LETTERS

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], Jacob Hanna[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] & Eric S. Lander[1,2,7,8]

DNA methylation is essential for normal development[1–3] and has been implicated in many pathologies including cancer[4,5]. Our knowledge about the genome-wide distribution of DNA methylation, how it changes during cellular differentiation and how it relates to histone methylation and other chromatin modifications in mammals remains limited. Here we report the generation and analysis of genome-scale DNA methylation profiles at nucleotide resolution in mammalian cells. Using high-throughput reduced representation bisulphite sequencing[6] and single-molecule-based sequencing, we generated DNA methylation maps covering most CpG islands, and a representative sampling of conserved non-coding elements, transposons and other genomic features, for mouse embryonic stem cells, embryonic-stem-cell-derived and primary neural cells, and eight other primary tissues. Several key findings emerge from the data. First, DNA methylation patterns are better correlated with histone methylation patterns than with the underlying genome sequence context. Second, methylation of CpGs are dynamic epigenetic marks that undergo extensive changes during cellular differentiation, particularly in regulatory regions outside of core promoters. Third, analysis of embryonic-stem-cell-derived and primary cells reveals that 'weak' CpG islands associated with a specific set of developmentally regulated genes undergo aberrant hypermethylation during extended proliferation *in vitro*, in a pattern reminiscent of that reported in some primary tumours. More generally, the results establish reduced representation bisulphite sequencing as a powerful technology for epigenetic profiling of cell populations relevant to developmental biology, cancer and regenerative medicine.

DNA methylation can be detected by sequencing genomic DNA that has been treated with sodium bisulphite[7]. It has been impractical to apply bisulphite sequencing at a genome-wide scale because polymerase chain reaction (PCR)-based[8] and whole-genome shotgun[9] approaches are currently too inefficient for comparative analysis across multiple cell states in large mammalian genomes. However, reduced representations can be generated to sequence a defined fraction of a large genome[6,10]. Computational analysis indicated that digesting mouse genomic DNA with the methylation-insensitive restriction enzyme MspI, selecting 40–220-base pair (bp) fragments, and performing 36-bp end-sequencing would cover ~1 million distinct CpG dinucleotides (4.8% of all CpGs), with roughly half located within 'CpG islands' (including sequences from 90% of all CpG islands) and the rest distributed between other relatively CpG-poor sequence features (Supplementary Fig. 1 and Supplementary Table 1). Notably, although CpGs are not distributed uniformly in the genome, every MspI reduced representation bisulphite sequencing (RRBS) sequence read includes at least one informative CpG position (Supplementary Fig. 2), making the approach highly efficient.

We validated high-throughout RRBS by sequencing MspI fragments from wild-type and methylation-deficient embryonic stem (ES) cells[6], using an Illumina Genome Analyser. We generated an initial set of ~21 million high quality, aligned RRBS reads. The reads from each cell type included ~97% of the predicted non-repetitive MspI fragments (12-fold and 8-fold median coverage, respectively). This demonstrates that RRBS library construction is relatively unbiased (Supplementary Fig. 3) and is insensitive to genome-wide CpG methylation levels (estimated by nearest-neighbour analysis as 72% and 0.5%, respectively). Reads from both cell types showed near complete (>99%) bisulphite conversion of non-CpG cytosines.

To investigate cell-type-specific DNA methylation patterns, we generated 140 million additional RRBS reads (5.8 gigabase (Gb); Supplementary Information) from ES-derived neural precursor cells (NPCs) and various primary cell populations (Supplementary Table 2). We also generated new chromatin-state maps of H3 lysine 4 mono- and di-methylation (H3K4me1 and H3K4me2) from ES cells, NPCs and whole brain tissue (Supplementary Table 3 and Supplementary Information), using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq)[11].

The methylation levels of CpG dinucleotides in wild-type ES cells display a bimodal distribution (Fig. 1), with most being either 'largely unmethylated' (<20% of reads showing methylation) or 'largely methylated' (>80% of reads). As expected[2,8,12], CpGs in regions of high CpG density (>7% over 300 bp) tend to be unmethylated, whereas CpGs in low-density regions (<5%) tend to be methylated. However, we noted that ~10% of CpGs in low-density regions were unmethylated, whereas ~0.3% of CpGs in high-density regions were methylated. We found that DNA methylation patterns were better explained by histone methylation patterns than by CpG density. Because genomic features tend to be associated with distinct histone methylation patterns[11], we analysed these features separately.

High-CpG-density promoters (HCPs) are associated with two classes of genes: ubiquitous 'housekeeping' genes and highly regulated 'key developmental' genes[13]. In ES cells, HCPs at housekeeping genes are enriched with the transcription initiation mark H3K4me3 ('univalent') and are generally highly expressed, whereas those at developmental genes are enriched with both H3K4me3 and the repressive mark H3K27me3 ('bivalent') and are generally silent[11,14].

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Molecular Pathology Unit and Center for Cancer Research, MGH, Charlestown, Massachusetts 02129, USA. [6]Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [8]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02114, USA.
*These authors contributed equally to this work.

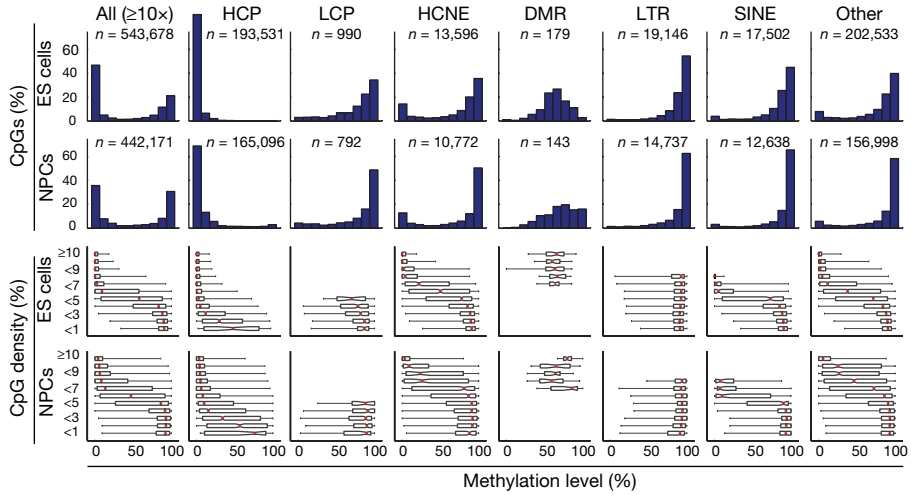

Figure 1 | CpG methylation levels in ES cells and NPCs for CpGs with ≥10-fold coverage. The top histograms show the distribution of methylation levels (%) across all CpGs, HCPs, LCPs, HCNEs, differentially methylated regions (DMRs), LTRs, SINEs and other genomic features (*n*, number of CpGs). Methylation levels are bimodal (except at DMRs, which have a unimodal distribution largely consistent with uniform sampling from the maternal and paternal alleles in ES cells and partial hypermethylation in NPCs). The bottom box plots show the distribution of methylation levels conditional on local CpG density (defined as fraction of CpGs in a 300-bp window; shown as percentage). The red lines denote medians, notches the standard errors, boxes the interquartile ranges, and whiskers the 2.5th and 97.5th percentiles.

Both types of promoters are also enriched with H3K4me2, which is associated with an open chromatin confirmation. Out of the 10,299 HCPs sampled (on average, 19 distinct CpGs per promoter), we found that virtually all contain a core region of unmethylated CpGs, regardless of their level of expression or H3K27me3 enrichment (Figs 1 and 2a)[12,14,15].

Low-CpG-density promoters (LCPs) are generally associated with tissue-specific genes. In ES cells, a small subset of LCPs are enriched with H3K4me3 (~7%) or H3K4me2 (~3%), and essentially none are enriched with H3K27me3 (ref. 11). We found that whereas most CpGs located in sampled LCPs (990 sites from 392 promoters) are methylated, those in LCPs enriched with H3K4me3 or H3K4me2 have significantly reduced methylation levels (Supplementary Fig. 4).

Distal regulatory regions such as enhancers, silencers and boundary elements are often required to establish correct gene expression patterns in mammalian cells[16]. *Cis*-regulatory elements active in a particular cell type are often associated with markers of open chromatin, such as H3K4me2 or H3K4me1 (refs 17, 18). We identified 25,051 sites of H3K4me2 enrichment in ES cells from 1 kb to >100 kb away from known promoters (most were also enriched with H3K4me1, but not with H3K4me3). CpGs sampled at H3K4me2-enriched sites (outside of promoters and CpG islands) had significantly lower methylation levels than those at unenriched sites (Fig. 2b). This relationship was particularly strong for CpGs located in highly conserved non-coding elements (HCNEs; Fig. 2c).

Imprinting control regions (ICRs) are CpG-rich regulatory regions that display allele-specific histone and DNA methylation[19]. Our RRBS library included sequences from 13 of ~20 known ICRs (on average, 13 distinct CpGs per ICR). CpGs within these elements display a unimodal distribution of methylation levels, with a median close to 50%, which is consistent with hypomethylation of the active allele marked with H3K4me3 and hypermethylation of the silenced allele marked with H3K9me3 (Fig. 1)[11].

Interspersed repeat families differ in their chromatin structure, with H3K9me3 enriched at active long terminal repeats (LTRs) and to a lesser extent at long interspersed elements (LINEs), but not at short interspersed elements (SINEs). Notably, CpGs located in LTRs and LINEs are generally hypermethylated, even in CpG-rich contexts

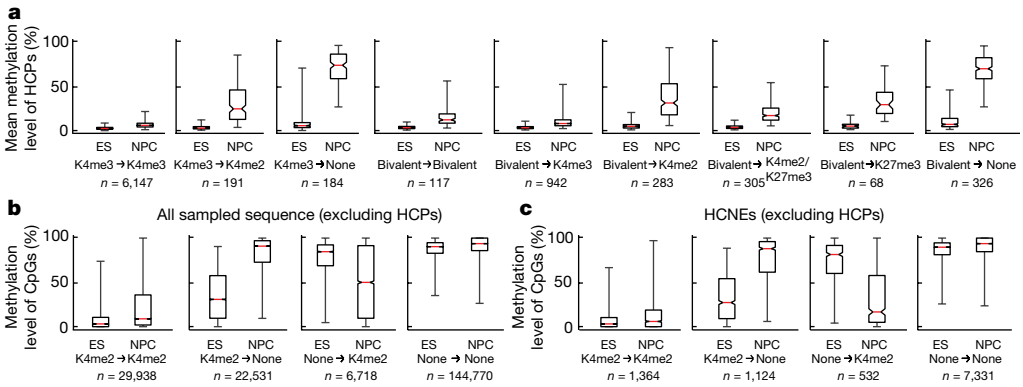

Figure 2 | Correlation between DNA and histone methylation. **a**, Mean methylation levels across CpGs within each profiled HCP (requiring ≥5-fold coverage of ≥5 CpGs), conditional on their histone methylation state in ES cells and NPCs (*n*, number of HCPs; those enriched with H3K4me3 are generally also enriched for H3K4me2, but not vice versa). Loss of H3K4 methylation, and to a lesser extent of H3K27me3, is correlated with gain of DNA methylation. **b**, Methylation levels of individual CpGs outside of HCPs, conditional on enrichment of H3K4me2 (*n*, number of distinct sites in each category). Changes in histone methylation state are inversely correlated with changes in DNA methylation. **c**, Methylation levels of CpGs in HCNEs not overlapping CpG islands, conditional on H3K4me2 enrichment. For **a**–**c**, the red lines denote medians, notches the standard errors, boxes the interquartile ranges, and whiskers the 2.5th and 97.5th percentiles. All pair-wise comparisons of methylation levels at sites with changing chromatin states are significant ($P < 10^{-20}$, Mann–Whitney U test).

(Fig. 1). In contrast, CpGs in SINEs show a correlation between methylation levels and CpG density that is comparable to non-repetitive sequences.

We conclude that in ES cells the presence of H3K4 methylation and the absence of H3K9 methylation are better predictors of unmethylated CpGs than sequence context alone. This is consistent with models in which *de novo* methyl-transferases either specifically recognize sites with unmethylated H3K4 (ref. 20) or are excluded by H3K4 methylation or associated factors. Similarly, H3K9me3 or associated factors may recruit methyl-transferases at ICRs and repetitive elements[21].

We next used RRBS to analyse how DNA methylation patterns change when ES cells are differentiated *in vitro* into a homogeneous population of NPCs (Supplementary Fig. 4)[22]. Whereas CpG methylation levels are highly correlated between the two cell types (rho = 0.81), there were clear differences: ~8% of CpGs unmethylated in ES cells became largely methylated in NPCs, whereas ~2% of CpGs methylated in ES cells became unmethylated; these changes were strongly correlated with changes in histone methylation patterns.

At both univalent and bivalent HCPs, we found that most CpGs remained unmethylated on differentiation, particularly within their core CpG island, but that loss of H3K4me3 and retention of H3K4me2 or H3K27me3 correlated with a partial increase in DNA methylation levels (median, ~25%; 2.9% and 32% of univalent and bivalent HCPs, respectively) and complete loss of H3K4 and H3K27 methylation correlated with DNA hypermethylation (median, ~75%; 2.8% and 16% of univalent and bivalent HCPs, respectively; Fig. 2).

Most LCPs marked by H3K4 methylation in ES cells lose this mark in NPCs; however, LCPs associated with genes expressed in NPCs gain this mark. Loss or gain of H3K4 methylation is a strong predictor of inverse changes in CpG methylation levels at these promoters (Supplementary Fig. 5).

Our chromatin-state maps revealed that 18,899 (75%) of putative distal regulatory elements enriched with H3K4me2 in ES cells lost this mark in NPCs, whereas 20,088 new H3K4me2 sites appeared, often in HCNE-rich regions surrounding activated developmental genes (Fig. 3). Loss or gain of H3K4 methylation were again inversely correlated with CpG methylation levels (Fig. 2b, c). In fact, these regions account for most observed de-methylation events. The presence of H3K27me3 alone did not correlate with lower methylation levels in CpG-poor regions (Supplementary Fig. 6).

The data support the notion that CpG-rich and -poor regulatory elements undergo distinct modes of epigenetic regulation[2,11,12]. Most (>95%) HCPs seem to be constitutively unmethylated and regulated by trithorax-group (trxG; associated with H3K4me3) and/or Polycomb-group (PcG; associated with H3K27me3) proteins, which may be recruited in part by means of non-specific unmethylated-CpG binding domains[23]. Hypermethylation of these CpG-dense regions leads to exclusion of trxG/PcG activity, heterochromatin formation and essentially irreversible gene silencing[2]. In contrast, regulatory elements in CpG-poor sequence contexts seem to undergo extensive and dynamic methylation and de-methylation. Hence, methylation of isolated CpGs may contribute to chromatin condensation or directly interfere with transcription factor binding[2], but does not necessarily prevent chromatin remodelling in response to activating signals.

As noted above, a small set of HCPs ($n = 252$; ~3%) became hypermethylated (>75% mean methylation across sampled CpGs) on *in vitro* differentiation of ES cells to NPCs. To investigate whether the observed pattern reflects an *in vivo* regulatory mechanism, we isolated NPCs from embryonic day (E)13.5 embryos and differentiated them into glial fibrillary acidic protein (Gfap)-positive astrocytes (with no more than two passages *in vitro*). We similarly differentiated the *in vitro*-derived NPCs into astrocytes (with these cells having undergone at least 18 passages; Supplementary Fig. 4), and compared the two populations using RRBS (Fig. 4a–f).

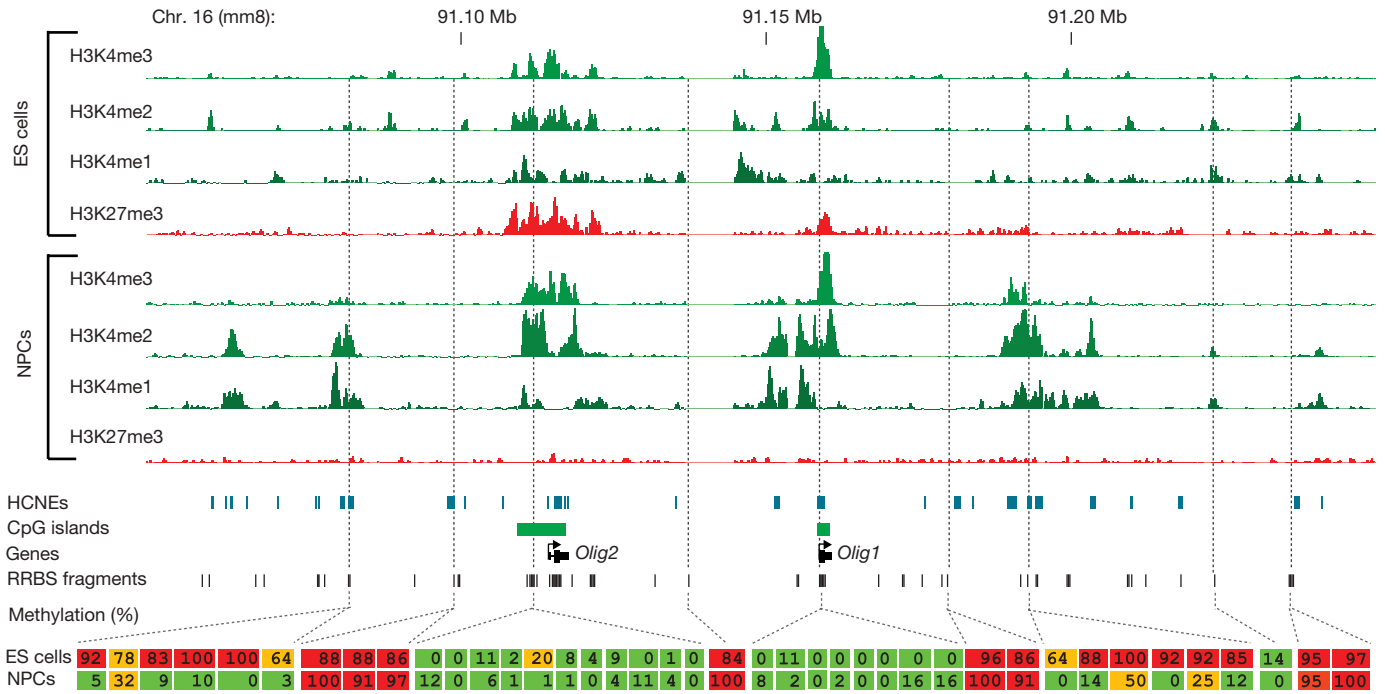

Figure 3 | Developmentally regulated de-methylation of highly conserved non-coding elements. Comparison of histone and DNA methylation levels across the *Olig1*/*Olig2* neural-lineage transcription factor locus. ChIP-Seq tracks for H3K4me1/2/3 and H3K27me3 in ES cells and NPCs are shown. The unmethylated CpG-rich promoters are bivalent and inactive in ES cells and resolve to univalent H3K4me3 on activation in NPCs. H3K4me2 enrichment appears over HCNEs distal to the two genes, and this correlates with CpG de-methylation. Inferred methylation levels for 40 out of 215 sampled CpGs are shown and colour-coded. Red indicates largely methylated (>80%); green indicates largely unmethylated (<20%), and orange indicates intermediate levels (≥20% and ≤80%).

The methylation levels of CpGs were highly correlated (rho = 0.85), but astrocytes obtained from *in vivo* NPCs displayed substantially less HCP hypermethylation than those obtained from ES cells (Fig. 4a). The *in vivo*-derived astrocytes showed hypermethylation at only 30 HCPs, largely associated with germline-specific genes (including *Dazl*, *Hormad1*, *Sycp1*, *Sycp2* and *Taf7l*), several of which also showed partial methylation in ES cells. In contrast, the *in vitro*-derived astrocytes showed hypermethylation of these and ~305 additional HCPs. This set includes some genes known to be expressed by at least some *in vivo* astrocytes (including *Isyna1*, *Gsn* and *Cldn5*; ref. 24) but that were silent in the ES-cell-derived astrocytes (Supplementary Information). However, the hypermethylated HCPs are significantly enriched for genes not expressed in NPCs or in the astrocyte lineage (Supplementary Tables 4–7). They include genes involved in development and differentiation of neuronal (*Lhx8*, *Lhx9*, *Moxd1*, *Htr1f* and *Slit1*), ependymal (*Otx2* and *Kl*) and unrelated lineages (including *Myod1*, *Dhh* and *Nkx3-1*). In fact, we found that 'key developmental' HCPs that are bivalent in ES cells are six times more likely to be included in the hypermethylated set compared to univalent HCPs. Moreover, univalent genes in the hypermethylated set are expressed at significantly lower levels in both ES cells and primary astrocytes, compared to those that remained hypomethylated (Fig. 4g). We also found that the hypermethylated HCPs tend to have a ~15% lower CpG density (Fig. 4h).

To investigate further the differences between *in vitro* and *in vivo* cell populations, we analysed whole brain tissue (representing cells of mainly glial origin). Virtually all (>99%) of sampled HCPs were unmethylated (Fig. 4c) and enriched with H3K4me3 and/or H3K27me3 (Supplementary Fig. 7), with ~20 germline-specific HCPs being the only clear exceptions. RRBS libraries from other *in vivo* sources (T cells, B cells, spleen, lung, liver and fibroblasts) also showed few hypermethylated HCPs (Supplementary Fig. 8). This suggests that—apart from silencing germline-specific[12], imprinted and X-inactivated (Supplementary Fig. 9) genes in somatic tissues—hypermethylation of HCPs is not a major mechanism of developmental regulation *in vivo*.

To test for a correlation between passage number and HCP hypermethylation, we examined independently derived *in vitro* NPCs collected after only 9 passages. These cells displayed hypermethylation at approximately half of the HCPs that are hypermethylated in the NPCs after 18 passages (Fig. 4d, e). To reduce time in culture further, we used Sox1–GFP (green fluorescent protein) ES cells[25] to isolate very early NPCs. These cells initially displayed virtually no HCP hypermethylation. However, after continued culturing they acquired hypermethylation at many of the same HCPs as the previous NPC populations (Supplementary Fig. 8). Finally, we grew the *in vivo*-derived NPCs for 11 passages *in vitro*, differentiated them into astrocytes and then examined the methylation pattern. Notably, these cells had also begun to acquire hypermethylation at a largely similar set of HCPs (Fig. 4a, b).

These results show that independently derived NPC populations from both *in vitro* and *in vivo* sources and different genetic backgrounds reproducibly undergo gradual hypermethylation at a characteristic set of HCPs. These observations have several implications.

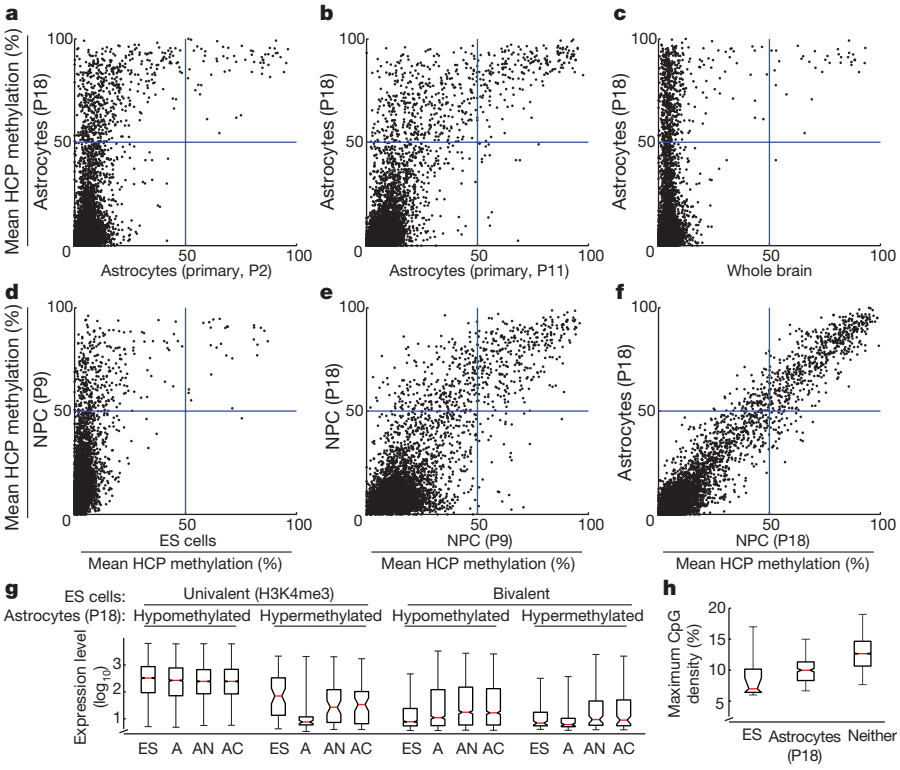

Figure 4 | HCP hypermethylation of cultured cells. Inferred mean methylation levels (%) across autosomal HCPs (requiring ≥5-fold coverage of ≥5 CpGs within the CpG island). **a**, ES-derived astrocytes contains roughly 10 times more hypermethylated HCPs than primary NPC-derived astrocytes after two passages (P) in culture. **b**, Continued passage of the primary cells lead to gradual hypermethylation of many of the same HCPs. **c**, Only a handful of mainly germline-specific HCPs display hypermethylation in a whole brain tissue sample. **d**, Most HCPs are unmethylated in ES cells, but a small subset gain significant methylation on differentiation to NPCs. **e**, Continued proliferation of NPCs leads to additional HCPs becoming hypermethylated after 18 passages. **f**, Differentiation of late-stage NPCs into astrocytes by growth factor withdrawal does not lead to additional HCP hypermethylation. **g**, Expression levels of genes associated with profiled HCPs for ES cells (ES), ES-derived astrocytes (A), primary neocortical astrocytes (AN) and cerebellar astrocytes (AC). Hypermethylation of HCPs is correlated with low expression levels in ES-derived astrocytes. HCPs that are univalent in ES cells and become hypermethylated in ES-derived astrocytes are associated with lower expression levels in both ES cells and primary astrocytes. **h**, The maximal CpG densities (300-bp window) of hypermethylated HCPs in ES cells or ES-derived astrocytes are significantly lower than for unmethylated HCPs. For **g** and **h**, the red lines denote medians, notches the standard errors, boxes the interquartile ranges, and whiskers the 2.5th and 97.5th percentiles.

First, aberrant epigenetic regulation in culture has raised concern over the accuracy of cellular models generated by *in vitro* differentiation or manipulation[26–28]. Both primary and transformed cell lines, including ES-derived NPC populations, tend to lose developmental potency after continued proliferation in culture[26,29]. Susceptibility to hypermethylation at key regulatory genes that are normally activated on differentiation could explain this phenomenon. Second, malignant cells are often found to harbour hypermethylated CpG islands[4,5]. Recently, genes known to undergo frequent hypermethylation in adult cancers were noted to be significantly enriched for genes with bivalent promoters in ES cells (reviewed in ref. 30). The similarities between hypermethylation in culture and in cancer may provide a useful *in vitro* model for studying a common underlying mechanism. Finally, the gradual hypermethylation of 'weak' HCPs hints at underlying kinetics. Because H3K4 methylases are targeted, at least in part, by non-specific CpG-binding domains[23], such HCPs may be particularly sensitive to imbalanced chromatin-modifying factors or other cancer- or culture-related perturbations.

More generally, RRBS makes it feasible to perform genome-scale bisulphite sequencing on large-mammalian genomes, providing a valuable tool for epigenetic profiling of cell populations. As sequencing capacity increases, genome coverage can be readily scaled in step by adding restriction enzymes, increasing the selected size range or using hybridization-based reduced representation strategies.

METHODS SUMMARY

ES cells and ES-derived neural cells were cultured as described previously[11,25]. Primary tissues were isolated from 4–6-week-old male 129SvJae/C57/B6 mice. Mouse embryonic fibroblasts (MEFs) and primary neural precursors were isolated from 129SvJae/C57/B6 E14.5 embryos.

RRBS libraries were prepared from 1–10 μg mouse genomic DNA digested with 10–100 Units MspI (NEB). Size-selected MspI fragments (40–120 bp and 120–220 bp) were filled in and 3′-terminal-A extended, extracted with phenol and precipitated with ethanol. Ligation to pre-annealed adapters containing 5′-methyl-cytosine instead of cytosine (Illumina) was performed using the Illumina DNA preparation kit and protocol. QIAquick (Qiagen) cleaned-up, adaptor-ligated fragments were bisulphite-treated using the EpiTect Bisulphite Kit (Qiagen). Preparative-scale PCR was performed and QIAquick-purified PCR products were subjected to a final size selection on a 4% NuSieve 3:1 agarose gel. SYBR-green-stained gel slices containing adaptor-ligated fragments of 130–210 bp or 210–310 bp in size were excised. Library material was recovered from the gel (QIAquick) and sequenced on an Illumina 1G genome analyser.

Sequence reads from bisulphite-treated Solexa libraries were identified using standard Illumina base-calling software and then analysed using a custom computational pipeline. ChIP-Seq experiments, sequencing, alignments and identification of significantly enriched regions were carried out as described previously[11].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March; accepted 21 May 2008.
Published online 6 July 2008.

1. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9,** 2395–2402 (2000).
2. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
3. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447,** 425–432 (2007).
4. Feinberg, A. P. The epigenetics of cancer etiology. *Semin. Cancer Biol.* **14,** 427–432 (2004).
5. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128,** 683–692 (2007).
6. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33,** 5868–5877 (2005).
7. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89,** 1827–1831 (1992).
8. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genet.* **38,** 1378–1385 (2006).
9. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452,** 215–219 (2008).
10. Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407,** 513–516 (2000).
11. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
12. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39,** 457–466 (2007).
13. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103,** 1412–1417 (2006).
14. Bernstein, B. *et al.* A bivalent chromatin structure marks key Developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
15. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6,** e22 (2008).
16. West, A. G. & Fraser, P. Remote control of gene transcription. *Hum. Mol. Genet.* **14** (Spec No 1) R101–R111 (2005).
17. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
18. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–181 (2005).
19. Edwards, C. A. & Ferguson-Smith, A. C. Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **19,** 281–289 (2007).
20. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
21. Esteve, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20,** 3089–3103 (2006).
22. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3,** e283 (2005).
23. Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A. & Skalnik, D. G. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* **20,** 2108–2121 (2000).
24. Sharma, M. K. *et al.* Distinct genetic signatures among pilocytic astrocytomas relate to their brain region origin. *Cancer Res.* **67,** 890–900 (2007).
25. Aubert, J. *et al.* Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from *Sox1–gfp* knock-in mice. *Proc. Natl Acad. Sci. USA* **100** (Suppl 1), 11836–11841 (2003).
26. Jones, P. A., Wolkowicz, M. J., Harrington, M. A. & Gonzales, F. Methylation and expression of the Myo D1 determination gene. *Phil. Trans. R. Soc. Lond. B* **326,** 277–284 (1990).
27. Smiraglia, D. J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10,** 1413–1419 (2001).
28. Shen, Y., Chow, J., Wang, Z. & Fan, G. Abnormal CpG island methylation occurs during *in vitro* differentiation of human embryonic stem cells. *Hum. Mol. Genet.* **15,** 2623–2635 (2006).
29. Bouhon, I. A., Joannides, A., Kato, H., Chandran, S. & Allen, N. D. Embryonic stem cell-derived neural progenitors display temporal restriction to neural patterning. *Stem Cells* **24,** 1908–1913 (2006).
30. Ohm, J. E. & Baylin, S. B. Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? *Cell Cycle* **6,** 1040–1043 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of the Broad Institute Genome Sequencing Platform for assistance with data generation and B. Ramsahoye for the nearest neighbour analysis. This research was supported by funds from the National Human Genome Research Institute, the National Cancer Institute, and the Broad Institute of MIT and Harvard.

Author Information All primary sequencing data have been submitted to the NCBI GEO repository under accession numbers GSE11034 (RRBS), GSE11172 (ChIP-Seq) and GSE11483 (gene expression microarrays). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.J. (jaenisch@wi.mit.edu) or E.S.L. (lander@broad.mit.edu).

METHODS

Cell culture and ES cell differentiation. V6.5 (129/B6), Sox1–EGFP knock-in (Sox1–GFP 129/129)[25] and methylation-deficient (Dnmt[1^{kd},$3a^{-/-}$,$3b^{-/-}$]) ES cells were expanded on γ-irradiated MEFs in DMEM plus 15% fetal bovine serum (FBS, Hyclone) supplemented with 1× MEM-nonessential amino acids (Life Technologies), 0.1 mM 2-mercaptoethanol and 10^3 Units ml^{-1} leukaemia inhibitory factor (LIF). After passaging onto gelatin-coated dishes (0.1% gelatin, Sigma), ES cells were trypsinized and transferred to bacterial dishes allowing embryoid body formation. Embryoid bodies were propagated for 4 days in the same medium in the absence of LIF and subsequently plated onto tissue culture dishes. One day after plating, the medium was replaced by ITSFn; that is, DMEM/F12 (Life Technologies) supplemented with 5 μg ml^{-1} insulin, 50 μg ml^{-1} human APO transferrin, 30 nM sodium selenite (all Sigma), 2.5 μg ml^{-1} fibronectin and penicillin/streptomycin (both Life Technologies). After 5–7 days, cells were trypsinized, triturated to a single cell suspension, replated on laminin-coated dishes (1 μg ml^{-1}, Life Technologies) and further propagated in N3 medium composed of DMEM/F12, 25 μg ml^{-1} insulin, 50 μg ml^{-1} transferrin, 30 nM sodium selenite, 20 nM progesterone, 100 nM putrescine (Sigma), 10 ng ml^{-1} Fgf2 (R&D Systems) and penicillin/streptomycin. Neural precursor cell proliferation was maintained by daily additions of Fgf2. Sox1–EGFP-positive neural precursors were isolated and FACS-purified (FACS Aria, Becton Dickinson) either from ITSFn cultures or after short-term expansion in Fgf2. Growth factor withdrawal of these cultures results in terminal differentiation into primarily neuronal cell populations. Neural precursor cell lines were obtained by sequential passaging and propagation in the presence of 20 ng ml^{-1} Egf and 10 ng ml^{-1} Fgf2 (both R&D Systems). Differentiation into astrocytes was induced by growth factor withdrawal and addition of 5% FBS for 5 days.

Primary tissues and cell types. Primary tissues were isolated from 4–6-week-old male 129SvJae/C57/B6 mice. MEFs and primary neural precursors were isolated from 129SvJae/C57/B6 E14.5 embryos. MEFs were generated according to standard protocols. *In vivo* neural precursors were isolated by disaggregating the whole brain and plating the suspension under the conditions described previously. Established lines were differentiated into astrocytes by growth factor withdrawal and addition of serum (see previously).

MspI RRBS library construction. 1–10 μg mouse genomic DNA was digested with 10–100 Units of MspI (NEB) in a 30–500 μl reaction 16–20 h at 37 °C. Digested DNA was extracted with phenol, precipitated with ethanol and size-selected on a 4% NuSieve 3:1 agarose gel (Lonza). DNA marker lanes were excised from the gel and stained with SYBR Green (Invitrogen). For each sample, two slices containing DNA fragments of approximately 40–120 bp and 120–220 bp, respectively, were excised from the unstained preparative portion of the gel. DNA was recovered using Easy Clean DNA spin filters (Primm Labs), extracted with phenol and precipitated with ethanol. The two size fractions were kept apart throughout the procedure, including during the final sequencing. Size-selected MspI fragments were filled in and 3′-terminal A extended in a 50 μl reaction containing 20 U Klenow exo$^-$ (NEB), 0.4 mM dATP, 0.04 mM dGTP and 0.04 mM 5-methyl-dCTP (Roche) in 1× NEB buffer 2 (15 min at 25 °C followed by 15 min at 37 °C), extracted with phenol and precipitated with ethanol using 10 μg glycogen (Roche) as a carrier. Ligation to pre-annealed Illumina adapters containing 5′-methyl-cytosine instead of cytosine (Illumina) was performed using the Illumina DNA preparation kit and protocol. QIAquick (Qiagen) cleaned-up, adaptor-ligated fragments were bisulphite-treated using the EpiTect Bisulphite Kit (Qiagen) with minor modifications: the bisulphite conversion time was increased to approximately 14 h by adding three cycles (5 min of denaturation at 95 °C followed by 3 h at 60 °C). After bisulphite conversion, the single-stranded uracil-containing DNA was eluted in 20 μl of elution buffer. Analytical (25 μl) PCR reactions containing 0.5 μl of bisulphite-treated DNA, 5 pmol each of genomic PCR primers 1.1 and 2.1 (Illumina) and 2.5 U PfuTurboCx Hotstart DNA polymerase (Stratagene) were set up to determine the minimum number of PCR cycles required to recover enough material for sequencing. Preparative-scale (8 × 25 μl) PCR was performed using the same PCR profile: 5 min at 95 °C, n × (30 s at 95 °C, 20 s at 65 °C, 30 s at 72 °C), followed by 7 min at 72 °C, with n ranging from 18 to 24 cycles. QIAquick-purified PCR products were subjected to a final size selection on a 4% NuSieve 3:1 agarose gel. SYBR-green-stained gel slices containing adaptor-ligated fragments of 130–210 bp or 210–310 bp in size were excised. RRBS library material was recovered from the gel (QIAquick) and sequenced on an Illumina 1G genome analyser.

Sequence alignments and data analysis. Sequence reads from bisulphite-treated Solexa libraries were identified using standard Illumina base-calling software and then analysed using a custom computational pipeline. Residual cytosines (Cs) in each read were first converted to thymines (Ts), with each such conversion noted for subsequent analysis. A reference sequence database was constructed from the 36-bp ends of each computationally predicted MspI fragment in the 40–220-bp size range. All Cs in each fragment end were then converted to Ts (only the C-poor strands are sequenced in the RRBS process; Supplementary Fig. 2).

The converted reads were aligned to the converted reference by finding all 12-bp perfect matches and then extending to both ends of the treated read, not allowing gaps (reverse complement alignments were not considered). The number of mismatches in the induced alignment was then counted between the unconverted read and reference, ignoring cases in which a T in the unconverted read is matched to a C in the unconverted reference. For a given read, the best alignment was kept if the second-best alignment had ≥2 more mismatches, otherwise the read was discarded as non-unique. Low-quality reads were identified and discarded if $\sum_{q \in Q} 10^{q/10} > 1{,}000$, where Q denotes the read quality scores at each mismatched position. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T, counting only reads with quality scores of ≥20 at the position.

HCP, ICP and LCP annotations were taken from ref. 11. CpG island and other annotations were downloaded from the UCSC browser (mm8). Estimation of methylation levels from individual CpGs was limited to those with ≥10-fold coverage. The methylation level of an HCP promoter was estimated as the mean methylation level across all CpGs with ≥5-fold coverage overlapping the annotated CpG island(s) in the promoter, requiring at least 5 such CpGs. HCPs were classified as hypermethylated if this mean methylation level was ≥75%.

Chromatin immunoprecipitation. H3K4me1 (ab8895), H3K4me2 (ab7766) and H3K4me3 (ab8580) antibodies were purchased from Abcam. ChIP experiments on mouse ES cells (H3K4me1/2), NPCs (H3K4me1/2) and whole brain tissue (H3K4me1/2/3), Illumina/Solexa sequencing, alignments and identification of significantly enriched regions (using 1 kb sliding windows and correction for alignability) were carried out as described previously[11].

Expression data. RNA expression data for ES-derived astrocytes were generated as described previously[11] and analysed using GenePattern (http://www.broad.mit.edu/cancer/software/genepattern/). Primary astrocyte data were obtained from ref. 24.