	4985630	0	0	0	0
KRT19	promoter	chr17	39683908	39684599	0.74	-3.96	-4.7	0.49
PCDHB16	promoter	chr5	140558931	140559789	0	-2.02	-2.15	-0.95
POU2F2	promoter	chr19	42700686	42700975	0	0	0	0
RASSF1	promoter	chr3	50377803	50378540	0.2	-4.11	-4.3	0.18
RB1	promoter	chr13	48877459	48878501	0	-0.66	-0.58	0
ACSS3	intragenic	chr12	81471569	81472119	0.34	-3.44	-3.78	0
CCND1	intragenic	chr11	69461060	69461378	1.62	-2.15	-3.77	-1.05
CCND1	intragenic	chr11	69468810	69469152	-0.57	-2.09	-1.52	-1.42
CCND2	intragenic	chr12	4383193	4384405	0	-0.74	-0.62	0
FGF6	intragenic	chr12	4554405	4554624	0	1.37	1.34	0
GNAS	intragenic	chr20	57415135	57417153	-0.37	-0.98	-0.6	0.19
GNAS	intragenic	chr20	57426729	57427047	-0.26	-2.06	-1.8	-0.44
GNAS	intragenic	chr20	57427691	57427995	0	0	0	0
GNAS	intragenic	chr20	57428308	57428516	0	0	0	0
GNAS	intragenic	chr20	57429024	57431239	-0.44	-0.43	0	0
GNAS	intragenic	chr20	57463652	57467739	0	0	0	0
HOXD3	intragenic	chr2	177004452	177004693	2.65	-1.92	-4.57	-0.2
HOXD3	intragenic	chr2	177005364	177005666	-0.54	-2.27	-1.73	-0.21
HOXD3	intragenic	chr2	177012371	177012675	15.7	-4.5	-20.2	-0.19
HOXD3	intragenic	chr2	177016416	177016632	2.24	-4.83	-7.07	-0.46
HOXD3	intragenic	chr2	177017266	177017489	36.34	-4.19	-40.53	-0.34
HOXD3	intragenic	chr2	177024501	177025692	1.19	-5.15	-6.34	-0.22
HOXD3	intragenic	chr2	177027617	177028014	35.59	-5.17	-40.76	0
HOXD3	intragenic	chr2	177029413	177029941	0	-5.31	-5.29	-0.28
HOXD3	intragenic	chr2	177036254	177037213	0	-6.61	-6.6	0
JAK2	intragenic	chr9	5040734	5041922	0	0	0	0
JAK2	intragenic	chr9	5042797	5043270	-2.07	-3.87	-1.8	-0.46
JAK2	intragenic	chr9	5110784	5112274	0	0	0	0
PCDHB10	intragenic	chr5	140573423	140574316	1.68	-1.15	-2.83	0
PCDHB11	intragenic	chr5	140580723	140581625	0.63	-1.48	-2.11	-0.36
PCDHB12	intragenic	chr5	140589744	140590657	1.03	-1.3	-2.33	-0.41
PCDHB13	intragenic	chr5	140595008	140595880	0.61	-1.95	-2.56	-0.47
PCDHB14	intragenic	chr5	140604453	140605304	0.66	-2.19	-2.85	0
PCDHB15	intragenic	chr5	140626444	140627373	1.36	-1.87	-3.23	-0.35
PCDHB16	intragenic	chr5	140563510	140564361	0	-1.47	-1.54	-0.6
PCDHB17	intragenic	chr5	140536896	140537766	0.4	-1.1	-1.5	-0.93
PCDHB2	intragenic	chr5	140475599	140476607	0	-2.3	-2.22	-0.4
PCDHB3	intragenic	chr5	140481546	140482460	0.28	-2.59	-2.87	-0.63
PCDHB4	intragenic	chr5	140502890	140503804	-35.23	-39.23	-4	-0.68
PCDHB5	intragenic	chr5	140516311	140517300	-1.51	-4	-2.49	-0.59
PCDHB6	intragenic	chr5	140531157	140532017	-0.35	-1.99	-1.65	0
PCDHB7	intragenic	chr5	140553681	140554637	0.61	-1.88	-2.49	-0.99
POU2F2	intragenic	chr19	42599507	42600522	-2.49	-3.51	-1.02	-0.52
POU2F2	intragenic	chr19	42637277	42637684	0	0	0	0
POU2F2	intragenic	chr19	42681304	42681513	0	0	0	0
RASSF1	intragenic	chr3	50374264	50375629	-0.33	0	0.24	0.42
RB1	intragenic	chr13	48890957	48891549	15.48	-2.79	-18.26	-0.4
RB1	intragenic	chr13	48892635	48893857	-0.2	-4.76	-4.56	-0.91
ZMYND10	intragenic	chr3	50382889	50383310	0	-3.19	-3.19	0