Supplementary Materials
for
Genome-scale DNA methylation maps of pluripotent and differentiated cells

Alexander Meissner (1,2,3)*, Tarjei S. Mikkelsen (2,4)*, Hongcang Gu (2), Marius Wernig (1), Andrey Sivachenko (2), Xiaolan Zhang (2), Bradley E. Bernstein (2,5,6), Chad Nusbaum (2), David B. Jaffe (2), Andreas Gnirke (2), Rudolf Jaenisch (1,7) and Eric S. Lander (1,2,7,8)

(1) Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge MA 02142
(2) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142
(3) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138
(4) Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139
(5) Molecular Pathology Unit and Center for Cancer Research, MGH, Charlestown, Massachusetts 02129
(6) Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115
(7) Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139
(8) Department of Systems Biology, Harvard Medical School, Boston MA 02114

Contents:

All data analyzed in this study can be obtained from
http://www.broad.mit.edu/seq_platform/methylation/

Supplementary Table S1 – RRBS coverage as a function of size selection

RR digest: Mouse (mm8)			Coverage		CpG islands		Enrichment[b]		
Enzyme	**Size range (bp)**	**Fragments**	**Sequence (Mb)[a]**	**CpGs**	**All**	**≥10 CpGs**	**CpG islands**	**TSS**	**Exons+ Introns**
MspI	40-120	186,429	13.4	853,075	13,105	12,303	63.1	7.8	1.4
MspI	100-220	185,349	13.3	700,518	12,152	10,492	31.0	5.1	1.3
MspI	220-400	144,683	10.4	472,895	7,840	3,783	11.7	3.6	1.4
MspI	40-220	333,104	24.0	1,383,382	14,353	13,633	47.5	6.5	1.3
MspI	40-400	476,883	34.3	1,853,073	15,015	14,200	36.7	5.6	1.3

The RRBS strategy can be applied to any mammalian genome. Due to the higher CpG and CpG island content of the human genome, the same size fractions will result in approximately twice as many fragments:

RR digest: Human (hg18)			Coverage		CpG islands		Enrichment[b]		
Enzyme	**Size range (bp)**	**Fragments**	**Sequence (Mb)[a]**	**CpGs**	**All**	**≥10 CpGs**	**CpG islands**	**TSS**	**Exons+ Introns**
MspI	40-120	369,554	23.4	1,808,076	22,434	21,069	41.7	6.9	1.5
MspI	100-220	337,756	24.3	1,463,283	21,064	18,206	19.2	5.3	1.4
MspI	220-400	232,189	16.7	843,688	14,415	7,542	8.6	3.8	1.4
MspI	40-220	647,902	43.5	2,985,666	24,633	23,303	30.0	6.9	1.5
MspI	40-400	878,491	60.1	3,823,195	25,783	24,336	24.1	6.1	1.4

[a] Total unique and repetitive sequence covered, assuming 36 bp end reads
[b] Relative to complete genome sequence

Supplementary Table S2 – RRBS libraries sequenced in this study

RRBS Library source	Total reads (number of lanes of Illumina sequencing)	Aligned reads [a]	Analyzed (high-quality) reads [b]	Distinct CpGs	Median cov. (x)	Median CpG meth. level (%)
Astrocytes (in vitro, P18)	22,792,761 (7)	9,304,473	9,037,586	951,422	7	70
Astrocytes (primary, P11)	25,931,105 (4)	10,080,532	9,638,968	928,227	10	42
Astrocytes (primary, P2)	27,578,312 (4)	10,452,548	9,783,816	919,407	10	25
B cells	15,742,954 (4)	6,623,398	6,416,120	894,879	7	17
Brain	32,871,302 (4)	12,007,722	11,472,495	906,010	14	10
CD4+ T cells	17,425,940 (4)	7,491,354	7,312,532	874,811	9	11
CD8+ T cells	12,714,727 (3)	5,673,776	5,540,188	821,388	6	10
ES cells	31,624,616 (10)	13,620,414	13,298,707	950,671	12	14
ES cells (Dnmt deficient)	23,899,182 (7)	8,276,492	8,062,719	908,483	8	0
Liver	20,644,156 (4)	8,248,332	7,983,808	668,614	8	9
Lung	17,469,597 (4)	9,085,895	9,017,768	796,645	6	8
Embryonic fibroblasts	21,326,932 (7)	9,507,186	9,289,500	903,921	8	23
NPC (P18)	20,086,908 (7)	9,388,129	9,118,163	921,136	9	40
NPC (P9)	28,748,784 (4)	11,496,175	11,150,501	912,408	9	55
Sox1+	31,030,042 (8)	11,621,399	11,314,731	972,024	11	29
Sox1+-derived NPCs	40,422,791 (6)	13,291,166	12,872,974	996,991	11	35
Spleen	18,630,637 (4)	9,985,486	9,887,195	799,684	7	8
Tail-tip fibroblasts	23,202,538 (7)	10,765,328	10,571,236	948,249	9	11

[a] Number of total sequenced reads that could be aligned to the reduced representation reference sequence such that the second best alignment has at least >= 2 mismatches more than the best alignment. Reads may not align due to low quality, repetitive content or because they do not correspond to *in silico* predicted MspI fragments in the 40-220 bp size range. Alignment efficiency increased from ~30% (e.g ES cells) in early runs up to ~50-60% after instrument upgrades (i.e. Lung, Spleen). This is comparable to ChIP-Seq (see Table S3), indicating that bisulfite treatment does not result in a substantial loss of alignment efficiency. At present, 4 lanes of Illumina sequencing (2 per size fraction) are generally sufficient for complete coverage of a single RRBS library.

[b] Number of aligned reads that begin with a CG or TG dinucleotide and for which $\sum_{q \in Q} 10^{q/10} > 1000$, where Q denotes the read quality scores at each mismatched position.

Supplementary Table S3 – ChIP-Seq libraries sequenced in this study

ChIP-Seq Library source	Epitope	Total sequences (number of lanes of	Aligned reads [a]
ES cells	H3K4me2	22 million (3)	4.1 million
ES cells	H3K4me1	14 million (2)	6.0 million
Neural Progenitor Cells	H3K4me2	27 million (4)	9.4 million
Neural Progenitor Cells	H3K4me1	23 million (4)	6.7 million
Whole brain tissue	H3K4me3	22 million (2)	9.2 million
Whole brain tissue	H3K4me2	27 million (2)	9.9 million
Whole brain tissue	H3K27me3	20 million (2)	9.7 million

[a] Number of total sequenced reads that could be aligned to the reference genome such that the second best alignment has at least >= 2 mismatches more than the best alignment, and the total number of mismatches is <= 6. A significant portion of unaligned reads are due to low quality sequences generated by process artifacts, rather than repetitive sequences.

Supplementary Table S4 – GO Categories enriched for HCPs with > 75% mean methylation in ES-derived astrocytes

GO category	Description	p-value[a]	Genes associated with methylated HCPs[b]
GO:0007126	Meiosis	6.19E-05	Msh4,Sycp3,Sycp2,Spo11,Syce2,Sycp1,Dmc1
GO:0007155	Cell adhesion	0.000141	Dsc2,Cd97,Dsg2,Nlgn2,Cpxm2,Gp1bb,Lamc2, Scarf2,Pkp1,Pgm5,Ctgf,Col9a3,Parvb,Aebp1,Itga4, Col2a1,Cldn11
GO:0007165	Signal transduction	0.001517	Cd97,Gpr176,Cspg4,F2rl1,Lep,Gpr64,Ptger2,Plcd1, Sstr1,Oxtr,Tnfrsf25,Fgfr4,Galr2,Fgf20,Sstr4,Gpr83, Irak3,Fgf17,Prokr1,Prlhr,Stat5a,Htr1f,Gna14,Tacr3, Tnfrsf10b,Htr6
GO:0007283	Spermatogenesis	0.002592	Taf7l,Sycp3,Spag6,Dazl,Dmc1,D1Pas1,Msh4, Cldn11,Spag16
GO:0009190	Cyclic nucleotide biosynthetic process	0.005906	Gucy2e,Npr1,Adcy7
GO:0006811	Ion transport	0.006922	Scnn1b,Grin2a,Clic6,Grik2,Atp2a3,Kcnj10, Slc34a2,Trpm6,Kcng1,Slc13a3,Kcna6,Bspry, Slc5a5,Pllp,Kcnb1,Tmem37,Mcoln2
GO:0001541	Ovarian follicle development	0.007887	Msh4,Dmc1,Spo11
GO:0006508	Proteolysis	0.008193	Mmp2,Ccdc79,Mmp23,A530088I07Rik,Agbl2, Mmp14,Casp8,Wdr31,Pgm5,Aebp1,Dhh,Adamts5, Npepl1
GO:0007275	Multicellular organismal development	0.00848	Taf7l,Cspg4,Myod1,Dkk3,Nnat,Hoxd12,Sema4b, Nodal,Lect1,Pgf,Dazl,Dhh,Amn,Dll3,Tnfaip2, Ddx4,Slit1,Nsd1,Hhat,Cdx1,Nkx3.1,Otx2,Rax, Ebf2,Heyl,Efnb3,Scx,D1Pas1,Shroom3
GO:0007218	Neuropeptide signaling pathway	0.011283	Npb,Sstr1,Gal,Gpr64,Cd97
GO:0001525	Angiogenesis	0.014419	Ctgf,Casp8,Tnfaip2,Pgf,Cspg4,Plcd1
GO:0007517	Muscle development	0.031826	Des,Ky,Myod1
GO:0016477	Cell migration	0.03212	Ctgf,Mmp14,Nodal,Itga4
GO:0006836	Neurotransmitter transport	0.03672	Slc18a2,Slc6a2,Slc6a11

[a] Nominal p-value of set enrichment based on Fisher's exact test (two-tailed)
[b] Based on GO annotations obtained form http://geneontoloy.org

Supplementary Table S5 – GO Categories enriched for HCPs with > 50% mean methylation in ES-derived astrocytes

GO category	Description	p-value [a]	Genes associated with methylated HCPs[b]
GO:0007165	Signal transduction	7.01E-06	Gpr37,Edaradd,Cd97,Grm8,Fgf15,Gpr176,Gpr101, Cspg4,F2rl1,Wnt3,Lep,Gpr64,Edg3,Ptger2,Plcd1, Gpr12,Sstr1,Oxtr,Grb10,Tnfrsf25,Gm266,Ltb4r2, Fgfr4,Galr2,Drd5,Gpr156,Hif3a,Wnt10b,Fgf20, Fgf16,Sstr4,Gpr26,Gpr83,Pacsin1,Irak3,Htr2c, Adora2a,Pde8a,Fgf17,Gnas,Ntsr1,Prokr1,Prlhr, Stat5a,Htr1f,Gpr150,Gna14,Tacr3,Tnfrsf10b,Htr6, Wnt2
GO:0007155	Cell adhesion	2.82E-05	Dsc2,Nlgn1,Cd97,Itgb4,Dsg2,Nlgn2,Cpxm2, Gp1bb,Lamc2,Pcdhac1,Scarf2,Sdk2,Pkp1,Pgm5, Ctgf,Col9a3,Parvb,Col18a1,Aebp1,Perp,Col12a1, 8430419L09Rik,F11r,Emilin2,Thbs4,Itga4,Col2a1, Cldn11
GO:0007275	Multicellular organismal development	2.82E-05	Ndrg2,Edaradd,Taf7l,Vamp5,Itgb4,Cspg4,Myod1, Lmx1b,Dkk3,Wnt3,Churc1,Nnat,Hoxd12,Bmp3, Sema4b,Snai1,Boll,Bmp8b,Nodal,Hic1,Hoxd9, Lect1,Pgf,Dazl,Dhh,Amn,Phox2a,Pitx2,T,Dll3, Gldn,Tnfaip2,Ddx4,Dlx4,Slit1,Wnt10b,Msx3,Nsd1, Mesp2,Htatip2,Hhat,Cdx1,Hoxd10,Hoxa11, 4930506M07Rik,Nkx3.1,Otx2,Rax,Ebf2,Heyl, Sfrp2,Nkx1.2,Efnb3,Scx,Hoxd1,D1Pas1,Olig2, Nav1,Shroom3,Wnt2
GO:0006817	Phosphate transport	6.48E-05	Scara3,Col2a1,Emid1,Col25a1,Col18a1,Col12a1, Slc34a2,Gldn,Emid2
GO:0007126	Meiosis	0.00013	Msh4,Sycp3,Sycp2,Spo11,Syce2,Smc1b,Sycp1, Dmc1,Boll
GO:0006811	Ion transport	0.00115	Scnn1b,Kcnf1,Grin2a,Clic6,Grik2,Atp2a3,P2rx5, Kcnj10,Slc34a2,Kcnk13,Trpm6,Kcng1,Kcnc4, Slc13a3,Slc39a8,Kcna6,Bspry,Slc5a5,Pkdrej,Pllp, P2rx2,Kcnb1,Kcnk4,Grin2c,Chrna3,Slc22a4, Grin3b,Gabrb1,Tmem37,Slc30a10,Mcoln2
GO:0030199	Collagen fibril organization	0.01084	Col2a1,Tnxb,Lox,Lmx1b
GO:0001525	Angiogenesis	0.01566	Ctgf,Col18a1,Casp8,Tnfaip2,Sox18,Htatip2,Pgf, Cspg4,Plcd1
GO:0051216	Cartilage development	0.01936	Bmp3,Gnas,Bmp8b,Lect1
GO:0007268	Synaptic transmission	0.02197	Chrna3,Grin2a,P2rx2,Nrxn2,Grik2,Grm8
GO:0001501	Skeletal development	0.02207	Rai1,Dll3,Hoxd10,Gnas,Hoxa11,Hoxd12,Pthlh

GO:0006629	Lipid metabolic process	0.02731	Fads3,Srebf1,Pcsk9,Acot12,Mlstd1,Slc27a3,Tnxb, Lep,Acot6,Slc27a2,Plcd1
GO:0006508	Proteolysis	0.02961	Ctsf,Mmp2,Gpr26,Dnahc11,Ccdc79,Mmp23, A530088I07Rik,Agbl2,Mmp14,Casp8,Wdr31, Pgm5,Aebp1,Pcsk9,Dhh,St14,Adamts5,Ctsh,Npepl1
GO:0007283	Spermatogenesis	0.04606	Boll,Bmp8b,Taf7l,Sycp3,Spag6,Dazl,Dmc1, D1Pas1,Msh4,Cldn11,Spag16
GO:0007517	Muscle development	0.04622	Des,Ky,Tagln2,Myod1
GO:0001541	Ovarian follicle development	0.04828	Msh4,Dmc1,Spo11

[a] Nominal p-value of set enrichment based on Fisher's exact test (two-tailed)
[b] Based on GO annotations obtained form http://geneontoloy.org

Supplementary Table S6 – GO Categories depleted for HCPs with > 75% mean methylation in ES-derived astrocytes

GO category	Description	p-value [a]	Genes associated with methylated HCPs[b]
GO:0006512	Ubiquitin cycle	0.000243	Parc
GO:0015031	Protein transport	0.001639	Lin7b,Rasef
GO:0006412	Translation	0.010517	Eef1a2
GO:0006886	Intracellular protein transport	0.016169	
GO:0008380	RNA splicing	0.023412	
GO:0006397	mRNA processing	0.031508	Papolb

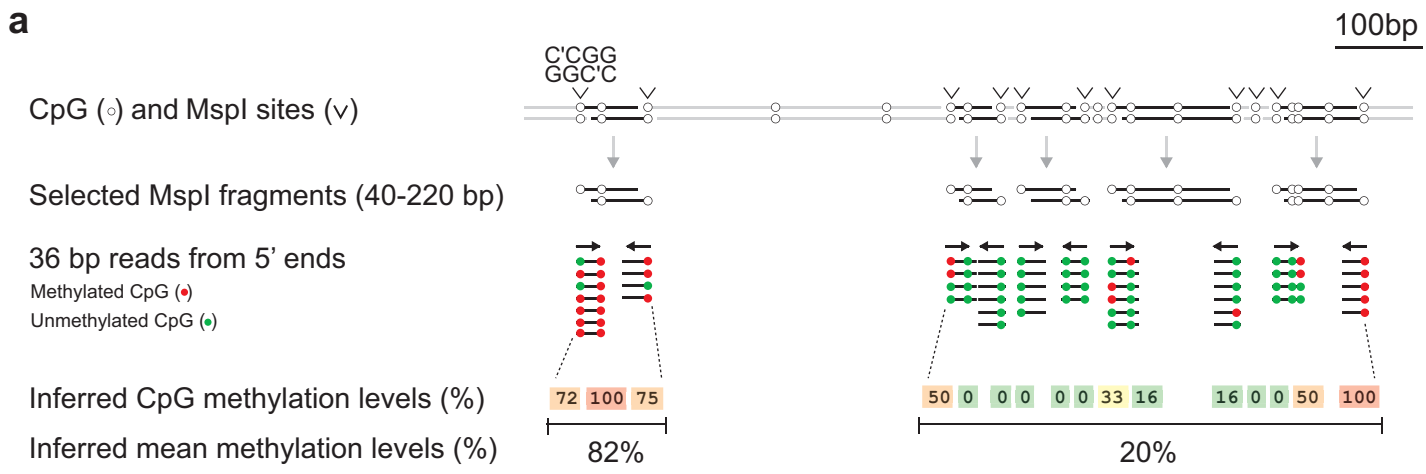
[a] Nominal p-value of set enrichment based on Fisher's exact test (two-tailed)
[b] Based on GO annotations obtained form http://geneontoloy.org

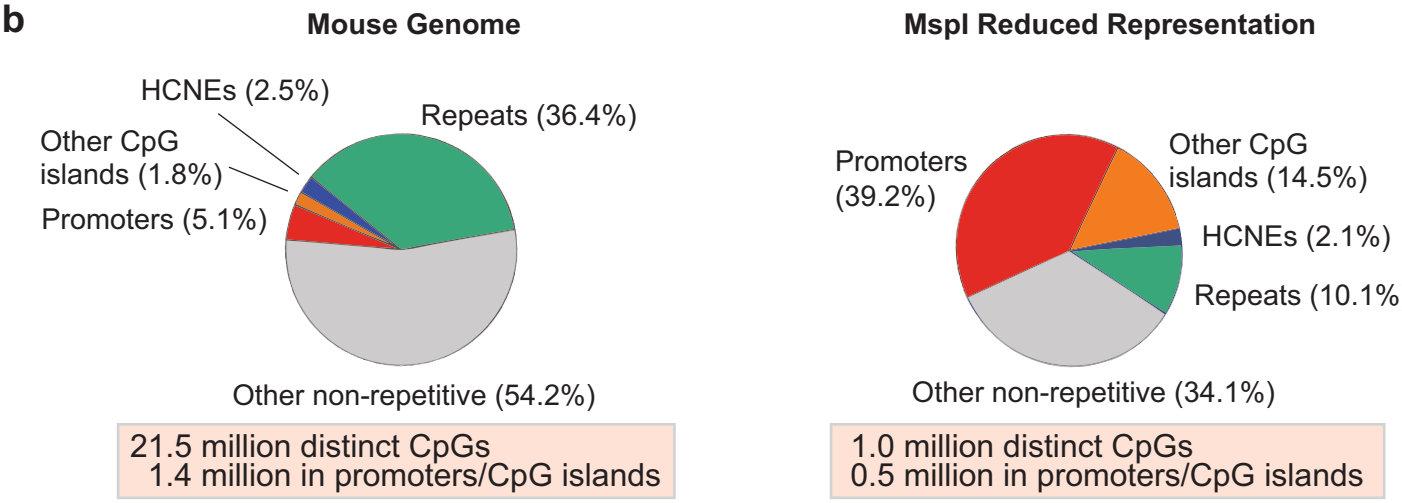
Supplementary Table S7 – GO Categories depleted for HCPs with > 50% mean methylation in ES-derived astrocytes

GO category	Description	p-value [a]	Genes associated with methylated HCPs[b]
GO:0006512	ubiquitin cycle	7.33E-08	Fbxo17,Parc
GO:0015031	protein transport	9.07E-06	Rab3b,Pitpnm1,Lin7b,Sec31b,Rasef
GO:0006412	translation	0.0001	Rps20,Eef1a2
GO:0006397	mRNA processing	0.00014	Papolb
GO:0008380	RNA splicing	0.00023	
GO:0006886	intracellular protein transport	0.001	Rab3b
GO:0006281	DNA repair	0.00444	Mpg
GO:0006974	response to DNA damage stimulus	0.00448	Mpg
GO:0006457	protein folding	0.0091	
GO:0042254	ribosome biogenesis and assembly	0.0141	
GO:0016568	chromatin modification	0.02006	Nsd1
GO:0051726	regulation of cell cycle	0.02915	Cdkl1
GO:0006511	ubiquitin-dependent protein catabolic process	0.03008	Parc
GO:0051301	cell division	0.03397	Sycp3,Sycp1,Syce2,Sycp2

[a] Nominal p-value of set enrichment based on Fisher's exact test (two-tailed)
[b] Based on GO annotations obtained form http://geneontoloy.org

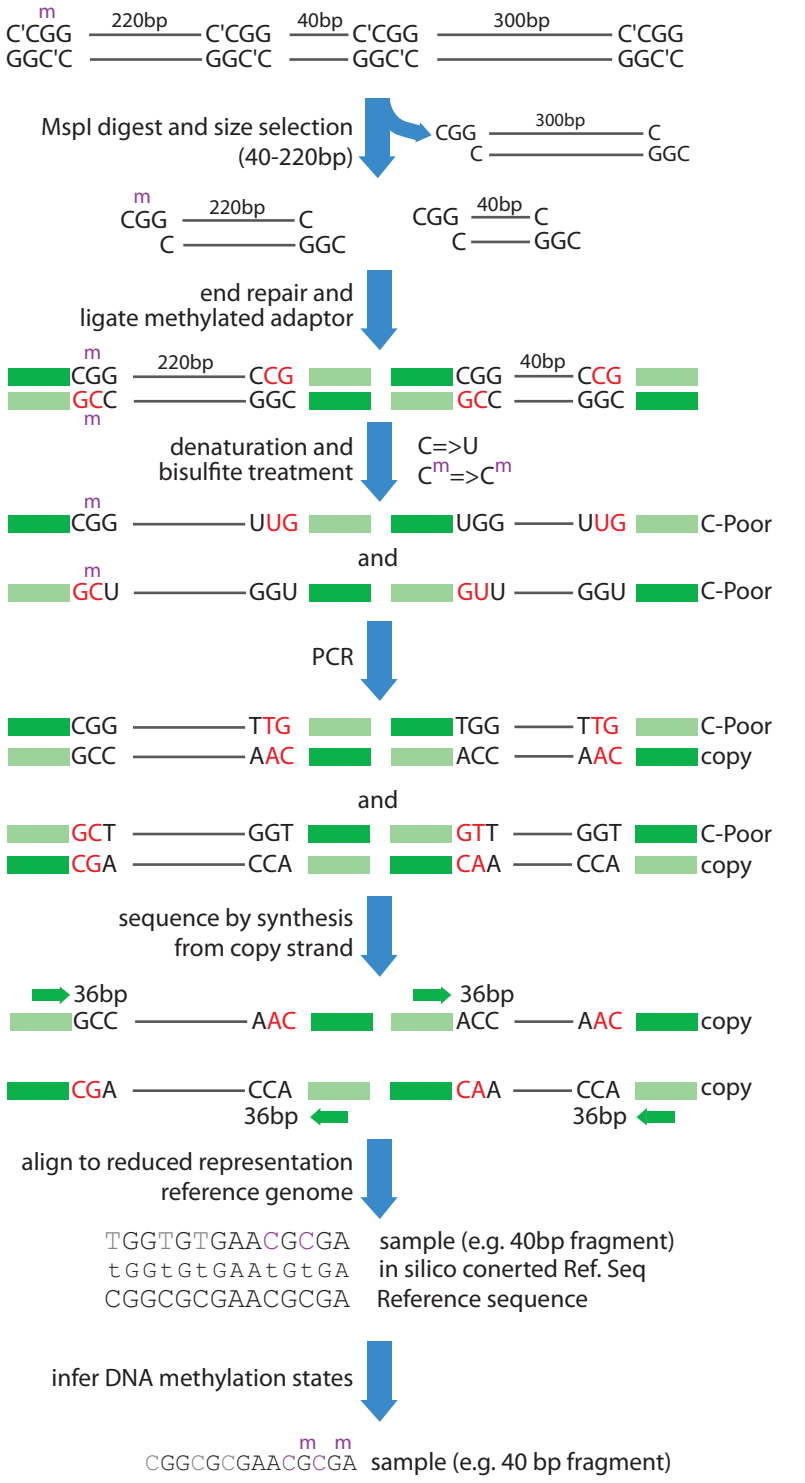
Supplementary Figure S1





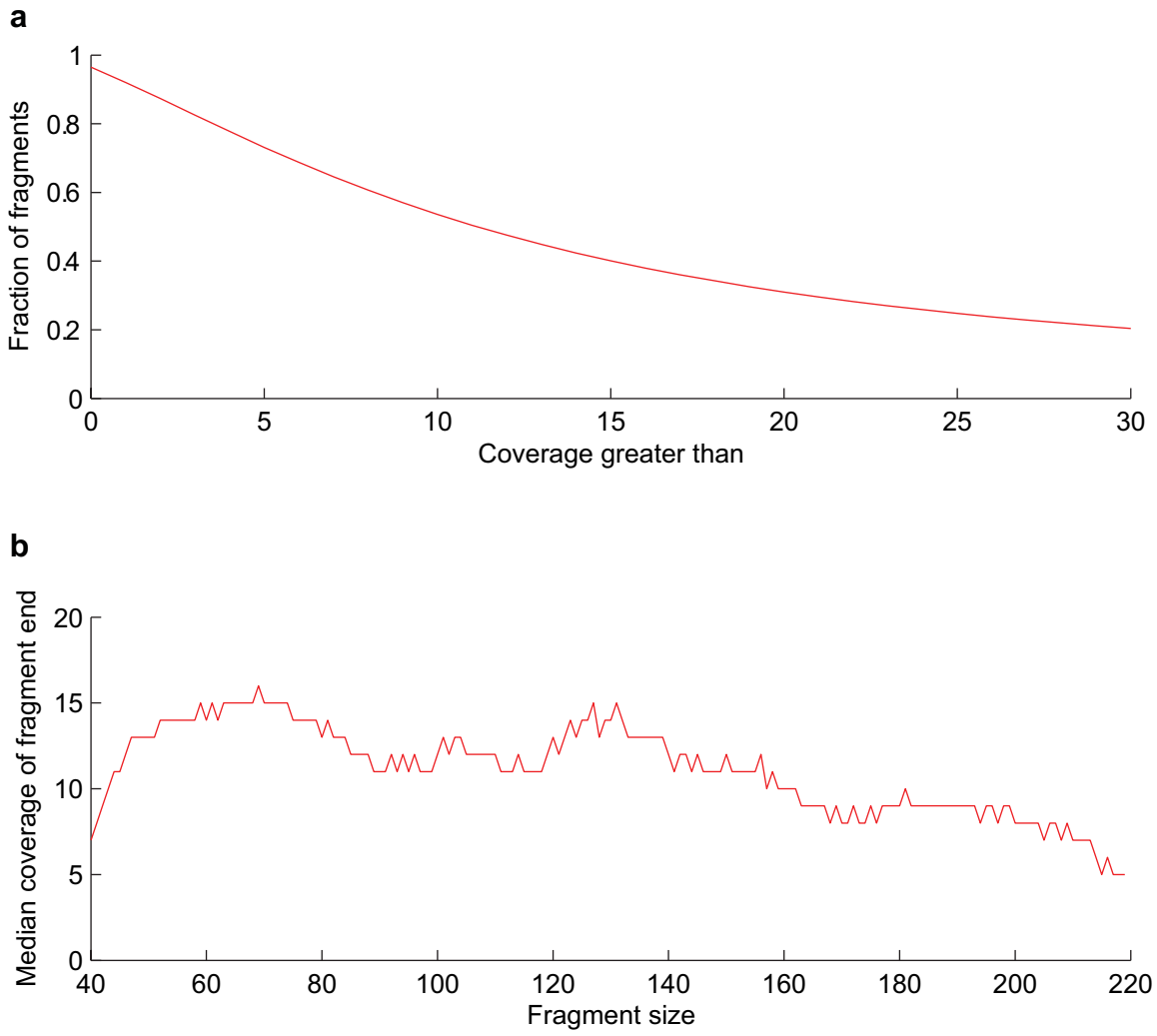
Reduced Representation Bisulfite Sequencing. **a**, Schematic overview of the RRBS approach. Genomic DNA is digested with methylation-insensitive MspI. Fragments between 40-220 bp are selected, treated with sodium bisulfite and 5' end-sequenced (see Supplementary Figure S2 for more details). CpGs are represented as open circles and MspI cut sites are indicated above (v). Filled circles represent either unmethylated (green) or methylated (red) CpGs at each sampled molecule. The methylation level of each CpG is inferred from the number of unconverted sites in reads overlapping that site. The inferred methylation level is shown below each CpG site. The color of the box ranges from green (<20% methylation) to red (>80% methylation). **b**, The MspI-based reduced representation fraction contains ~4.8% of all CpGs in the mouse genome, but is significantly enriched for HCPs and other CpG-rich sequence features.

Supplementary Figure S2



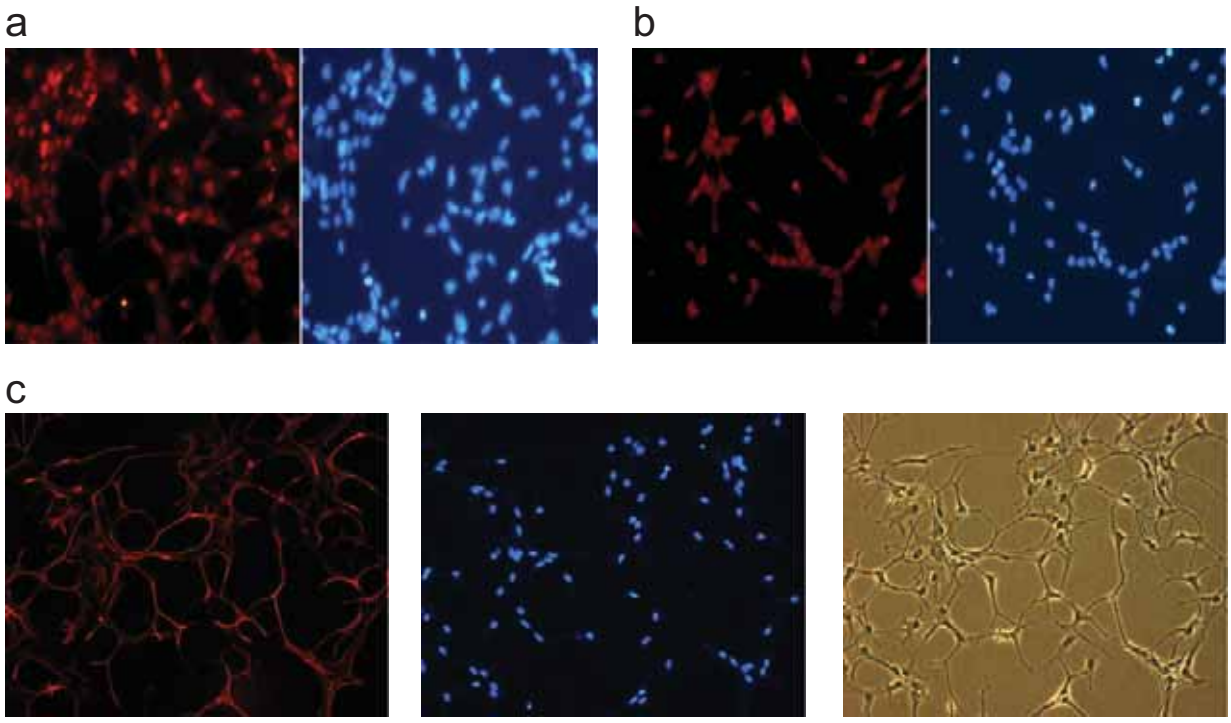
Overview of the RRBS process. Genomic DNA is digested with MspI, size selected, end-repaired and fitted with methylated Illumina/Solexa adapters prior to sodium bisulfite treatment and PCR enrichment. Sequenced reads are aligned to a reference genome digest to infer methylation levels.

Supplementary Figure S3



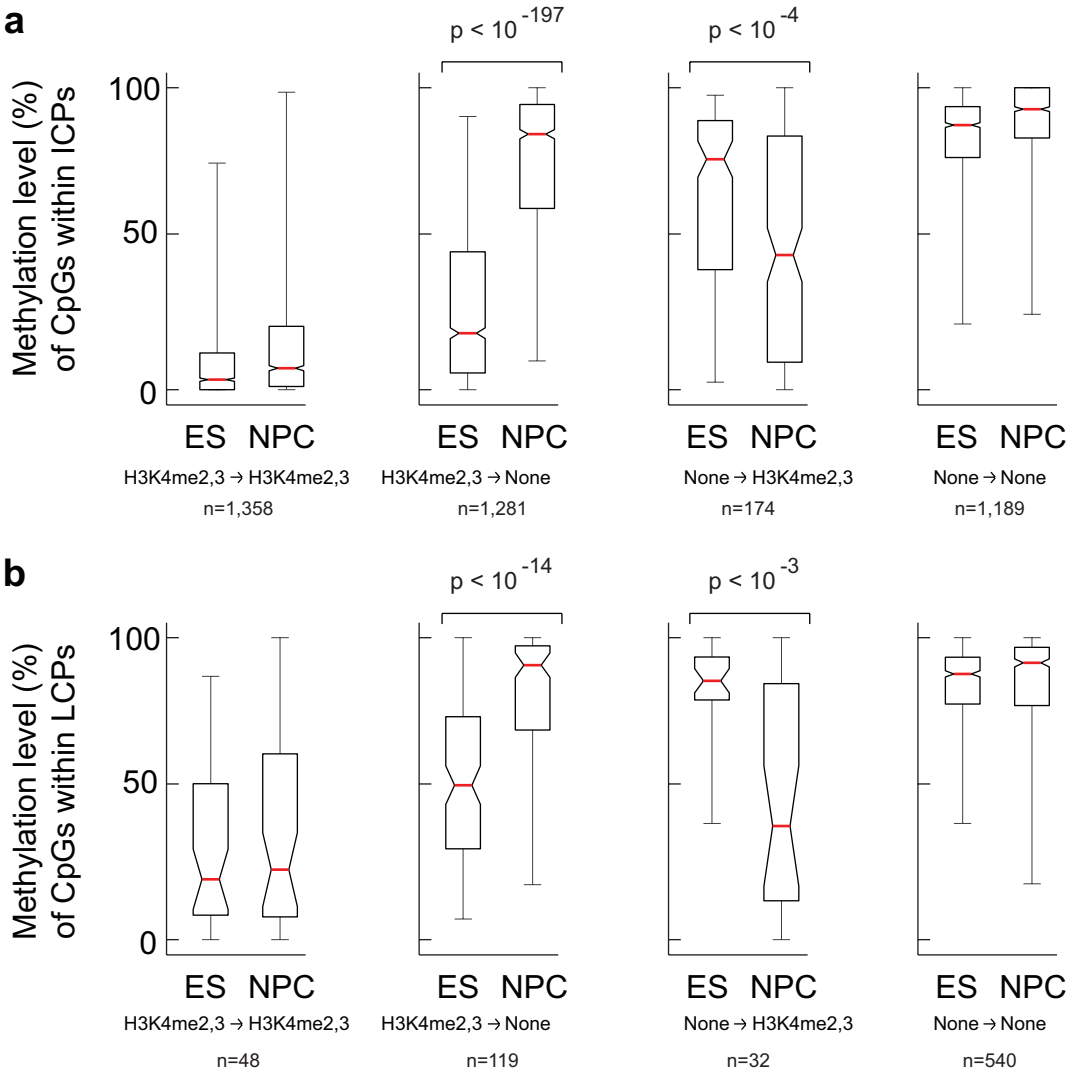
RRBS Library representation from ES cells. **a**, The majority (97%) of non-repetitive MspI fragment ends were observed at least once among 13 million aligned reads, and the median coverage was 12X. **b**, Median coverage was relatively similar for fragments of different lengths.

Supplementary Figure S4



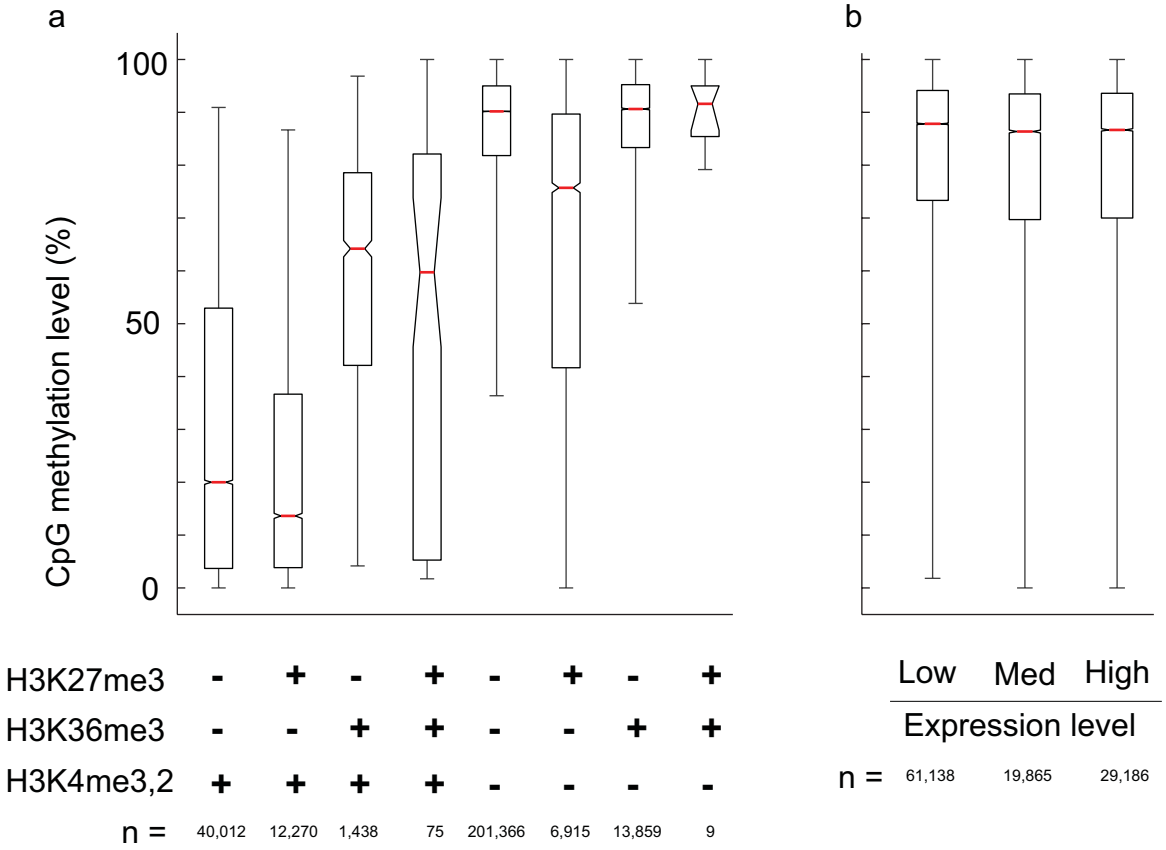
Immunohistochemistry of ES-derived neural progenitor cells (NPCs) and NPC-derived astrocytes. **a**, Sox2 (left) and DAPI (right) staining of NPCs. **b**, Brn2 (left) and DAPI (right) staining of NPCs. **c**, GFAP (left), DAPI (middle), brightfield (right) of astrocytes.

Supplementary Figure S5



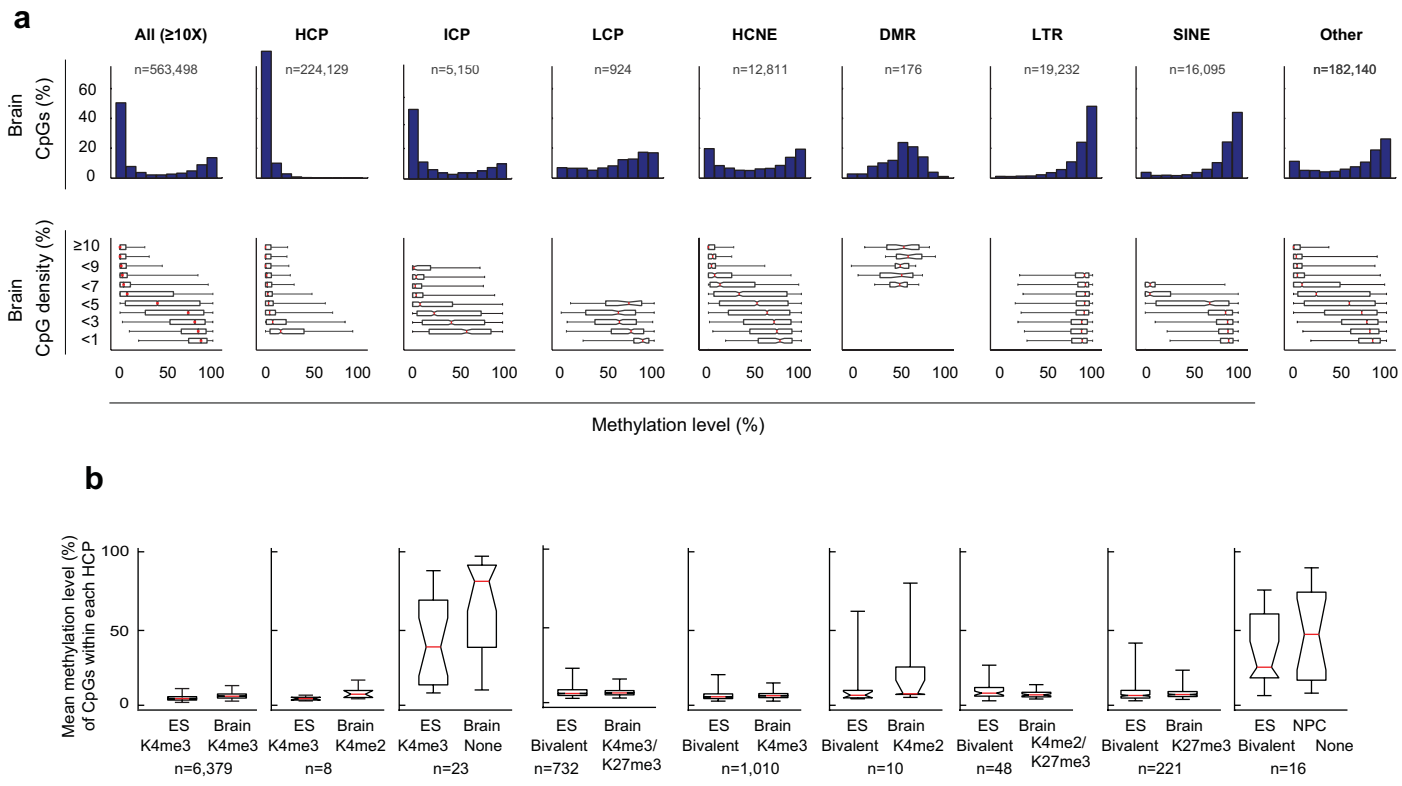
Distribution of CpG methylation levels for (**a)** intermediate CpG-density promoters (ICPs) and (**b**) low CpG-density promoters (LCPs), conditional on histone methylation states in ES cells and neural progenitor cells (NPCs). Changes in H3K4 methylation are significant correlated with inverse changes in DNA methylation levels (Mann-Whitney's U test).

Supplementary Figure S6



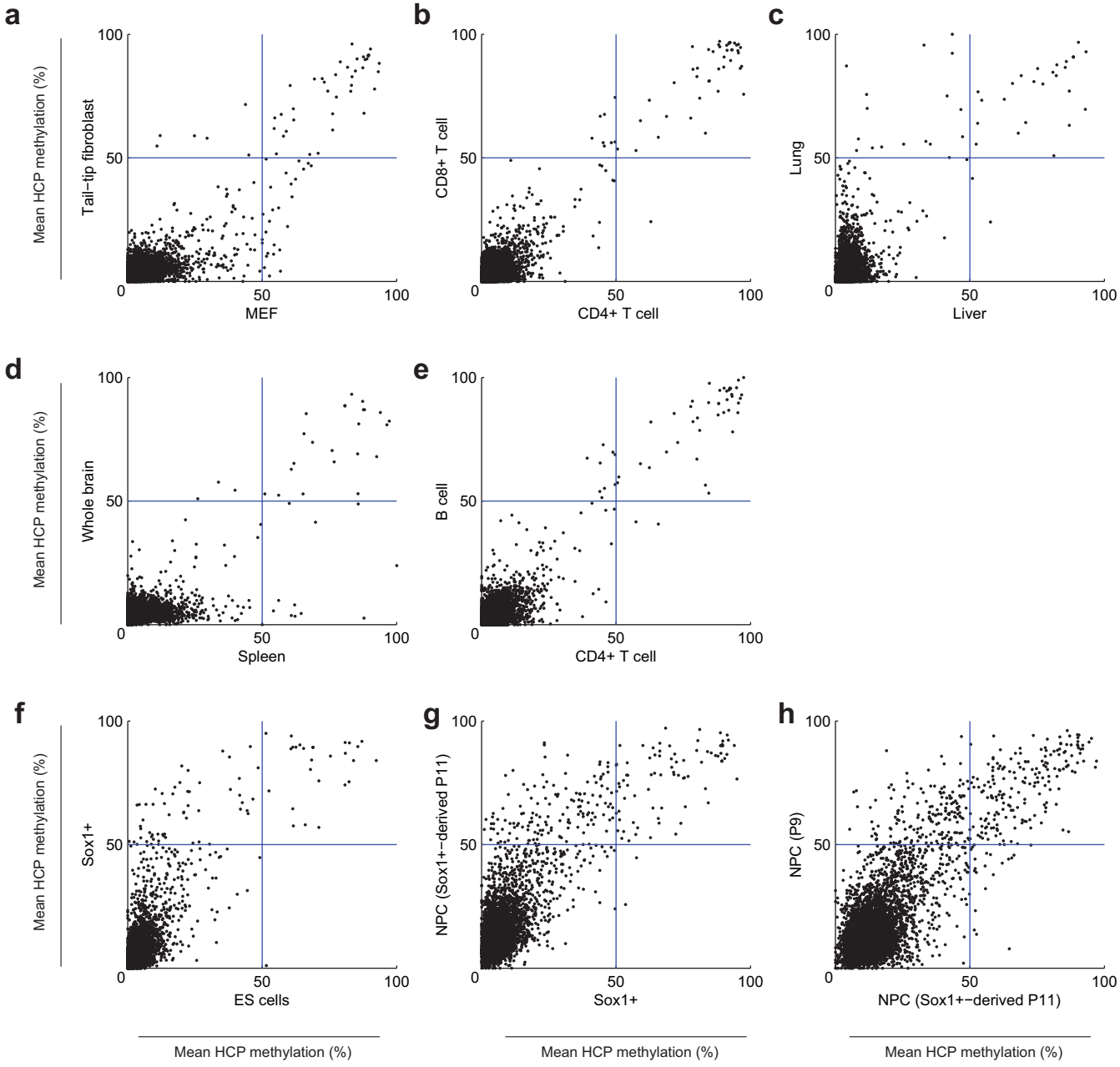
Correlations between histone methylation, expression levels and CpG methylation levels outside of annotated promoters and CpG islands in ES cells. **a**, H3K4me3 or H3K4me2 are correlated with low DNA methylation, whereas H3K36me3 and H3K27me3 alone is correlated with high DNA methylation levels. **b**, Distribution of methylation levels for CpGs overlapping known genes (excluding promoter regions), conditional on expression levels. Low = normalized absolute expression level < 50; Med >= 50 and < 200; High >= 200.

Supplementary Figure S7



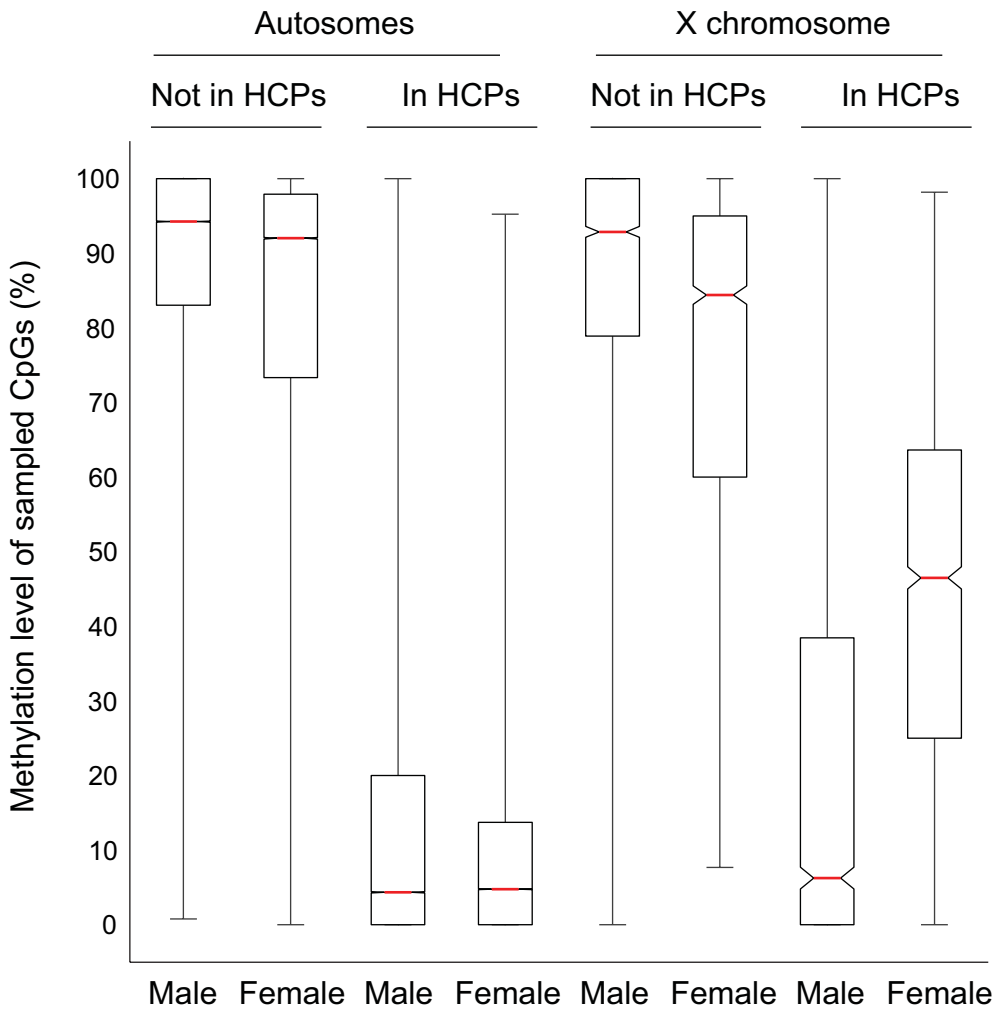
Distribution of CpG methylation levels inferred from a whole brain RRBS library. **a**, Distribution of inferred methylation levels for all CpGs with ≥10X coverage in either ES cells or NPCs. The top histograms show and the distribution of methylation levels (%) across all CpGs, high CpG density promoters (HCP), intermediante CpG density promoters (ICP), low CpG density promoters (LCP), highly conserved non-coding elements (HCNE), differentially methylated regions (DMR), long terminal repeats (LTR), short interspersed elements (SINE) and other genomic features (n gives the number of CpGs in each category).The distribution of methylation levels is bimodal and correlated with CpG density and genomic features in a pattern similar to the observed in ES cells (see main text). **b**, The distribution of CpG and histone methylation states for HCPs in ES cells and whole brain. The vast majority of HCPs that are univalent (H3K4me3) in ES cells also show this state in the brain sample. The vast majority of HCPs that are bivalent in ES cells, retain at least one of these marks in the brain sample (enrichment of H3K4me3 and H3K27me3 may not represent bivalency due to heterogeneity). The absence of both H3K4me3 and H3K27me3 correlates with hypermethylation. The red lines denote medians, notches the standard errors, boxes the interquartile ranges, and whiskers the 2.5th and 97.5th percentiles.

Supplementary Figure S8



Inferred mean methylation levels (%) of autosomal HCPs compared across different primary and ES-derived cell populations. **a-e**, primary cell types contain only ~20-30 hypermethylated HCPs, largely associated with germline-specific genes. **f-h**, Progressive hypermethylation of HCPs during continued proliferation of Sox1+ progentior cells. Sox1+ is the earliest known marker of neural progenitors and therefore allows isolation of a differentiated ES-derived population after minimal time in culture. There is initially little methylation in these cells, but after 11 passages in culture, many of the same HCPs that were methylated in the original NPC populations have also become methylated in Sox1+-derived NPCs.

Supplementary Figure S9



Comparison of methylation levels (%) for CpGs within and outside of HCPs in male and female cell populations (ES-derived and primary astrocytes, respectively). CpG islands show an average of ~50% methylation in the female population, consistent with hypermethylation of HCPs on the inactivated X-chromosome.