Global Epigenomic Reconfiguration During Mammalian Brain Development

Ryan Lister,* Eran A. Mukamel, Joseph R. Nery, Mark Urich, Clare A. Puddifoot, Nicholas D. Johnson, Jacinta Lucero, Yun Huang, Andrew J. Dwork, Matthew D. Schultz, Miao Yu, Julian Tonti-Filippini, Holger Heyn, Shijun Hu, Joseph C. Wu, Anjana Rao, Manel Esteller, Chuan He, Fatemeh G. Haghighi, Terrence J. Sejnowski, M. Margarita Behrens,* Joseph R. Ecker*

Introduction: Several lines of evidence point to a key role for dynamic epigenetic changes during brain development, maturation, and learning. DNA methylation (mC) is a stable covalent modification that persists in post-mitotic cells throughout their lifetime, defining their cellular identity. However, the methylation status at each of the ~1 billion cytosines in the genome is potentially an information-rich and flexible substrate for epigenetic modification that can be altered by cellular activity. Indeed, changes in DNA methylation have been implicated in learning and memory, as well as in age-related cognitive decline. However, little is known about the cell type–specific patterning of DNA methylation and its dynamics during mammalian brain development.

Methods: We performed genome-wide single-base resolution profiling of the composition, patterning, cell specificity, and dynamics of DNA methylation in the frontal cortex of humans and mice throughout their lifespan (MethylC-Seq). Furthermore, we generated base-resolution maps of 5-hydroxymethylcytosine (hmC) in mammalian brains by TAB-Seq at key developmental stages, accompanied by RNA-Seq transcriptional profiling.

Results: Extensive methylome reconfiguration occurs during development from fetal to young adult. In this period, coincident with synaptogenesis, highly conserved non-CG methylation (mCH) accumulates in neurons, but not glia, to become the dominant form of methylation in the human neuronal genome. We uncovered surprisingly complex features of brain cell DNA methylation at multiple scales, first by identifying intragenic methylation patterns in neurons and glia that distinguish genes with cell type–specific activity. Second, we report a novel mCH signature that identifies genes escaping X-chromosome inactivation in neurons. Third, we find >100,000 developmentally dynamic and cell type–specific differentially CG-methylated regions that are enriched at putative regulatory regions of the genome. Finally, whole-genome detection of 5-hydroxymethylcytosine (hmC) at single-base resolution revealed that this mark is present in fetal brain cells at locations that lose CG methylation and become activated during development. CG-demethylation at these hmC-poised loci depends on Tet2 activity.

Discussion: Whole-genome single-base resolution methylcytosine and hydroxymethylcytosine maps revealed profound changes during frontal cortex development in humans and mice. These results extend our knowledge of the unique role of DNA methylation in brain development and function, and offer a new framework for testing the role of the epigenome in healthy function and in pathological disruptions of neural circuits. Overall, brain cell DNA methylation has unique features that are precisely conserved, yet dynamic and cell-type specific.

The DNA methylation landscape of human and mouse neurons is dynamically reconfigured through development. Base-resolution analysis allowed identification of methylation in the CG and CH context (H = A, C, or T). Unlike other differentiated cell types, neurons accumulate substantial mCH during the early years of life, coinciding with the period of synaptogenesis and brain maturation.

Fig. 1. Methylcytosine in mammalian frontal cortex is developmentally dynamic and abundant in CG and CH contexts.

Fig. 2. mCH is positionally conserved and is the dominant form of DNA methylation in human neurons.

Fig. 3. mCH is enriched in genes that escape X inactivation.

Fig. 4. Cell type–specific and developmental differences in mC between mouse neurons and glia.

Fig. 5. hmCG is enriched within active genomic regions in fetal and adult mouse brain.

Fig. 6. Developmental and cell type–specific differential mCG.

SUPPLEMENTARY MATERIALS
Materials and Methods Figs. S1 to S12 Tables S1 to S5 References (63–78)
Global Epigenomic Reconfiguration During Mammalian Brain Development

Ryan Lister,1,2,*† Eran A. Mukamel,3, Joseph R. Nery,1 Mark Urlich,1 Clare A. Puddifoot,3 Nicholas D. Johnson,1 Jacinta Lucero,3 Yun Huang,4 Andrew J. Dwork,5,6 Matthew D. Schultz,1,7 Miao Yu,8 Julian Tonti-Filippini,2 Holger Heyn,9 Shijun Hu,10 Joseph C. Wu,10 Anjana Rao,4 Manel Esteller,7,11 Chuan He,8 Fatemeh G. Haghihighi,5 Terrence J. Sejnowski,3,12,13 M. Margarita Behrens,4,† Joseph R. Ecker1,3,†

DNA methylation is implicated in mammalian brain development and plasticity underlying learning and memory. We report the genome-wide composition, patterning, cell specificity, and dynamics of DNA methylation at single-base resolution in human and mouse frontal cortex throughout their lifespan. Widespread methylome reconfiguration occurs during fetal to young adult development, coincident with synaptogenesis. During this period, highly conserved non-CG methylation (mCH) accumulates in neurons, but not glia, to become the dominant form of methylation in the human neuronal genome. Moreover, we found an mCH signature that identifies genes escaping X-chromosome inactivation. Last, whole-genome single-base resolution 5-hydroxymethylcytosine (hmC) maps revealed that hmC marks fetal brain cell genomes at putative regulatory regions that are CG-demethylated and activated in the adult brain and that CG demethylation at these hmC-poised loci depends on Tet2 activity.

Dynamic epigenetic changes have been observed during brain development, maturation, and learning (1–6). DNA methylation (mC) is a stable covalent modification that persists in postmitotic cells throughout their lifetime, defining their cellular identity. However, the methylation status at each of the ~1 billion cytosines in the genome is potentially an information-rich and flexible substrate for epigenetic modification that can be altered by cellular activity (7, 8). Changes in DNA methylation were implicated in learning and memory (9, 10), as well as in age-related cognitive decline (11). Mice with a postnatal deletion of DNA methyltransferases Dnmt1 and Dnmt3a in forebrain excitatory neurons, or with a global deletion of methyl-Cpg-binding protein 2 (MeCP2), show abnormal long-term neural plasticity and cognitive deficits (2, 12).

DNA methylation composition and dynamics in the mammalian brain are highly distinct. A modification of mC catalyzed by the Tet family of mC hydroxylase proteins, 5-hydroxymethylcytosine (hmC), accumulates in the adult brain (13–15) along with its more highly oxidized derivatives 5-formylcytosine and 5-carboxylcytosine. These modifications of mC were implicated as intermediates in an active DNA demethylation pathway (16–19). In addition, methylation in the non-CG context (mCH, where H = A, C, or T) is also present in the adult mouse and human brains (20, 21) but is rare or absent in other differentiated cell types (22, 23). Little is known about cell type–specific patterning of DNA methylation and its dynamics during mammalian brain development. Here, we provide integrated empirical data and analysis of DNA methylation at single-base resolution, across entire genomes, with cell type- and developmental specificity. These results extend our knowledge of the unique role of DNA methylation in brain development and function and offer a new framework for testing the role of the epigenome in healthy function and in pathological disruptions of neural circuits.

Accumulation of Non-CG DNA Methylation During Brain Development

To identify the composition and dynamics of transcription and methylation during mammalian brain development, we performed transcriptome profiling (mRNA-Seq) and whole-genome bisulfite sequencing [MethylIC-Seq (24)] to comprehensively identify sites of cytosine DNA methylation (mC and hmC) and mRNA abundance at single-base resolution throughout the genomes of mouse and human frontal cortex (table S1). DNA methylation in embryonic stem (ES) cells occurs in both the CG (mCG) and non-CG (mCH) contexts, but mCH is largely lost upon cell differentiation (22, 23, 25, 26). We found that although mCH levels are negligible in fetal cortex, abundant mCH occurs in adult frontal cortex (Fig. 1A). mCH has previously been identified throughout the genome of the adult mouse brain (20) and at several hundred genomic positions in the human adult brain (21). Supporting previous studies, we found that mammalian brain mCH is typically depleted in expressed genes, with genic mCH level inversely proportional to the abundance of the associated transcript (Fig. 1, A and B) (20). This pattern is the opposite of that observed in ES cells (22) and suggests that genic mCH in the brain may inhibit transcription. The absence of mCH in fetal brain suggests that this signature for gene repression is added to the genome at a later developmental stage.

We performed MethylIC-Seq on mouse and human frontal cortex during early postnatal, juvenile, adolescent, and adult stages (Fig. 1C). CH methylation level, defined as the fraction of all base calls at CH genome reference positions that were methylated (denoted mCH/CH), accumulates in mouse and human brain during early postnatal development to a maximum of 1.3 to 1.5% genome-wide at the end of adolescence before diminishing slightly during aging. mCH increases most rapidly during the primary phase of synaptogenesis in the developing postnatal brain, from 2 to 4 weeks in mouse (27) and in the first 2 years in humans (28), followed by slower accumulation of mCH during later adolescence. mCH accumulation initially parallels the increase in synapse density within human middle frontal gyrus (synaptogenesis lasts from birth to 5 years), but it subsequently continues to increase during the period of adolescent synaptic pruning, which in humans occurs between 5 and 16 years of age (Fig. 1C). Notably, the accumulation of mCH in mice from 1 to 4 weeks after birth coincides with a transient increase in abundance of the de novo methyltransferase Dnmt3a mRNA and protein (Fig. 1D). Analysis of the context of mCH sites showed that it is mainly present in the CA context (fig. S1, A to F), as previously reported for mCH (20, 22, 23, 26).

Overall, genomes in the frontal cortex are highly methylated. Whereas CG partially methylated domains (PMDs) account for about a third of the genome of various differentiated human cells (22, 25), human brain genomes have negligible CG PMDs, resembling pluripotent cell methylomes (25) (fig. S1, G and H). Given the high spatial concordance of CG PMDs and nuclear lamina-associated domains reported previously (29), the
Fig. 1. Methylcytosine in mammalian frontal cortex is developmentally dynamic and abundant in CG and CH contexts. (A) Browser representation of mC and mC transcript abundance in human and mouse frontal cortex and human ES cells. Chr2, chromosome 2. (B) mC/CH within gene bodies exhibits opposite correlation with gene expression in ES cells (ESC) and brain. Contours show data point density, and red line shows smoothed mC/CH as a function of mRNA. a.u., arbitrary units. (C) Synaptic density quantitation from De Felipe et al. (27) and Huttenlocher and Dabholkar (28). (D) DNA methyltransferase mRNA and protein abundance (mean ± SEM) in mouse frontal cortex through development. FPKM, fragments per kilobase of exon per million fragments mapped. (E) Fraction of cytosine base calls with each modification in fetal and adult mouse frontal cortex. (F) Cortex mC level in CG and CH contexts throughout mouse and human chromosome 12 in 100-kb bins smoothed with ~1-Mb resolution. CEN, centrosome. (G) Transcript abundance, chromatin accessibility [8-week mouse cortex ChIP input and DNaseI hypersensitivity (HS) normalized read density], and mC levels in 5-kb bins at the mouse Immunoglobulin VH locus. (H) Density (z) plot of 10-week mouse frontal cortex mC level (x) versus 8-week mouse cortex ChIP-input normalized read density (y) for all 10-kb bins of the mouse genome.
Fig. 2. mCH is positionally conserved and is the dominant form of DNA methylation in human neurons. (A) Browser representation of mCG and mCH in NeuN+ and NeuN– cells. Human NeuN+ vs NeuN– samples: R1, 53-year-old female; R2, 55-year-old male. Mouse NeuN+ vs NeuN– samples: R1, 7-week males; R2, 6-week females; R3, 12-month females (not shown). (B) Percentage of methylated base calls in each sequence context throughout the genome. (C) Box and whisker plot of mCG and mCH level in neurons and glia at genomic regions bound by Dnmt3a versus a random set. Whiskers indicate 1.5 times the interquartile range. (D) mCH correlation between NeuN+ and NeuN– cells in mouse and human, measured in 10-kb bins. (E) Browser representation of mCH sites in neurons. Scatter plots (right) show consistent mCH/CH at all single sites in a 20-kb window overlapping the example region (left). (F) Correlation analysis of methylation state at single sites between neurons and ES cells in human and mouse. Correlation values are normalized by a simulation (62).
mCH = 25.6%, hmC = 17.2%). These data suggest that the steady-state population of hmC in the adult brain is not an intermediate stage in the demethylation of mCH. However, these steady-state measurements do not preclude the possibility that hmC could be rapidly turned over after conversion from mCH, leading to negligible detected hmC despite Tet-mediated demethylation at CH sites.

Protection of Inaccessible Genomic Regions from de Novo Methylation

mCH accumulates in parallel across most of the genome (Fig. 1F). However, we found numerous (36 in human, 34 in mouse) noncentromeric, megabase-sized regions that do not accumulate mCH. These regions, which we termed mCH deserts, are enriched for large gene clusters that encode proteins involved in immunity and receptors required for sensory neuron function (table S2). One mCH desert spans the immunoglobulin V_H locus, which encodes variable domains of the immunoglobulin heavy chain that rearrange in B lymphocytes. The V_H locus is transcriptionally quiescent in the frontal cortex of 10-week-old mice, and the chromatin state is highly inaccessible, as inferred from deoxyribonuclease I (DNaseI) hypersensitivity profiling (35) and chromatin immunoprecipitation (ChIP) input sequence read density data (36, 37) (Fig. 1G). In contrast, mCG is not depleted in mCH deserts.

Genome-wide detection of hmC by cytosine 5-methylenesulphonate immunoprecipitation (CMS-IP) (38, 39) revealed that hmC is also strongly depleted in the V_H locus. mCH deserts are observed at other loci in the genome, including olfactory receptor gene clusters that form heterochromatic aggregates required for monoallelic receptor expression in olfactory sensory neurons (40, 41). Genome-wide comparison of mCH/CH with chromatin accessibility, as inferred from ChIP input read density (36, 37), for all 10-kb windows of the mouse genome revealed two discrete groups of genomic regions (Fig. 1H). Low-accessibility regions tend to contain minimal mCH, whereas more-accessible regions of the genome show a proportional relationship between genome accessibility and mCH levels. Thus, although mCG is unaffected in these regions, lower chromatin accessibility appears to be highly inhibitory to deposition of mCH and hmC, potentially via inaccessibility to de novo methyltransferases and Tet mC hydroxylases. Furthermore, this indicates that accumulation of mCH and hmC during mammalian brain development occurs via processes that are at least partly independent from methylation at CG dinucleotides.

Fig. 3. mCH is enriched in genes that escape X inactivation. (A) Browser representation showing mCH-hypermethylated female human and mouse genes that escape X inactivation (shaded genes). (B) Box and whisker plots of gender differences in promoter mCG and intragenic mCH in inactivated and escapee genes on human chrX. (C) Scatter plot of gender differences in mCG and mCH in human chrX genes. Reported X inactivated and escapee genes: *Carrel and Willard (49); †Sharp et al. (50); ‡predicted escapee genes, and autosomal (Chr2) genes are indicated. (D) Discriminability analysis of genes that escape female X inactivation using mC data, showing correct versus false detection rate mapped for all possible mC/C thresholds.
Cell Type–Specific DNA Methylation Patterns in Neurons and Glia

The diversity of neuronal and glial cells in the frontal cortex raises the question of which features of DNA methylation are found in specific cell types. We isolated populations of nuclei by fluorescence-activated cell sorting that were highly enriched for neurons (NeuN+) or glia (NeuN–) from human and mouse adult frontal cortex tissue. An additional glial population was isolated from mice expressing enhanced green fluorescent protein (eGFP) under the S100b promoter. MethylC-Seq revealed differences in the composition and patterning of mCG and mCH in neurons and glia (Fig. 2A). Whereas differential mCG between neurons and glia was restricted to localized regions, neurons were globally enriched for mCH compared with glia. Indeed, we discovered that the level of mCH in glia is similar to that of fetal and early postnatal cortical tissue, whereas adult neurons have the greatest frequency of mCH that has been observed in mammalian cells. This indicates that the rapid developmental increase in mammalian brain mCH that coincides with the period of synaptogenesis is primarily due to mCH accumulation in neurons. Furthermore, our data show that in human neurons mCH is the dominant form of methylation in the genome: It is more abundant than mCG and occurs in 5% of CH and 10% of CA sites (Fig. 2B and fig. S1, A, B, and H). Of the total methylated fraction of adult human neuronal genomes, mCH accounts for ~53%, whereas mCG constitutes ~47%.

Although sparse in glia, mCH enrichment occurs within genes that are CH-hypomethylated in neurons, such as Mef2c (Fig. 2A), a transcriptional activator that plays critical roles in learning and memory, neuronal differentiation (42).
synaptic plasticity (43), and regulation of synapse number and function (44). Genome-wide surveys identified 174 mouse genes in which glia were hypermethylated relative to neurons in the CH context (table S3). Unbiased gene ontology analysis revealed that these glial hyper-mCH genes are highly enriched for roles in neuronal and synaptic development and function (table S3). These genes also overlapped significantly with a set of 461 genes expressed at higher levels in neurons than in astrocytes (13-fold higher overlap than chance, P < 10^{-30}, Fisher exact test, FET) (45) and 233 developmentally up-regulated genes (7.5-fold, P < 10^{-7}, FET). These genes show hypomethylation of CG and CH in neurons and hypermethylation of CH in glia (fig. S3A), consistent with a potential role of mCH in transcriptional repression of neuronal genes in the glial genome. Furthermore, genes associated with oligodendrocyte or epithelial function accumulate mCH through development (fig. S3B), with oligodendrocyte or epithelial function accumulating mCH, whereas epithelial genes display mCH hypermethylation in both neuronal and glial populations. Consistent with CH methylation requiring Dnmt3a, glial hyper-mCH genes frequently intersect areas of the genome bound by Dnmt3a in mouse postnatal neural stem cells (46). Dnmt3a-binding regions are greatly enriched for mCH, particularly in glia, whereas mCG is not enriched in Dnmt3a-binding regions in glia and is depleted in neurons (Fig. 2C). Thus, there is an association with Dnmt3a binding sites specific to mCH and not mCG, suggesting partial independence between these two marks.

mCH Position Is Highly Conserved

We examined whether the position of DNA methylation is stochastic or precisely controlled at different genomic scales. The level of mCH in 10-kb windows throughout the genome was highly reproducible between independent samples of the same cell type, with lower, but substantial, correlation between cell types (Fig. 2D). Closer inspection revealed consistency between the methylation level at individual mCH sites in neurons from different individuals in both mice and humans (Fig. 2E). At single-base resolution (fig. S4), perfect correlation between individuals would not be observed even if the true methylation level was identical at each site because of the stochastic effect of a finite number of sequenced reads. To correct for this, we normalized the observed correlation by that of simulated data sets with the same coverage per site as each of our experimental samples but with identical methylation levels (Fig. 2F and fig. S4). To assess statistical significance, we used a permutation test, which compared the data correlation with the correlation after randomly shuffling the relative positions of CH sites in each sample (fig. S4). This revealed that autosomal CG and CH sites have nearly identical methylation levels in neuronal populations isolated from different individuals of the same species. Observed differences could be explained by stochastic sampling rather than true individual variation. Unexpectedly, normalized per-site correlation is higher for mCH than mCG between neuronal populations isolated from the frontal cortices of different human individuals, and mouse neuronal mCG and mCH per-site
correlations are equivalent. Per-site correlation between two human ES cell lines (H1 and HUES6) is also high (>0.8) for both mCG and mCH.

The high interindividual correlation of mCH at the kilobase and single-site scales indicates that methylation of CH positions, particularly in mammalian neurons, is a highly controlled process. It is not consistent with a stochastic event that takes place at any available CH position in a particular genomic region that accumulates mCH. Comparison of mCH between human and mouse neurons at conserved exonic CH positions revealed a low but significant interspecies correlation (Fig. 2F; P < 0.005, shuffle test), possibly indicating conservation of the cellular processes that precisely target or restrict mCH at these positions. Last, per-site mCG and mCH correlation between human ES cells and neurons is significantly lower, likely because of differences in the processes governing methylation of particular genomic features in the distinct cell types, for example, enrichment and depletion of mCH in highly transcribed genes in ES cells and neurons, respectively (Fig. 1, A and B).

The precise conservation of mCH position may be partly caused by the physical configuration of DNA within nucleosomes. Consistent with this, neuronal mCH patterns contain robust periodic components at the scale of nucleosome spacing (~170 base pairs (bp), fig. S5A) and the DNA helix coil length (~10.5 bp, fig. S5B). Such periodic components may arise from sequence-dependent constraints on mCH position, which would be the same in every neuronal cell. Alternatively, epigenetic heterogeneity within the population of NeuN+ nuclei in our sample may lead to stronger correlation for CH sites located on the same physical chromosome, compared with the correlation between the same locations on chromosomes from different cells. To test this, we measured the cross-correlation within individual reads, revealing a contribution of within-chromosome correlation to the periodic methylation pattern (fig. S5C).

Gender-Specific DNA Methylation Patterns on the X Chromosome

Interindividual correlation of mCG and mCH on chromosome X (ChrX) is frequently lower than on autosomes (Fig. 2F), prompting a closer analysis of ChrX mC patterns. ChrX mCG and mCH levels were generally lower in females compared with males, presumably because of the effect of ChrX inactivation (fig. S5, F and G) (47, 48). However, a subset of genes in both humans and mice have significantly greater intragenic mCH levels in females compared with males (Fig. 3A). Inspection of these genes revealed that most were previously found to escape inactivation in human females (X-escapees), displaying biallelic expression (49) and a reduction in promoter mCG hypermethylation, a DNA methylation signature of inactivated alleles (50). Quantification of human gender differences in neuronal DNA methylation for ChrX genes previously characterized as showing biallelic expression (49) revealed that females have reduced promoter mCG and a large increase in intragenic mCH but not intragenic mCG (Fig. 3, B and C, and fig. S5, D and E). The sequence composition of mCH is very similar in the whole genome, within autosomal genes bodies, and within X-chromosome inactivated and escape gene bodies (fig. S5H). Analysis of gender-specific methylation in additional human cell types revealed that female promoter mCG hypomethylation is observed at X-escapee genes in glia and human embryonic stem cells (fig. S6). Intragenic mCH hypermethylation of X-escapees was also observed in female glia, albeit to a lesser extent than in neurons, but was not present in ES cells. Thus, X-escapee mCH hypermethylation may be a feature that is specific to neuronal cell types. Although both promoter CG hypomethylation and intragenic CH hypermethylation provide significant information for discriminating X-escapees [Fig. 3D, discriminability index (area under the curve, AUC) = 0.75 and 0.78, respectively], combining both mCG and mCH measurements boosts discriminability (AUC = 0.88). By using this intragenic mCH hypermethylation signature, we identified seven new putative X-escapee genes (table S4). On the basis of these data, we hypothesize that intragenic CH hypermethylation in neurons may play a compensatory role in genes that fail to acquire repressive CG hypermethylation in the promoter region, restoring equal gene expression between male and female cells (51).

Distinct Genic DNA Methylation States Demarcate Functionally Relevant Gene Clusters

DNA methylation within promoter regions and in gene bodies is implicated in regulation of gene expression (22, 52), suggesting that the precisely conserved, cell type–specific DNA methylation patterns may be related to specific neuronal and glial cellular processes. We therefore used an unbiased approach to classify patterns of mCG and mCH within each annotated gene body and in flanking regions extending 100 kb up- or downstream. After normalizing the methylation pattern around each autosomal gene by the local baseline mCG or mCH level in each adult neuronal or glial sample, we combined these features into a large data matrix containing 4200 individual DNA methylation measurements for each gene [seven samples, two contexts (CG and CH), 300 1-kb bins within and around each gene]. Using principal component (PC) analysis, we extracted five methylation features (PCs) that together account for 46% of the total data set variance (fig. S7A). Gene sets with specific neuronal or astrocytic expression, as well as ChrX genes, segregate within PC space (fig. S6B). We then used k-means clustering to classify all genes into 15 clusters on the basis of their mCG and mCH patterns (Fig. 4A and fig. S8). Several dominant patterns of DNA methylation and transcript abundance and dynamics between developmental and cellular states are evident. A cluster of genes that progressively loses gene-body mCG and mCH through development contains constitutively highly expressed genes that are strongly enriched for neuronal function and depleted for astrocyte-specific roles (Fig. 4A, box 1). These genes show intragenic mCG enrichment in glia and depletion in neurons (box 2), indicating that glial gene body mCG resembles that of the neural precursor cells that predominate the fetal brain. This indicates that the loss of mCG in brain tissue during development is due to CG hypomethylation in mature neurons. These constitutively highly expressed genes enriched for neuronal function also show extensive intragenic mCH hypomethylation in neurons in contrast to glia (box 3), and they are enriched for hmtCG (box 4) as previously described (32, 33). Genes that are not as highly transcribed, but that are associated with neuronal function and are developmentally up-regulated, also show intragenic mCG and mCH hypomethylation in neurons but not glia (box 5). For these gene sets, mCG and mCH enrichment or depletion is precisely localized to transcribed regions, suggesting that this modification of genomic mC is tightly coupled to transcription. Notably, the bodies of constitutively high genes that are not enriched for neuronal function (box 6) do not show marked fetal/glial mCG enrichment or neuronal mCH depletion, indicating that this differential methylation is specific for genes enriched for neuronal function and not simply an association with particular levels of transcriptional activity. Genes associated with astrocyte function show an opposite pattern to genes associated with neuronal function: a progressive increase in intragenic mCG and mCH in frontal cortex tissue over development, neuronal mCG and mCH hypomethylation, and glial mCG and mCH hypomethylation (Fig. 4A, boxes 7 to 9). Last, genes with constitutively low expression do not show developmental or cell type–specific DNA methylation patterns (box 10), demonstrating that dynamic DNA methylation in genes is highly associated with differential transcriptional activity in mammalian brain development and neural cell specialization. Each of the gene clusters identified in our unbiased analysis was significantly enriched or depleted for cell type–specific function [neuronal or astrocytic genes (45)] or particular expression patterns (constitutively high or low expression, developmentally up- or down-regulated) (Fig. 4B). Profiling the median mCG and mCH of genes within each of these categories allows direct comparison of developmental and cell type–specific DNA methylation in mouse (Fig. 4C) and human (fig. S3C). This analysis recapitulates many of the conclusions of the unbiased clustering (Fig. 4A).

The inverse relationship observed between genic mCH level and transcriptional activity is consistent with a model whereby intragenic accumulation of mCH impedes transcriptional activity. Alternatively, the process of transcription could interfere with mCH de novo methylation or induce active mCH demethylation, although these are not consistent with the DNMT3A-dependent intragenic mCH in human embryonic stem cells that is positively correlated with gene expression.
Fig. 6. Developmental and cell type–specific differential mCG. (A) Heat map of absolute mCG level in CG-DMRs identified between neurons and glia and over development in mouse (left) and human (right). (B) Fraction of all CG-DMRs located in distinct genomic features in mouse. (C) Enrichment or depletion of distinct cell type–specific and developmental CG-DMRs sets within genomic features from (B). (D and E) Intersection of developmentally dynamic CG-DMRs and (D) DNaseI hypersensitive sites or (E) enhancers in mouse brain in thousands. (F) Browser representation of mouse developmentally dynamic CG-DMRs and quantification of local enrichment of chromatin modifications, genome accessibility, and mC. (G) TAB-Seq reads showing fetal-specific hmCG in the Fetal > Adult CG-DMR in mouse. (H) Proportion of mouse developmental CG-DMRs where mCG/CG is significantly increased or decreased in Tet2 knockout mice. (I) Distribution of mCG level difference between wild-type (WT) and Tet2 mutant at mouse CG-DMRs. Significantly different DMRs are indicated by coloration.
Overall, glial mCG and mCH patterns closely resemble those of the fetal and the early postnatal brain, indicating that DNA methylation in early mammalian brain developmental stages may be a default state that largely persists through to maturity in glial cells, whereas neuronal differentiation and maturation involve extensive re-configuration of the DNA methylome that is highly associated with cell type–specific changes in transcriptional activity.

**hmCG Is Enriched Within Active Genomic Regions in Fetal and Adult Mouse Brain**

Our base-resolution analysis of hmC using TAB-Seq revealed that intragenic and global hmCG levels are largely equivalent between chromosomes, whereas hmCG/CpG is 22% lower on the male ChrX, consistent with previous reports from enrichment-based detection of hmC (32, 33). Analysis of hmCG levels in different genomic regions revealed that, although adult hmCG/CpG is similar across transcriptional end sites and intragenic, DNase-hypersensitive (DHS), and enhancer regions, the fetal frontal cortex shows a relative enrichment of hmCG in DHS regions and enhancers, in particular enhancer regions that are unique to the fetal developmental stage (Fig. 5B).

The inverse pattern can be observed for adult mCG levels, which are lower in DHS regions and enhancers (Fig. S9, A and B), suggesting that regions of relatively high hmCG levels in the fetal brain show relatively low mCG levels in the adult brain. Analysis of intragenic hmCG enrichment relative to flanking genomic regions, for cell type–specific or developmentally dynamic gene sets (Fig. 5C), showed that neuronal and astrocyte gene bodies that are highly enriched with hmCG in adult are also highly enriched at the fetal stage. Thus, despite lower absolute levels of intragenic hmCG in the fetal stage, the adult patterns of hmCG enrichment at these cell type–specific genes are already forming in utero. Constitutively lowly expressed genes show intragenic depletion of hmCG, in contrast to constitutively highly transcribed genes, which show localized enrichment of hmCG throughout part or all of the gene body. Developmentally down-regulated genes show enrichment of hmCG in the fetal frontal cortex but not in adults, indicating that reduced transcription is accompanied by a loss of hmCG enrichment. Overall, transcriptional activity is associated with intragenic hmCG enrichment, as reported (33), with utero establishment of adult hmCG patterns for cell type–specific genes and loss of hmCG enrichment associated with developmentally coupled transcriptional down-regulation.

Measurement of mC and hmC in all genes in fetal and adult mouse frontal cortex indicated that both mCg and hmCg are depleted at promoters and in gene bodies of lowly expressed genes, whereas hmCg is enriched throughout the gene bodies of more highly transcribed genes (Fig. 5D). The most highly expressed genes in the adult frontal cortex show intragenic mCG hypomethylation (Figs. 4 and 5D) but still retain high intragenic hmCG. Ranking all genes by transcript abundance, it is evident that the highest mean intragenic hmCG levels, which occur in the most highly transcribed genes, correspond to hmCG/CpG ~0.25 and mCG/CpG ~0.5 (fig. S10D). Frontal cortex development is accompanied by increased enrichment of hmCG at intragenic regions that are already hyper-hydroxymethylated at the fetal stage (Fig. 5D and fig. S10), demonstrating that adult patterns of genic hmC are already evident in the immature fetal brain.

**CG Differentially Methylated Regions Enriched in Regulatory Regions**

Because differences in genic mCG were observed over development and between neuronal and glial cell populations (Fig. 4), we scanned the human and mouse methylomes to comprehensively identify CG differentially methylated regions (CG-DMRs) throughout the genome. CG-DMRs were identified between fetal and adult frontal cortex, neurons and glia, and combined into four sets: neuronal and glial hyper- and hypo-methylated CG-DMRs. In total, 267,799 human and 142,835 mouse CG-DMRs were identified (median lengths: for mouse, 473 bp; human, 533 bp), revealing several predominant dynamics in mCG during brain development and cellular specialization (Fig. 6A and fig. S11). Neuronal CG-DMRs are the most numerous in both mice and humans, because of the very distinct mCG patterns that emerge during neuronal differentiation and maturation. At these sites, CG methylation in adult neurons is distinct compared with those in glial and/or fetal and early postnatal development frontotemporal cortex tissue samples. Neuronal hypermethylated CG-DMRs also show mCH hypermethylation (fig. S11). In mouse, mCG/Cg with intragenic hypomethylated CG-DMRs declines to a stable level by 1 week after birth. In contrast, neuronal hypermethylated CG-DMRs do not begin to change until 1 week after birth, after which they accumulate mCg until 2 to 4 weeks of age. These data indicate that increases in neuronal mCg occur during synaptogenesis after most decreases in neuronal mCg have already occurred. Furthermore, we found that hydroxymethylation in the adult cortex is highest in CG-DMRs that show neuronal hypermethylation and is depleted from CG-DMRs that display neuronal hypomethylation (Fig. 6A). This suggests that hmCg may be most abundant in neurons, rather than glial cells, in the frontal cortex.

Analysis of the genomic features in which CG-DMRs are located revealed that although half are found within gene bodies, they are not common within promoters and transcriptional start and end regions. Instead, they are disproportionately located at DHS regions and enhancers unique to fetal or adult brain (Fig. 6B). Closer inspection of the enrichment and depletion of these CG-DMRs revealed that fetal enhancers and DHS sites unique to the fetal brain are enriched for hypermethylation in adult brain but not in the fetal brain (Fig. 6C). In contrast, adult enhancers and unique adult DHS sites are highly associated with CG hypermethylation in fetal stages but are not associated with hypermethylated CG-DMRs in the adult brain and in neurons. Thus, developmentally dynamic enhancers and DHS sites in frontal cortex have dynamic CG methylation that is depleted where chromatin accessibility and regulatory element activity increase, consistent with a range of human cell lines (53).

To characterize gene functions associated with the CG-DMRs, we analyzed the association between proximal genes (transcriptional start site within 5 kb of the DMR) and cell type–specific or developmentally dynamic gene sets (fig. S12A). We observed an inverse relationship between methylation and gene function. Genes associated with neuronal function and up-regulation during development are enriched for promoter hypermethylation in glia and hypomethylation in neurons, whereas genes down-regulated during brain development and those related to astrocyte function are enriched for promoter hypomethylation in neurons and hypomethylation in glia. Genes that are constitutively expressed at either high or low levels are not associated with promoter/transcription start site CG-DMRs, indicating that dynamic CG methylation is highly associated with changes in transcriptional activity and cell type–specific transcriptional regulation.

Because the majority of all developmentally dynamic CG-DMRs are associated with DHS sites, we examined the directional relationships between dynamic mCG and DNA accessibility states over development (Fig. 6D). Notably, DNS sites unique to fetal frontal cortex overlap with 28% of CG-DMRs that gain methylation through development (Adult>Fetal). However, these sites only overlap 7.3% of CG-DMRs that lose mCg during development (Fetal>Adult). Similarly, DNS sites unique to adult frontal cortex rarely overlap Adult>Fetal CG-DMRs. A similar analysis of developmentally dynamic enhancers active in only one of the developmental stages (37) (Fig. 6E) showed that enhancer activation is associated with mCg hypomethylation of the enhancer, whereas enhancer inactivation is associated with enhancer mCg hypermethylation.

This inverse relationship between genome accessibility and mCg level at putative functional regions of the genome suggests that nuclear factors that bind the region and increase accessibility may cause localized reduction in mCg, as previously reported for a small number of DNA binding proteins (54). Alternatively, mCg hypermethylation may cause reduced genome accessibility by direct inhibition of DNA–protein interactions or induction of chromatin compaction, with loss of mCg enabling increased chromatin accessibility and genome interaction with DNA binding factors.

Discrete regions that show increased or decreased CG methylation through development are associated with specific local chromatin modifications. We found that CG-methylated regions of the fetal frontal cortex that become hypomethylated in the adult (Fig. 6F, Fetal>Adult) gain
Tet2 of genomic DNA from frontal cortex tissue of adult mice to hmC should be necessary for mCG hypomethylation in Tet2−/− relative to wild type (6 weeks, 10 weeks, and 22 months), >fourfold more numerous than hypomethylated CG-DMRs (3099). This further indicates a role for Tet2 in mediating mCG demethylation during brain development.

**A Hydroxymethylation Signature of Developmentally Activated Regions**

Adult→Fetal CG-DMRs show broad low-level hmCG enrichment flanking the CG-DMRs and a localized depletion of hmC at the center in both fetal and adult genomes (Fig. 6F), and the absolute abundance of hmC is several-fold lower in fetal compared with adult frontal cortex (Fig. 1E). This suggests that although Adult→Fetal CG-DMRs gain mCG through development, they tend to be refractory to conversion to hmC, potentially because of lower accessibility to the Tet hydroxylases. In contrast, Fetal→Adult CG-DMRs have a local enrichment of both mCG and hmCG in the fetal cortex that becomes a local depletion in the adult. Two enrichment-based genome-wide hmC profiling techniques, CMS-IP (38) and bioinformatic analysis (31), confirmed the localized enrichment of hmC at Fetal→Adult CG-DMRs (fig. S12B). The localized enrichment of hmC at these inaccessible and quiescent genomic regions, which lose mCG and hmCG later in development, indicates that they may be premodified with hmCG in the fetal stage to create a dormant state that is poised for subsequent demethylation and activation at a later developmental stage. Closer inspection of base-resolution hmC data revealed that 4% of the hmC bases that have significantly higher hmC levels in fetal compared with adult [false discovery rate (FDR) 0.05] directly overlap with Fetal→Adult CG-DMRs, far exceeding the number expected by chance (0.5%). This indicates that despite lower global levels of hmC in the fetal brain, developmentally demethylated CG-DMRs are enriched for hmC bases that are more highly hydroxymethylated in fetal than in adult brain (Fig. 6G). The localized enrichment of mCG at these CG-DMRs in the fetal cortex indicates that CG-demethylation has not yet taken place.

If fetal hmCG is poised at dormant genomic regions in order to facilitate active DNA demethylation at later developmental stages, then the Tet hydroxylase enzymes that catalyze conversion of mC to hmC should be necessary for mCG hypomethylation in the adult frontal cortex at these regions. To test this, we performed MethylC-Seq of genomic DNA from frontal cortex tissue of adult Tet2−/− mice. Adult→Fetal CG-DMRs, which gain mCG through development, are largely unaffected in Tet2−/− compared with wild-type adult mice (Fig. 6, H and I; 3.6% hypermethylated, FET, FDR 0.05). By contrast, a substantial fraction of Fetal→Adult CG-DMRs are hypermethylated in Tet2−/− (19.7%) versus wild type. The mutant shows a small but significant increase in mCG at Fetal→Adult CG-DMRs (Fig. 6I and fig. S12C) (Tet2−/−: 7.9% ± 4.6%, P < 10−11, Wilcoxon signed rank test). The partial effect of the mutation on CG methylation is not unexpected given that all three Tet genes are expressed in the brain (fig. S9C) and may exhibit some functional redundancy. Additionally, a genome-wide search identified 14,340 CG-DMRs hypermethylated in Tet2−/− relative to wild type (6 weeks, 10 weeks, and 22 months), >fourfold more numerous than hypomethylated CG-DMRs (3099). This further indicates a role for Tet2 in mediating mCG demethylation during brain development.

**Discussion**

The essential role of frontal cortex in behavior and cognition requires the coordinated interaction, via electrical and chemical signaling, of multiple neuronal cell types and a diverse population of glial cells. Individual brain cells have unique roles within circuits that are defined by their location and pattern of connections as well as by their molecular identity. The development and maturation of the brain’s physical structure and the refinement of the molecular identities of neurons and glial cells occur in parallel in a finely orchestrated process that starts early during the embryonic period and continues, in humans, well into the third decade of life (55, 56). An early postnatal burst of synaptogenesis is followed by activity-dependent pruning of excess synapses during adolescence (28, 57, 58). This process forms the basis for experience-dependent plasticity and learning in children and young adults (59), and its disruption leads to behavioral alterations and neuropsychiatric disorders (60). During this period, profound transcriptional changes lead to the appearance of adult electrophysiological characteristics in neocortical neurons.

Our study suggests a key role of DNA methylation in brain development and function. First, CH methylation accumulates significantly in neurons through early childhood and adolescence, becoming the dominant form of DNA methylation in mature human neurons. This shows that the period of synaptogenesis, during which the neural circuit matures, is accompanied by a parallel process of large-scale reconfiguration of the neuronal epigenome. Indeed, central nervous system deletion of Dnmt3a during late gestation induces motor deficits, and animals die prematurely (61). However, mice with a postnatal deletion restricted to the pyramidal cell population (complete recombination around 1 month old) do not show overt behavioral or transcriptional alterations (2). Our data suggest that expression of Dnmt3a specifically around the second postnatal week may be critical for establishing a normal brain DNA methylation profile and allowing healthy brain development.

Second, the precise positioning of mCG and mCH marks, which are conserved between individuals and across humans and mice, is consistent with a functional role. Whether this is the case, or whether the conserved patterns are instead a reflection of conserved nucleosome position or chromatin structure, requires further investigation. Third, the relationship between DNA methylation patterns and the function of neuron- or astrocyte-specific gene sets suggests a role for DNA methylation in distinguishing these two broad classes of cortical cells. DNA methylation could therefore play a key role in sculpting more-specific cellular identities. If this is the case, we expect that purified subpopulations will reveal high specificity of methylation at specific sites for particular cell types. Thus, the observation that most CH sites with nonzero methylation are methylated in >20 to 25% of sampled cells (fig. S1H) could be explained by the heterogeneity of these brain circuits rather than by stochastic methylation within each cell. These conclusions obtained from our genome-wide, base-resolution, cell-type-specific DNA methylation maps for brain cells through key stages of development are the first steps toward unraveling the genetic program and experience-dependent epigenetic modifications leading to a fully differentiated neuronal system.

### References and Notes


11. A. M. M. Oliveira, T. J. Hemsdottir, H. Bading, Rescue of aging-associated decline in Dnmt3a expression restores...


13. N. W. Penn, R. Suwalski, C. O. Hughes Medical Institute to T.J. and J.R.E.; the Gordon and Betty Moore Foundation (GBMF3034) to J.R.; NIH grant HG006827 to C.H.; NIH RO1 grants AI44432, HD065812, and 93730; Biobank, which is part of the BrainNet Europe Bank funded by the European Union’s Health research framework programme (HEALTH-F2-2008-200001) and the European Commission (LSHM-CT-2004-503039). We thank D. Chambers for technical assistance with animal breeding; D. Chambers for technical reading of this manuscript; J. Chambers for technical assistance; W. F. Loomis for critical reading of this manuscript; H. T. Chugani, A critical period of brain development: Relevance for the pathophysiology of schizophrenia and neurodevelopmental hypothesis. Schizophr. Bull. 37, 514–523 (2011). doi: 10.1093/schbul/sbr034; 21505118.


TRP 6187-12 from the Leukemia and Lymphoma Society (to A.R.). R.L. was supported by an Australian Research Council Future Fellowship. Support was also provided by the Government of Western Australia through funding for the Western Australia CoE for Computational Systems Biology. E.A.M. was supported by the Center for Theoretical Biological Physics, University of California San Diego, and NIH (National Institute of Neurological Diseases and Stroke grant K99NS080911). W.A.P. was supported by a predoctoral graduate research fellowship from the NSF. Analyzed data sets can be accessed at http://neomorph.salk.edu/brain_methylomes. Sequence data can be downloaded from National Center for Biotechnology Information GEO (GSE47966). Tet2 mutant mice are available under a material transfer agreement from the La Jolla Institute for Allergy and Immunology. Correspondence and requests for materials should be addressed to R.L. (ryan.lister@uwa.edu.au), M.M.B. (mbehrens@salk.edu), and J.R.E. (ecker@salk.edu).

Supplementary Materials
www.sciencemag.org/content/341/6146/1237905/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S5
References (63–78)

15 March 2013; accepted 7 June 2013
10.1126/science.1237905
Supplementary Materials for

Global Epigenomic Reconfiguration During Mammalian Brain Development

Ryan Lister,* Eran A. Mukamel, Joseph R. Nery, Mark Urich, Clare A. Puddifoot, Nicholas D. Johnson, Jacinta Lucero, Yun Huang, Andrew J. Dwork, Matthew D. Schultz, Miao Yu, Julian Tonti-Filippini, Holger Heyn, Shijun Hu, Joseph C. Wu, Anjana Rao, Manel Esteller, Chuan He, Fatemeh G. Haghighi, Terrence J. Sejnowski, M. Margarita Behrens,* Joseph R. Ecker*

*Corresponding author. E-mail: ryan.lister@uwa.edu.au (R.L.); mbehrens@salk.edu (M.M.B.); ecker@salk.edu (J.R.E.)

Published 4 July 2013 on Science Express
DOI: 10.1126/science.1237905

This PDF file includes:

Materials and Methods
Figs. S1 to S12
Table S1 to S5 captions
References

Other Supplementary Material for this manuscript includes the following:
available at www.sciencemag.org/cgi/content/full/science.1237905/DC1

Tables S1 to S5
Materials and Methods

Animals
The mouse lines C57Bl/6 and S100b-eGFP (B6;D2-Tg(S100B-EGFP)1Wjt/J), both from Jackson Laboratories, ME were bred and maintained in our animal facility in 12 h light/dark cycles with food ad libitum. To produce the S100b-eGFP animals used for FACS, animals were crossed to C57BL/6 to produce heterozygotes for eGFP expression. Tet2-/ mice were produced by targeted disruption of the Tet2 gene as described previously (63). Animals were weaned in groups of 3-4 per cage, and used at the postnatal time-points indicated. Three animals obtained from separate litters were processed together for DNA and RNA isolations. All protocols were approved by the Salk Institute’s Institutional Animal Care and Use Committee (IACUC).

Human samples
Human brain tissue was obtained from the NICHD Brain and Tissue Bank for Developmental disorders at the University of Maryland, Baltimore, MD, the IDIBELL Biobank (Barcelona, Spain), which is part of the BrainNet Europe Bank (Munich, Germany) funded by the European Commission (LSHM-CT-2004-503039). Procedures were approved by the Institutional Review Board of the Salk Institute.

Tissue production
Animals were anesthetized with isoflurane and decapitated. Brains were quickly extracted and placed in dissection media (64). The frontal cortex was obtained by slicing the adult brains coronally at Bregma 1 mm, and dissecting the frontal cortical tissue, being careful to avoid contamination with olfactory or putamen regions, under a dissection microscope. Meninges were then dissected out from the frontal cortex, which was rapidly frozen on dry ice or minced and processed for nuclei isolation and FACS. Brain dissections in younger animals occurred at a similar coordinate (i.e., when the genu of the corpus callosum unites between the two hemispheres). For the embryonic cortex, the front third of the cortical plate was dissected under a microscope.

Protein isolation and Western blot
Tissue dissected as above was disrupted in RIPA buffer (Hepes 50 mM, pH 8.0; 1% NP-40; 0.7% Deoxycholate; 0.5 M LiCl; complete protease inhibitors) with the addition 0.5% SDS, by incubation on ice for 30 min followed by centrifugation at 15,000 x g to separate nucleic acids. Protein was determined using a commercial kit (BCA, Pierce, Rockford, IL)). Fifty micrograms of protein for each timepoint were separated in 7.5% SDS-PAGE gels and transferred onto nitrocellulose membranes. Dnmt3a was detected by incubation with anti-Dnmt3a antibody (1:300, Imgenex Cat# IMG-268A, San Diego, CA) followed by HRP conjugated secondary antibodies and chemiluminescence (Cell Signaling, Danvers, MA) in four independent experiments.

Nuclei isolation
**Mouse:** Frontal cortices from four animals for each time-point were processed as described (65) with the following modification: isolated nuclei from 4 animals were incubated for 1 h at 4°C in 1 ml of PBS + 2% horse serum containing a 1:1000 dilution of AlexaFluor488 conjugated anti-NeuN monoclonal antibody (Millipore Cat# MAB377X, Temecula, CA) before FACS.

**Human:** production of nuclei followed a similar procedure as for mice. Freshly frozen samples of middle frontal gyrus were obtained from the NICHD Brain and Tissue Bank for Developmental disorders at the University of Maryland. Sample demographics are described in Table S5.

**Dissociation of frontal cortex tissue for FACS**
Frontal cortices from S100b-eGFP mice at 7-8 week old, obtained as above were minced and placed in oxygen bubbled dissection media containing 1 mg/ml freshly prepared pronase. After incubation at 37°C for 30 min, the tissue was washed thoroughly with the same media, resuspended in 5 ml PBS containing 1 % horse serum and gently triturated using a 5 ml pipette. The solution was then filtered through a 70 µm nylon mesh and subjected to FACS.

**FACS**
Cells (S100b-eGFP) or nuclei (NeuN) were sorted using a FacsVantage SE DiVacell sorter (BD Biosciences, San Jose, CA) using an 80 µm tip. Singlet cells were gated based on forward scatter pulse height and pulse width characteristics. eGFP positive cells or AlexaFluor488 labeled nuclei were discriminated from autofluorescence by plotting green fluorescence versus orange fluorescence, and bright expressors were chosen for sorting. Sorted cells or nuclei were centrifuged and pellets rapidly frozen in dry ice until processed for nucleic acid isolation.

**MethylC-Seq**
MethylC-Seq library generation, read mapping, processing, and analysis were performed as described previously (25), aligning reads to the mouse mm9 and human hg18 reference genomes, except a previous filter that excluded reads containing >3 cytosine bases in the CH context was not applied in this study. Library amplification was performed with either PfuTurboCx Hotstart DNA polymerase (Agilent Technologies, Santa Clara, CA) or KAPA HiFi HotStart Uracil+ DNA polymerase (Kapa Biosystems, Woburn, MA).

**mRNA-Seq**
mRNA-Seq libraries were generated from total RNA with polyA+ selection of mRNA using the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA). Strand-specific libraries were constructed using a dUTP methodology as described previously (66), while non-strand-specific libraries were constructed with the TruSeq RNA Sample prep Kit vs as per manufacturer’s instructions. TopHat2 and Cufflinks2 packages (67, 68) were used to map sequence reads to the mouse mm9 and human hg18 reference genomes and quantitate differential gene expression.

**Enrichment-based detection of 5-hydroxymethylcytosine (hmC)**
Cytosine 5-methylenesulphonate immunoprecipitation (CMS-IP): mouse E13 fetal cortex or 10 wk frontal cortex genomic DNA fragments (100 - 200 bp) were first end repaired, 3’ adenylated, ligated to methylated Illumina adaptor oligonucleotides and bisulfite converted as per the MethylC-Seq protocol (25). Bisulfite conversion of hmC to cytosine 5-methylenesulphonate (CMS) was followed by immunoprecipitation of gDNA fragments containing CMS with a specific antiserum to CMS. Bisulfite converted gDNA was first denatured for 10 minutes at 95°C (0.4 M NaOH, 10 mM EDTA), neutralized by addition of cold 2 M Ammonium Acetate pH 7.0, incubated with anti-CMS antiserum in 1x IP buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% Triton X-100) for 2 h to overnight at 4°C, and then precipitated with Protein G beads. Precipitated DNA was washed 3 times with 1x IP buffer and eluted with Proteinase K, then purified with Phenol Chloroform. Following immunoprecipitation, sequencing libraries were amplified by 8 cycles of PCR and DNA sequencing was performed using a HiSeq 2000 Genome Analyzer (Illumina) as described previously (38).

Biotin-glucosyl tagging and enrichment of hmC (biotin-gmC): the Hydroxymethyl Collector Kit (Active Motif, Carlsbad, CA) was used to selectively tag hmC bases within mouse E13 fetal cortex or 6 wk frontal cortex genomic DNA fragments (100 - 200 bp) with glucose and biotin moieties to form biotin-N3-5-gmC (biotin-gmC), as described previously (31). Genomic DNA fragments containing biotin-gmC were then enriched and purified by high-affinity capture using streptavidin magnetic beads. Enriched DNA containing biotin-gmC was used to generate Illumina DNA sequencing libraries with the TruSeq DNA Sample Prep Kit v2 and sequenced using a HiSeq 2000 Genome Analyzer (Illumina).

TAB-Seq
Performed as described in Yu et al. 2012 (34). Conversion and protection control sequences, read filtering, and quantitation of hmC protection and mC non-conversion are detailed below (Correction for non-conversion and protection).

Data Analysis Methods

Bisulfite non-conversion rate estimation for MethylC-Seq
To estimate the bisulfite non-conversion frequency, the frequency of all cytosine basecalls at reference cytosine positions in the lambda genome (unmethylated spike in control) was normalized by the total number of base-calls at reference positions in the lambda genome. This was performed for all cytosines as well as specific sequence contexts (CG, CH, CA, CC, CT, see Table S1).

Identification of CG Differentially Methylated Regions (CG-DMRs)
To identify DMRs, we used a two-step process. The first step involved performing a root mean square test, as outlined previously (69), on each individual CG. For this test, we constructed a contingency table where the rows indicated a particular sample and the columns indicated the number of reads that supported a methylated cytosine or an unmethylated cytosine at this position in a given sample. The p-values were simulated.
using 5000 permutations. For each permutation, a new contingency table was generated by randomly assigning reads to cells with a probability equal to the product of the row marginal and column marginal divided by the total number of reads squared. To speed up this process, if a p-value returned 50 permutations with a statistic greater than or equal to the original test statistic, we stopped running permutations, thus using adaptive permutation testing. To determine a p-value cutoff that would control the false discovery rate (FDR) at our desired rate (see details for each DMR analysis, below), we used the procedure outlined previously (70). Briefly, this method first generates a histogram of the p-values and calculates the expected number of p-values to fall in a particular bin under the null. This expected count is computed by multiplying the width of the bin by the current estimate for the number of true null hypotheses (m0), which is initialized to the number of tests performed. It then looks for the first bin where the expected number of p-values is greater than or equal to the observed value, starting from the most significant bin and working its way towards the least significant. The differences between the expected and observed counts in all the bins up to this point are summed, and a new estimate of m0 is generated by subtracting this sum from the current total number of tests.

This procedure was iterated until convergence, which we defined as a change in the m0 estimate less than or equal to the FDR. With this m0 estimate, we were able to estimate the FDR of a given p-value by multiplying the p-value by the m0 estimate (the expected number of positives at that cutoff under the null hypothesis) and dividing that product by the total number of significant tests we detected at that p-value cutoff. We chose the largest p-value cutoff that still satisfied our FDR requirement. Once this p-value cutoff was chosen, for mCG-DMR analysis, significant sites were combined into blocks if they were within 500 bases of one another and had methylation changes in the same direction, for example if at both sites sample A was hypermethylated and sample B was hypomethylated. Furthermore, blocks that contained fewer than 4 differentially methylated sites were discarded. Finally, CG-DMR blocks were filtered based on a requirement for a minimum number of samples to all show the same significant differential methylation patterns. These sample comparison details for the CG-DMR sets described in this study are described below.

- **Human fetal vs adult frontal cortex CG-DMRs** were identified (FDR = 0.05) where fetal frontal cortex was differentially methylated compared to both 16 yr and 25 yr frontal cortex samples.

- **Human NeuN+/NeuN- vs fetal frontal cortex CG-DMRs** were identified (FDR = 0.05) between NeuN+/NeuN- samples and fetal frontal cortex, requiring both human NeuN+/NeuN- samples (53 yr female R1, 55 yr male R2) to be differentially methylated compared to fetal.

- **Human NeuN+ vs NeuN- CG-DMRs** were identified (FDR = 0.05) between NeuN+ and NeuN- samples, requiring both human individuals (53 yr female, 55 yr male) to be differentially methylated between cell types.

- **Mouse fetal vs adult frontal cortex CG-DMRs** were identified (FDR = 0.01) where fetal frontal cortex was differentially methylated compared to both 6 wk and 10 wk frontal cortex samples.

- **Mouse NeuN+/NeuN- vs fetal frontal cortex CG-DMRs** were identified (FDR = 0.01) between NeuN+/NeuN- samples and fetal frontal cortex, requiring ≥2 of 3 mouse
NeuN+/NeuN- samples (7 wk male R1, 6 wk female R2, 12 mo female R3) to be differentially methylated compared to fetal.

- **Mouse NeuN+ vs NeuN- CG-DMRs** were identified (FDR = 0.01) between NeuN+ and NeuN- samples, requiring $\geq 2$ of 3 mouse samples (7 wk male R1, 6 wk female R2, 12 mo female R3) to be differentially methylated between cell types.

- **Mouse Tet2$^{-/-}$ vs wild-type CG-DMRs** were identified (FDR = 0.01) where Tet2$^{-/-}$ was differentially methylated compared to $\geq 2$ of the 6 wk, 10 wk and 22 mo wild-type adult frontal cortex samples.

- **Mouse fetal hmCG vs 6 wk hmCG CG-DMRs** were identified (FDR = 0.05) as above except individual significantly differentially hydroxymethylated bases were identified and no block joining was performed.

**DNaseI hypersensitivity**

Data generated by the Mouse ENCODE Consortium (35) was mapped to the mouse mm9 reference genome with Bowtie (71) and peaks of read enrichment were identified using MACS (72).

**ChIP-Seq and enhancers**

Mouse E14.5 embryonic brain predicted enhancer regions, adult cortex predicted enhancer regions, and sequence reads mapped to mm9 for H3K4me1, H3K27ac and input were from Shen et al. 2012 (37). Coordinates of predicted enhancer regions that overlapped with DNaseI hypersensitive regions were adjusted to center the enhancers on the midpoint of DNaseI hypersensitive sites.

**Dnmt3a binding sites**

Liftover was used to convert the coordinates of Dnmt3a binding sites identified in mouse postnatal neuronal stem cells (46) to the mm9 reference.

**IP detection of hmC**

CMS-IP reads were mapped to the mouse mm9 genome as described previously for MethylC-Seq sequencing data (25). Biotin-gmC reads were mapped to the mouse mm9 genome using Bowtie2 (71).

**Estimation of methylation level (mC/C) with correction for non-conversion and protection**

The majority of our analyses of DNA methylation are based on the level (mC/C, mCG/CG, or mCH/CH), which is an estimate of the fraction of cytosines in the sequenced population which are methylated. To estimate mC/C, we computed the fraction of all MethylC-Seq base-calls at cytosine reference positions that were cytosine (protected from bisulfite conversion). We then corrected these estimates for known experimental biases. Both the whole-genome bisulfite sequencing and TAB-Seq techniques used in this study are subject to a low error rate, $r$, due to the failure of the chemical conversion of unmethylated cytosine to uracil. Such errors leads to false positive detections of mC. In addition, TAB-Seq suffers from a small rate of error due to non-protection, $s$, which is the probability that a hydroxymethylated cytosine is converted
by the bisulfite reaction and leads to a false negative (missed call). These rates were calibrated as described previously (34), summarized as follows.

**TAB-Seq cytosine non-conversion and hmC protection rates:** M.SssI-converted pUC19 DNA (NEB), fully methylated (mC) at all CG cytosines, was spiked into the mouse genomic DNA sample at 0.5% (w/w) prior to library preparation (34). The frequency of C basecalls at pUC19 CG reference positions was calculated from reads uniquely mapped to the pUC19 reference, following removal of clonal reads and reads containing >3 cytosine basecalls in the CH context (34). The cytosine non-conversion rate represents the aggregate frequency of failed mC conversion by Tet1 and failure of bisulfite conversion.

**hmC non-protection rate:** The 38-48 kb region of the unmethylated lambda phage genome (Promega, Madison, WI) was amplified using ZymoTaq DNA polymerase in 5 separate, nonoverlapping 2 kb PCR amplicons in the presence of dATP, dGTP, dTTP and d5hmCTP (Zymo Research), as described previously (34). PCR products purified by agarose-gel electrophoresis were spiked into the mouse genomic DNA sample at 0.5% (w/w) prior to library preparation. The frequency of C basecalls at CG reference positions (protection frequency) was calculated from all reads uniquely mapped to the 38-48 kb region of the lambda genome, only considering CG reference positions where the closest neighboring reference C is at least 4 bases away (e.g. hmCGNNhmC), in order to reflect the subset of bases in the similar hmC context as mammalian genomes (34). Finally, given the commercial source of 5hmCCTP used to synthesize the control contains ~5% dCTP, which does not get protected, the protection frequency was increased by 5%. Thus, the non-protection frequency (1 - protection frequency) represents the rate at which hmC bases failed to be glucosylated and protected from Tet1 conversion and subsequent bisulfite conversion.

We corrected the observed number of converted and protected reads (a, c respectively) to take these rates into account. If the true rate of methylation or hydroxymethylation at a given site is p, the probability of observing (c, a) is Binom[c|a+c,q], where q = p(1-s) + (1-p)r is the probability of a read being sequenced as cytosine. The maximum likelihood estimate of p is therefore \( \hat{p} = \frac{q-r}{1-r-s} \). When \( \hat{p} < 0 \) we set \( \hat{p} = 0 \). In all of our samples, r, s \( \ll 1 \). For some analyses, we therefore used an approximation valid for low non-conversion rate, \( \hat{p} \approx q - r = \frac{c}{a+c} - r \).

**Supplementary Text**

**Figure-specific data analysis details:**

Fig. 1A, 2A, 2E and Fig. 3A (browser representations) The AnnoJ browser (24) shows tracks of hg18 and mm9 UCSC gene annotations, mRNA-Seq read density, and DNA methylation on the Watson and Crick strands (W, C). DNA methylation at each CG or CH site is represented as a vertical tick (green or blue, respectively) whose height corresponds to the fraction of all reads mapped to that site.
which were protected from bisulfite conversion (mC/C, range from 0 to 1). Regions which are homologous in human and mouse are indicated by light-blue shading.

Fig. 1B
CH methylation level (mCH/CH) for each gene as a function of the rank of mRNA transcript abundance (mRNA-Seq fragments per kilobase of mapped read (FPKM)). We analyzed all genes which are expressed (FPKM > 0.01), whose length > 1 kb, mCH/CH < 0.03, and we omitted genes with no MethylC-Seq coverage for CG or CH reference positions. Note: mCG/CG and mCH/CH are defined as the total number of mC basecalls in each context, normalized by the total number of basecalls in the same context (both methylated and unmethylated), corrected by subtraction of context specific non-conversion rates (see above).

Fig. 1E
We calculated the fraction of all cytosine basecalls in each context, considering only sites covered by at least 10 reads in the mouse MethylC-Seq and TAB-Seq samples (fetal and 6wk). The proportion of basecalls was corrected for each sample’s non-conversion rate as described above.

Fig. 1F
For all 100 kb contiguous windows of mouse and human chromosome 12, mCG/CG and mCH/CH were calculated, with subtraction of mCG or mCH non-conversion values.

Fig. 1G
For all 5 kb contiguous windows of the mouse genome in chr12:112,250,000 - 119,500,000, the following mouse data was summarised:
- mRNA-Seq (10 wk frontal cortex): fraction of all reads located in window.
- ChIP-input reads (8 wk cortex): fraction of all reads located in window.
- DnaseI HS (8 wk cortex): fraction of all reads located in window.
- mCG/CG and mCH/CH: fraction of basecalls in window that are methylated in each context.
- hmC CMS-IP (10 wk frontal cortex): [fraction of all CMS-IP reads located in window] / [fraction of all CMS-input reads located in window].
  - Note: mRNA, ChIP input and DnaseI HS tracks use arbitrary units proportional to read density.

Fig. 1H
Chromatin accessibility vs CH methylation level (mCH/CH). For all 10 kb contiguous windows of the mouse genome, we calculated mCH/CH level from 10 wk mouse frontal cortex MethylC-Seq and estimated chromatin accessibility as ChIP-Seq input read density (reads per 10 kb per million total reads) for 8 wk mouse cortex (37). We omitted any 10 kb regions where ChIP-seq input read density = 0 or mCH/CH = 0. Density of bins is shown for mCH/CH ≤ 0.02 and ChIP-seq input read density ≤ 0.5.

Fig. 2B: as per Fig. 1E. Only sites that were covered by at least one read in all samples were included, and correction for non-conversion was applied.
For each mouse NeuN+, NeuN- and glia sample, mCG/CG and mCH/CH were calculated for all Dnmt3a binding regions \( (46) \) (\( n = 67,050 \)) and a set of 100,000 regions randomly distributed throughout the genome, each the size of the mean length of Dnmt3a binding regions (555 bp). Data is presented in a box and whisker plot with outliers not shown.

Heatmaps show Pearson correlation of mCH/CH within bins of size 10kb throughout the human and mouse genome. The median total coverage in each bin (i.e., the number of CH sites times the number of reads overlapping that site) was >24,000 for all samples. This provides confidence that the estimated mCH/CH was not severely affected by stochastic sampling of reads. Finally, we applied hierarchical clustering to all the data sets with complete linkage and distance given by one minus the sample correlation.

We developed statistical procedures to analyze the base-resolution, or per-site, correlation of methylation level between samples within the same species (intra-species) and across species (inter-species comparison of mouse and human). For this analysis, we focused on samples with substantial mCH, including NeuN+ in mouse (R1, R2, R3) and human (R1, R2), as well as human stem cell lines (H1, HUES6).

**Intra-species correlation:** To reduce the effect of stochastic sampling, we set a minimum coverage threshold of 10 reads per site. Any site which had fewer than 10 reads in any of the samples was excluded. To account for slight differences in the bisulfite non-conversion between samples, we corrected the methylation level (mC/C) using the procedure described above. We then calculated the Pearson correlation of the corrected mC/C across all single sites between each pair of samples. These values defined the correlation values for “individual replicates,” shown as open circles in Fig. S4. In some cases more than two samples were available for comparison, so we computed a group average correlation coefficient by summing each of the pairwise covariances of each independent pair of samples, then normalizing by the appropriate sum of variances. For the three mouse NeuN+ replicates, let the three replicates be m1, m2 and m3. The average correlation, which determines the bar height in Fig. 2F and S4, is then:

\[
r_{ave} = \frac{(\sigma_{m1,m2} + \sigma_{m2,m3} + \sigma_{m3,m1})}{(\sigma_{m1,m1} + \sigma_{m2,m2} + \sigma_{m3,m3})}
\]

**Inter-species correlation:** To compare methylation patterns between mouse and human, we first identified homologous CH sites. We restricted analysis to CH sites of high coverage in all three mouse NeuN+ samples (≥20 reads). For each of these sites, we used the UCSC liftOver utility to determine whether a unique homologue site exists in the human (73). We then filtered these homologous sites to retain only those which were CH sites in both species, and which had a minimum coverage in all human NeuN+ or ESC samples (≥10 reads).

We identified 66,489 autosomal and 3,323 ChrX sites for comparison of human and mouse NeuN+; for comparison of mouse NeuN+ with human H1 ESC we used slightly
more (88,771 autosomal, 5,120 ChrX). The correlation coefficients for individual pairwise comparisons of replicates were then calculated as for the intra-species correlations (above).

For the human-mouse NeuN+ comparison, we combined 6 total comparisons (human replicates r1={h1,h2} vs. mouse replicates r2={m1,m2,m3}) to form an average correlation coefficient as follows:

\[
    r_{\text{ave}} = \frac{\left( \Sigma r_1 \Sigma r_2 \sigma_{r1,r2} \right)}{\left( \Sigma r_1 \Sigma r_2 \sigma_{r1,r1} \right) \left( \Sigma r_1 \Sigma r_2 \sigma_{r2,r2} \right)}
\]

This procedure was also used to compute an average correlation for comparison of the two embryonic stem cell lines with the human and mouse NeuN+ replicates.

To assess the magnitude and significance of the correlation values we observed, we carried out two control analyses for each of the above correlation measurements.

**Simulation of full correlation to assess correlation magnitude:** Because our estimate of methylation level at each site is based on the observed frequency of methylated reads (mC/C), this measurement suffers from stochastic noise arising from the random sampling of reads from the population of DNA fragments. Such noise lowers the correlation coefficient we observe between each pair of samples. To correct for this, we performed a simulation to determine how much the correlation coefficient is reduced by sampling noise. We did this by simulating pairs of samples with identical methylation patterns, but independent sampling noise. At each site, we combined the number of observed methylated and unmethylated reads from two experimental samples \((a_1, a_2, c_1, c_2, \text{respectively})\) and determined the total empirical frequency of methylated reads: \(f = (c_1 + c_2)/(a_1 + a_2 + c_1 + c_2)\). We then simulated random samples from a binomial distribution with probability of success \(f\) and total trials \((a_1 + c_1)\) or \((a_2 + c_2)\), respectively. Finally, we computed the correlation coefficient using these simulated samples. Fig. S4 shows that these simulated correlation coefficients were ~0.8 for most comparisons. Since the true correlation of the simulated samples, which would be observed in the limit of very high coverage, is 1, in Fig. 2F we normalized the observed correlation of the original data by the simulated correlation. Fig. S4 shows the correlation of the data, without normalization.

**Shuffle control to assess correlation significance:** A second control analysis was conducted to compare the observed correlation with what would be expected in case the true correlation was 0. The finite size of the data set leads to a non-zero observed correlation, particularly for inter-species comparisons that cannot take advantage of large populations of homologous sites. In addition, long length-scale correlations related to genomic regions such as CG-islands may induce correlation that is not specific to a particular site (Fig. S5A). For each comparison of a pair of samples, we performed a set of 200 shuffle controls in which the methylation level at a given site in sample 1 was compared with the methylation level at a randomly chosen site in sample 2, constrained to be <1 Mb distant. After shuffling sites, we computed the correlation coefficient exactly as for the unshuffled data.
Fig. 3: Analysis of gender-specific methylation patterns on X-chromosome

We compared the brain DNA methylation patterns in our human female and male NeuN+ samples with two published surveys of human X-linked genes. The first of these studies measured expression from the active and inactive ChrX alleles using a human/rat somatic cell hybrid assay (49). The authors of that study assigned each surveyed gene a score in the range 0-9 corresponding to the number of individual hybrid lines in which expression from the inactivated ChrX was detected. A score of 0 thus corresponds to complete inactivation, whereas 9 corresponds to complete escape from X-inactivation. We remapped these genes to the reference used in this study (hg18) using the UCSC liftOver utility (73). We also compared our data with a second survey that used an immunoprecipitation based method (MeDIP) to profile DNA methylation from peripheral blood in normal human females (karyotype 46, XX) as well as Turner syndrome patients (45, X) (50). That study reported that inactivated genes feature hypermethylation of promoter region CG islands, whereas escapee genes show low levels of promoter methylation. These two studies using complementary techniques each provide a list of putative X-inactivated and X-escapee genes, which we used to compare with the mCG and mCH patterns in our neuronal samples. To avoid noise and bias, we excluded short genes (<2 kb in length) as well as any gene with low coverage (<50% of the mean read density for the entire ChrX).

Fig. 3B and S4D

For each of the genes assayed in (49), we examined the total mCG/CG within the promoter region, defined to be a 2 kb region centered on the TSS, and the total mCG/CG or mCH/CH within the gene body. The box plot shows the difference between female and male methylation level for genes ranked according to the X-inactivation status index. For each box, the central line is the group median, the box edges are the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum values (including all “outliers”).

Fig. 3C and S5E

Scatter plots of human female vs. male NeuN+ mCG/CG within the promoter region (TSS±1kb) of each gene, or intragenic mCG/CG or mCH/CH.

Fig. 3D

Receiver operating characteristic (ROC) analysis was used to assess how well DNA methylation patterns allow discrimination of X-escapee genes. For this analysis, we used only those genes whose expression was measured by (49). We defined escapee genes as those which were expressed from both the active and inactivated ChrX in all of the cell lines tested (score = 9). Genes with score 0-8 were defined as non-escapee. We tested discriminability using the female-male difference in promoter region mCG/CG, which was previously shown to correlate with X-inactivation status (50), as well as the difference in neuronal intragenic mCH/CH. In addition, we tested a linear combination of four different features (female-male promoter region mCG/CG and mCH/CH, female-male intragenic mCG/CH and mCH/CH) with coefficients determined by linear regression against the X-inactivation score (49). For each of these three measures of
DNA methylation, we created an ROC curve by plotting the fraction of escapee genes that are correctly discriminated, as well as the fraction of false detections, at each value of a discrimination threshold. The area under the ROC curve (AUC) is a statistical measure of discriminability, which ranges from 0.5 when little or no discrimination information is present to 1 for perfect discriminability.

Fig. 4A, S7: Unbiased clustering of gene sets based on DNA methylation profiles
To identify sets of genes that share similar DNA methylation patterns in an unbiased fashion, we developed a two-stage procedure that first represents all annotated genes within a low dimensional feature space, and then groups nearby genes in this space to create gene clusters.

The starting point is a large data array with one entry for each gene in each of 7 samples (6 wk, 7 wk and 12 mo NeuN+ and NeuN-, and glia (S100b+)) and in each sequence context (CG and CH). For each entry, we profiled the methylation level (mCG/CG and mCH/CH) in bins of size 1 kb starting 100 kb upstream of the TSS and ending 100 kb downstream of the transcription end site (TES). To compare genes with different lengths, we divided each gene body into 10 non-overlapping bins of equal size extending from the TSS to the TES. We then linearly interpolated the gene-body mC/C data at 100 evenly spaced bins within the gene body in order to give roughly equal weight to the gene-body and flanking methylation data. The absolute mCG/CG and mCH/CH level in each sample for each gene was normalized by the median value in the distal flanking region (50-100 kb upstream of TSS or downstream of TES). Normalized mC/C values were then log-transformed. Combining all of the normalized methylation data points, we obtained a matrix of 4,200 features for each of 25,260 genes. Any bins with missing data due to insufficient coverage in one of the samples (5.9% of the total) were replaced with the median value of the entire data set. We performed singular value decomposition on this data matrix to identify the linear combinations of methylation features that account for the largest fraction of the total data variance [principal components (PCs), Fig. S7]. We retained the top 5 PCs as a low-dimensional representation of robust genomic methylation features, accounting for 46% of the total data variance. Although the remaining PCs contain a great deal of meaningful data which may be useful for more refined analysis of DNA methylation features, our analysis focused on this reduced space.

Next, we used k-means clustering to estimate gene sets with highly similar within-set methylation patterns. We chose to extract k=15 clusters to capture a diverse range of methylation features, while still allowing visualization and statistical enrichment analysis of functional association for each gene set. We repeated the clustering procedure 10 times using random initialization of the cluster centers, choosing as the final estimate the run with the smallest within-cluster sum of distances from each point to the cluster centroid. To display the heatmaps of mRNA-Seq and mC/C patterns for each of 25,260 genes, we smoothed and downsampled the genes 40-fold to allow representation of genome-scale features.

Fig. 4B: Enrichment analysis of functional gene categories
To relate DNA methylation profiles with cell-type specific gene functions and with specific patterns of developmental regulation, we defined six mouse gene sets, as follows:
1. Neuronal genes (461) which were reported to have >3-fold enrichment in mouse neurons (45). To increase the specificity of this list, we also required that genes were homologous to rat genes which are highly expressed in neuronal somata (74).

2. Astrocytic genes (2,618), reported to have >20-fold enriched expression in astrocytes (45).

3. Constitutively high genes (60) which did not appear in the neuronal or astrocytic gene set; whose transcripts were highly abundant (mRNA-Seq FPKM ≥ 10) in brain tissue samples from seven developmental stages (fetal, 1, 2, 4, 6 and 10 wk, and 22 mo); and which had ≤20% variation between the maximum and minimum abundance in these samples.

4. Constitutively low genes (5,278) which did not appear in the neuronal or astrocytic gene set; whose transcript abundance remained below a threshold (FPKM ≤ 0.5); and which had ≤20% variation between the maximum and minimum abundance in these samples.

5. Upregulated genes (233) which are either neuronal or astrocytic genes and which have >10-fold increased expression at 2 wk compared with fetal.

6. Down-regulated genes (207) which are either neuronal or astrocytic genes and which have >5-fold lower expression at 2 wk compared with fetal.

For each functional gene set, we computed the overlap with each of the 15 gene clusters identified by \( k \)-means (Fig. 4A). We compared this number with the expectation for random gene sets with the same size to compute a fold-enrichment value, and we assessed statistical significance using Fisher’s exact test. We applied a Benjamini-Hochberg procedure (75) to control the false discovery rate < 0.05.

**Fig. 4C**
For each gene set, we computed the median of the flank-normalized mC/C profiles in each developmental and cell-type specific sample. The mRNA-Seq expression data (right-hand column) show the median expression level over all genes in each set.

**Fig. 5A**
Total hmCG/CG and hmCH/CH are shown for autosomes and ChrX in 6 wk cortex. Methylation values have been corrected for the biases introduced by non-conversion and non-protection (see above).

**Fig. 5B**
We profiled the total hmCG/CG within the following regions: CG islands (UCSC), promoter regions (TSS±2kb), gene ends (TES±2kb), gene bodies (TSS-TES), DHS (35) identified in adult or fetal brain, DHS identified exclusively in adult or fetal, and enhancers identified in adult or fetal brain (37). The plot shows the median over all the individual regions of each type, and the error bars indicate the 32-68th percentile range corresponding to ~1 standard deviation.

**Fig. 5C**
Heatmaps show the median profile of flank-normalized hmCG/CG around six categories of genes. Methods are the same as Fig. 4C.
Fig. 5D
Heatmaps showing the profile of mC/C and hmC/C for each of 16,077 genes expressed in adult mouse cortex (mRNA-Seq FPKM > 0.1 in 6 wk tissue). Genes are ranked by their expression in adult cortex and the flank-normalized mC/C or hmC/C pattern around each gene is displayed, as described in the methods for Fig. 4A (above).

Fig. 6A
Starting with six distinct categories of CG-DMRs for each species (described above), we excluded those on the X and Y chromosomes and organized the remaining DMRs into 4 largely non-redundant sets comprising most of the specific CG-DMR types. These were defined as follows:

*NeuN+ hyper-mCG*: This set includes CG-DMRs where mCG/CG is larger in NeuN+ compared with NeuN– and Fetal; larger than Fetal only; or larger than NeuN– only

*NeuN– hyper-mCG*: This set includes CG-DMRs where mCG/CG is larger in NeuN– compared with NeuN+, and those in which mCG/CG is larger in NeuN– than in fetal.

*NeuN+ hypo-mCG*: This set includes CG-DMRs where mCG/CG is smaller in NeuN+ compared with NeuN– and Fetal; or smaller compared with Fetal only.

*NeuN– hypo-mCG*: This set includes CG-DMRs where mCG/CG is smaller in NeuN– compared with NeuN+ and Fetal, or Fetal only.

We also created two additional DMR sets (heatmaps not shown):

*Fetal hyper-mCG*: This set includes CG-DMRs where mCG/CG is larger in fetal compared with adult NeuN+ and/or NeuN–.

*Adult hyper-mCG*: This set includes CG-DMRs where mCG/CG is larger in adult NeuN+ and/or NeuN– compared with fetal.

To display heatmaps of the methylation level in each CG-DMR, we first sorted the DMRs within each set by the difference between NeuN+ and NeuN– mCG/CG, averaged over replicates. We then smoothed and down-sampled the ordered DMR methylation patterns by 10 DMRs. The line plots below each heatmap show the median methylation level for each CG-DMR category, while the shaded regions show the 32nd-68th percentiles.

Fig. 6B
To assess the genomic distribution of CG-DMRs, we computed the fraction of all CG-DMRs (the union of the six sets listed above) that overlap the genomic regions defined above (see Fig. 5B and associated methods).

Fig. 6C
We compared the number of CG-DMRs of each type overlapping each region, with the expected number that would be observed if both sets were randomly distributed. Suppose
there are \( N \) CG-DMRs and \( R \) regions, and let the total genomic length of the CG-DMRs be \( L_N \) and the total genomic length of the regions be \( L_R \). Then the expectation value of the number of overlapping regions is \( n_0 = (NL_R + RL_N)/G \), where \( G \) is the total length of the genome. In the limit \( L_N, L_R \ll G \), the number of observed overlapping regions will be Poisson distributed with mean \( n_0 \). To assess the significance of enrichment or depletion, we therefore compared the actual number of observed overlaps with the Poisson cumulative distribution function. No multiple-comparison correction was applied to this analysis.

**Fig. 6D,E**
For each category of developmental CG-DMRs (Fetal > Adult or Adult > Fetal), we calculated the fraction that overlap a fetal or adult DHS site or a fetal- or adult-specific enhancer. The Venn diagrams represent these overlaps, and the relative sizes and positions of the circles are chosen to approximate the true proportions as closely as possible. Exact proportionality of the Venn diagrams with the true overlaps is sometimes not geometrically feasible.

**Fig. 6F**
For each CG-DMR, data was analyzed in contiguous 100 bp bins from 2.5 kb upstream to 2.5 kb downstream of the central position of the CG-DMR. H3K4me1, H3K27ac and CMS-IP read density profiles were normalized to their respective input control samples. The DNaseI HS read density profile uses arbitrary units proportional to read density.

**Fig. 6H**
To test the hypothesis that Tet2 is involved in active demethylation during development of brain cells, we analyzed differences between mCG/CG in adult wild-type mice (6 wk) and \( \text{Tet2}^{-/-} \) knockout mice (63). At each developmental CG-DMR, we counted the total methylated and unmethylated reads within a window of size 1 kb centered on the midpoint of the DMR. We then used Fisher’s exact test to assess the significance of differences between the 6 wk wild-type sample and each of the other four samples. We applied a Benjamini-Krieger-Yekutieli procedure to control FDR\(<0.05 \) (MATLAB procedure fdr_bky by David M. Groppe) (76). The bars in Fig. 6H show the proportion of developmental CG-DMRs that had higher or lower methylation in the \( \text{Tet2}^{-/-} \) mice compared to the 6 wk wild-type.

**Fig. 6I**
Histograms show the distribution of the difference in mCG/CG within 1 kb windows centered on each developmental CG-DMR between adult WT (6 wk) and two groups: control (4 wk, 10 wk and 22 mo WT), as well as the \( \text{Tet2}^{-/-} \) knockout. Filled bars show the magnitude of differences at CG-DMRs that were significantly altered (see statistical procedure above, Fig. 6H).

**Fig. S1A-D**
The fraction of all basecalls methylated in each context was calculated as follows (where \( \text{CN}=\text{CC}, \text{CA}, \text{CT} \) or \( \text{CG} \)): 
Total mCG basecalls (corrected for CG BS non-conv.):  
\[ m_{\text{CG basecalls-corr}} = \sum m_{\text{CG basecalls}} - \sum m_{\text{CG BS-non-conv}} \]

Total mCA basecalls (corrected for CA BS non-conv.):  
\[ m_{\text{CA basecalls-corr}} = \sum m_{\text{CA basecalls}} - \sum m_{\text{CA BS-non-conv}} \]

Total mCC basecalls (corrected for CC BS non-conv.):  
\[ m_{\text{CC basecalls-corr}} = \sum m_{\text{CC basecalls}} - \sum m_{\text{CC BS-non-conv}} \]

Total mCT basecalls (corrected for CT BS non-conv.):  
\[ m_{\text{CT basecalls-corr}} = \sum m_{\text{CT basecalls}} - \sum m_{\text{CT BS-non-conv}} \]

Fraction of all C basecalls that are mCG:  
\[ m_{\text{CG fraction}} = \frac{m_{\text{CG basecalls-corr}}}{\sum m_{\text{CN basecalls}}} \]

Fraction of all C basecalls that are mCA:  
\[ m_{\text{CA fraction}} = \frac{m_{\text{CA basecalls-corr}}}{\sum m_{\text{CN basecalls}}} \]

Fraction of all C basecalls that are mCC:  
\[ m_{\text{CC fraction}} = \frac{m_{\text{CC basecalls-corr}}}{\sum m_{\text{CN basecalls}}} \]

Fraction of all C basecalls that are mCT:  
\[ m_{\text{CT fraction}} = \frac{m_{\text{CT basecalls-corr}}}{\sum m_{\text{CN basecalls}}} \]

Fraction of all C basecalls that are mCH:  
\[ m_{\text{CH fraction}} = m_{\text{CA fraction}} + m_{\text{CC fraction}} + m_{\text{CT fraction}} \]

Fig. S1G: Partially methylated domains were identified as described previously (22).

Fig. S2A,B  
Overall, before correcting for the rate of false positives introduced by bisulfite non-conversion and failure of Tet1 oxidation, we measured 0.5262% non-converted cytosines in TAB-Seq reads in the CH context in adult (6wk) mouse cortex. Our empirical control measurement of non-conversion in unmethylated pUC19 DNA in the CH context yielded a rate of 0.5245%, with a 99% confidence interval spanning 0.467 - 0.587% (based on fit of a binomial distribution using the Clopper-Pearson method; MATLAB function “binofit”). We thus estimate the global hmCH/CH to be 0.017%, with a 99% confidence interval 0 - 0.059%.

To test whether individual CH sites contain statistically significant hmCH, we developed a statistical model based on the null hypothesis that CH sites are composed of 1.4% mCH, 98.6% CH, and 0% hmCH. The non-conversion rate for unmethylated CH sites was assumed to be 0.52%, whereas the sum of bisulfite non-conversion and Tet1 non-oxidation for mC sites was set at 2.08% based on measurement of pUC19 DNA (Table S1). Our model assumes that the number of non-converted reads will follow a mixture of two binomial distributions with these two rates. Based on this model we calculated the p-value associated with the reads observed at each CH site across the genome. We detect only 50 significant sites [FDR=0.01, Benjamini-Hochberg-Yekutieli procedure (76)]. These sites were marginally significant, with hmCH/CH ~ 0.5.

Fig. S2C  
Our base-resolution measurement of hmC (TAB-Seq) allowed us to examine the relationship between the hydroxymethylation level (hmCG/CG) and the total methylation level (mC/C as measured in bisulfite sequencing, which is the sum of the frequency of hydroxymethylated and methylated sites). We restricted analysis to sites covered by at least 10 reads in the TAB-Seq data set. We sorted each of these sites into ten equally spaced bins according to the total level of methylation observed in bisulfite sequencing (mC/C) of the same sample. Within each bin, we plot the proportion of sites that have statistically significant 5hmC/C (uncorrected p<0.01, Fisher’s exact test). Significance
was assessed by comparing the number of hmC reads with the cumulative binomial distribution with mean parameter given by the bisulfite non-conversion rate.

**Fig. S3A and Table S2**
For each of the three adult mouse brain cell-type specific methylomes (R1, R2, R3) we used Fisher’s exact test to assess whether intragenic mCH/CH is larger in glia (NeuN–) vs. neurons (NeuN+) for each gene. A final list of 174 glial hyper-mCH genes was obtained by requiring a significant effect (FDR=0.05, Benjamini-Hochberg) in all three comparisons.

**Fig. S3B**
We profiled mC/C around genes reported to be enriched in oligodendrocytes (45) and in brain epithelial cells (78). The top oligodendrocyte genes were those with at least 20-fold enrichment (45).

**Fig. S3C**
We repeated our analysis of methylation patterns around functional gene categories in human brain samples by mapping each mouse gene to a human ortholog using the gene-oriented ortholog database (GOOD) (77).

**Fig. S8:** Methods as in Fig. 4.
Fig. S1.
(A,B) Global frequency of mC within all cytosine basecalls, categorized by sequence context, for each human and mouse sample. (C,D) Frequency of mC in neuronal samples within each CHN context. (E) Human neuron or (F) or glial cell sequence conservation at highly-methylated and unmethylated CH positions. Ideograms show sequence conservation and composition at positions flanking the most highly methylated CH sites (top row), unmethylated sites (middle), or all CH positions (bottom). (G) Total length of partially methylated domains in human genomes. (H) Proportion of methylated bases in CG or CH context that are methylated at a given level in human cortex and neuron samples.
**Fig. S2**

Detection of hydroxymethylcytosine in the CH context by TAB-Seq, and the relationship between mCG and hmCG level. Using sequencing of pUC19 control DNA, we calibrated the rate of false positive hmC basecalls due to non-conversion of fully unmethylated CH sites ($r_{CH}=0.0052$); as well as the non-conversion and/or non-oxidation of fully methylated (mC) CG sites ($r_{CG}=0.0208$). These two parameters define the likelihood function for the data under the null hypothesis. (A) Analysis of the pUC19 control data with multiple-comparison correction (Bonferroni, Benjamini-Hochberg (BH) (75) or Benjamini-Hochberg-Yekutieli (BHY) (76)) shows no false positives. (B) The same analysis applied to adult mouse TAB-Seq data results in ~50 significant sites throughout the genome. Note that the BH procedure assumes statistical independence among all sites, and is therefore a less conservative approach. (C) The fraction of mCG sites that show any significant hmC signal ($p<0.01$, binomial exact test) from the TAB-Seq data, given a specific level of mCG.
Fig. S3

mCG, mCH throughout gene body and flanking 100 kb, for mouse genes that show (A) glial mCH hypermethylation and (B) oligodendrocyte and epithelial genes. Transcript abundance (RNA-Seq FPKM) over development is shown for the same gene sets. (C) mCG, mCH throughout gene body and flanking 100 kb, for indicated gene sets in human. Transcript abundance (RNA-Seq FPKM) over development is shown for the same gene sets.
Fig. S4
Pearson correlation coefficient of methylation state at single sites between neurons and ES cells in human and mouse. Data are the same as in Fig. 2F, but shown with no normalization by the simulated full correlation value. Shuffled data are represented by the median over 200 random shuffles and error bars show the 5-95\textsuperscript{th} percentiles.
(A-B) Periodicity of mCG and mCH in human NeuN+ cells (53 yr female). (C) Within-read autocorrelation of mCH/CH (red) is enhanced relative to the correlation of summed methylation patterns (blue). (D) Box and whisker plots of gender-specific differences in intragenic mCG in inactivated and escapee genes on human chrX. (E) Scatterplot of gender differences in intragenic mCG in human chrX genes. Reported chrX inactivated and escapee genes (* Carrel et al.; § Sharp et al.), predicted escapee genes (†), and autosomal (Chr2) genes are indicated. (F) mCG and mCH within 1 kb bins is highly consistent between male and female humans. (G) ChrX mCG and mCH are reduced in female compared with male. (H) ChrX inactivated genes in female neurons have a slightly lower fraction of mCA dinucleotides.

Fig. S5
Fig. S6

Gender-specific methylation patterns across cell types. Male vs. female promoter mCG/CNG (top row), genic mCH/CNG (middle row) and genic mCG/CNG (bottom row) in human adult frontal cortex neurons (A), glia (B), and embryonic stem cell lines (C). Different symbols correspond to ChrX genes reported to be inactivated or to escape inactivation (* Carrel et al.; † Sharp et al.), predicted escapee genes (‡, this study), and autosomal (Chr2) genes. Neurons and glia, but not embryonic stem cells, are hypermethylated specifically in female chrX-escapee genes.
**Fig. S7**
Principal component (PC) analysis of mC patterns in 25,260 mouse genes. (A) Cumulative fraction of data variance explained by each PC. The first 5 PCs were retained for cluster analysis. (B) Scatter plots of the projection of mC pattern of each gene onto the space defined by two of the principal components (PCs). Colored points show chromosome X, neuronal and astrocytic genes, illustrating the segregation of gene mC patterns according to function and regulation.
Fig. S8

(A,B) The data from Fig. 4 are shown here with an absolute (unnormalized) scale for mC/C.
Fig. S9

(A) Total mCG/CG, mCH/CH, hmCG/CG and hmCH/CH in each genomic region. Values have been corrected for the sample-specific rate of bisulfite non-conversion. (B) Median hmC and mC level in different genome features (error bars 32-68th percentile). hmC data are duplicated from Fig. 5B. (C) Transcript abundance during development for the three Tet family genes measured by mRNA-Seq.
Fig. S10
Relationship between mC and hmC and gene expression. (A) Transcript abundance for each of 20,184 genes in 6 wk mouse frontal cortex with expression level FPKM>0.1. (B) Absolute level of methylation (mCG/CG, mCH/CH) and hydroxymethylation (hmCG/CG) within each gene and in flanking 100 kb regions. (C) mC and hmC level normalized by the flanking region. (D,E) Total intragenic mC/C is compared with hmCG/CG.
Fig. S11
Heatmaps of mCG and mCH at CG-DMRs identified between neurons, glia, and frontal cortex through development in mouse and human.
Fig. S12
Enrichment/depletion of CG-DMRs in at the transcriptional start site of cell-type specific and developmentally dynamic genes in mouse. (B) Quantification of local enrichment of hmC at developmentally dynamic CG-DMRs by hmC enrichment methodologies. (C) Density maps showing all developmental CG-DMRs (left-hand plots), or only those which have significantly more or less mCG/CG compared with a wild-type adult brain (6 wk). The Tet2−/− mutant mice have increased mCG/CG compared with WT, particularly for Fetal>Adult CG-DMRs, indicating Tet-dependent demethylation at these sites during development.
Table S1 (separate file)
Sample information and sequencing metrics.

Table S2 (separate file)
mCH deserts in human and mouse and the genes located within these domains.

Table S3 (separate file)
Glial mCH-hypermethylated genes.

Table S4 (separate file)
Gender-specific methylation on human X chromosome.

Table S5 (separate file)
Sample demographics.
References and Notes


62. Materials and methods are available as supplementary material on Science Online.


