SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Genomic profiling of DNMT3A in the developing brain, Related to Figure 1.

(A) Scatterplots comparing DNMT3A binding between wild type replicates (top) and between wild type and *Dnmt3a* cKO (bottom) in the cortex and hippocampus in 5 kilobase tiles across the genome.

(B) Genome browser views of DNMT3A binding in wild type and *Dnmt3a* cKO two-week cortex.

(C) Average DNMT3A distribution across defined DNMT3A-enriched regions in the two week old cortex (N=22,223) and hippocampus (N=6,556). DNMT3A signal was normalized to input DNA (log2 ratio). DNMT3A ChIP from wild type (*Dnmt3a* $^{fl/fl}$) and *Dnmt3a* cKO (*Nestin-cre*; *Dnmt3a* $^{fl/fl}$) cortex and hippocampus was plotted.

(D) Average DNMT3A distribution across all genes in the cortex and hippocampus. DNMT3A signal was normalized to input DNA, and the enrichment of DNMT3A signal in the wild type (*Dnmt3a* $^{fl/fl}$) compared to *Dnmt3a* cKO controls (*Nestin-cre*; *Dnmt3a* $^{fl/fl}$) (log2 ratio) was plotted.

(E) Average DNMT3A distribution across putative distal enhancers (Nord et al., 2013) (N=38,620). The enrichment of DNMT3A signal in the wild type (*Dnmt3a* $^{fl/fl}$) compared to *Dnmt3a* cKO controls (*Nestin-cre*; *Dnmt3a* $^{fl/fl}$) (log2 ratio) was plotted.

(F) Western blots of DNMT3A on cortical extracts across different stages of postnatal development, with histone H3 as a loading control.

(G) Scatterplot comparing DNMT3A binding at 2 weeks and 8 weeks across the genome in 5 kilobase tiles.

(H) Quantification of amount of histone H3 pulled down by immunoprecipitating DNMT3A. The relative amount of H3 in IP compared to input were calculated and normalized to that of wild type.

(I) Genome browser view of ChIP-seq data in the two-week cortex.

(J) Average distributions of histone marks over DNMT3A-enriched regions. The log2 ratio of histone marks to H3, and log2 ratio of DNMT3A to input DNA are shown.

(K) Correlation between H3K36me3 and DNMT3A binding within gene bodies. The average H3K36me3 density was binned according to DNMT3A density within genes (200 genes per bin).

(L) Average distributions of DNMT3A over regions of the genome enriched in the indicated histone marks or DNA-binding factor. Number of defined enriched regions (see STAR Methods): H2A.Z (17,020), H3K4me3 (30,072), H3K27ac (58,746), H3.1 (2,631), H3.3 (23,577), H3K9me3 (24,552), H3K27me3 (11,375), H3K36me3 (17,869), H3K9me2 (23,739), H3K27me2 (9,354), MECP2 (1,626), RNA POL II (All) (12,330), RNA POL II Ser2P (13,756).

(M) Prediction performance for DNMT3A binding based on various histone modifications. Gene promoters, TSS - 500 bp to TSS + 500 bp. Gene bodies, TSS + 3 kilobases to TTS. Intergenic, regions >10 kilobases away from annotated genes. Prediction was performed using linear models in 1 kilobase tiles spanning TSS for promoters, 5 kilobase tiles within gene bodies and 5 kilobase tiles across intergenic regions of the genome (see STAR Methods).

(N) Correlation between gene expression and DNMT3A binding within gene bodies. The average DNMT3A density was binned according to gene expression levels (200 genes per bin). Note decrease in DNMT3A occupancy in the extremely lowly expressed genes on the left (e.g. heterochromatic genes) as well as highly expressed genes on the right.

Figure S2. *In vitro* and *in vivo* characterization of DNMT3A catalytic activity, Related to Figure 2.

(A) Unbiased assessment of the sequence preference of DNMT3A activity. *In vitro* methyltransferase activity of human DNMT3A protein. DNA methylation levels measured by bisulfite sequencing were quantified for each cytosine context along the unmethylated lambda phage genome, and compared with lambda DNA incubated without the DNMT3A protein. Error bars represent s.d. between two technical replicates.

(B) Genome-wide DNA methylation levels in one-week, two-week, and eight-week cortex and hippocampus.

(C) Genome browser views of DNMT3A binding and DNA methylation at different developmental stages in the cortex and hippocampus.

(D) Average mCA patterns across genes of different expression levels in the wild type cortex and hippocampus.

(E) Methylation calls for spike-in controls of known methylation patterns were calculated. Scheme illustrating the oxBS-seq, adapted from (Song et al., 2012) also shown.

(F) Average distribution of mCG and oxidized forms of mCG over genes, determined by OxBS-seq.

(G) Correlation between two week DNMT3A binding and mCA levels across the genome at different stages of development of the hippocampus.

(H) Distribution of mCA in eight-week hippocampus over DNMT3A-enriched regions in the 2-week hippocampus (N=6,565).

(I) Boxplots of eight-week mCA and two-week DNMT3A in cortex and hippocampus within genomic tiles defined to be differentially CA methylated (see STAR Methods). mCA cortex > hippocampus: N=13,438; hippocampus > cortex: N=58. *P=2.3e-278, **P<1.7e-5, Wilcoxon rank-sum test.

(J) Genome browser views of DNA methylation in wild type and *Dnmt3a* cKO cortex (10 weeks).

(K) Distribution of mCA in wild type and *Dnmt3a* cKO cortex binned by DNMT3A densities across the genome in 5 kilobase tiles (5,000 tiles per bin).

(L) Distribution of mCG in wild type cortex (left) and 10-week wild type and *Dnmt3a* cKO cortex (right) binned by DNMT3A densities across the genome in 5 kilobase tiles (5,000 tiles per bin).

Figure S3. Causal relationship between DNMT3A binding and gene expression, Related to Figure 3.

(A) Genome browser view for DNMT3A binding and gene expression with KA treatment in two-week hippocampus on the activity-dependent *Nr4a1* locus (top). Quantification of the average densities of DNMT3A ChIP-seq reads within the *Fosb* and *SikI* gene bodies (bottom). Error bars represent s.d. for two biological replicates.

(B) Average distribution of 2-week hippocampal DNMT3A ChIP-seq reads across all genes and KA-induced genes (N=51).

(C) Gene expression levels of genes defined as KA-induced. RNA-seq data for two-week *Dnmt3a* cKO and littermate wild type controls are also shown.

(D) Average distribution of DNMT3A over genes of different expression levels in the wild type visual cortex.

(E) Effect of light exposure on DNMT3A binding over all genes and light-induced genes (N=33). DNMT3A ChIP-seq and input DNA read densities of two-week visual cortex within gene bodies were calculated, and log2 ratios between light exposed (light) and control (dark) samples are shown. *P=0.01, Wilcoxon rank-sum test.

(F) Box plots of gene expression levels of genes defined as light-induced. RNA-seq data for two-week *Dnmt3a* cKO and littermate wild type controls are also shown.

(G) KA-induced genes are still inducible after ten days of daily KA administration. Boxplots of gene expression levels of genes defined as KA-induced for daily KA-treated and control hippocampus at the final day of injections (P20). The RNA were collected an hour after injections. Data for two biological replicates are shown.
(H) Model of activity-dependent gene induction disrupting DNMT3A binding during early life.

Figure S4. Effect of a genetic mutation in the *Ezh2* gene on DNMT3A and mCA deposition, Related to Figure 4.

(A) Top, average H3K27me3 distribution in *Ezh2* cKO cortex compared to wild type across defined H3K27me3-enriched regions in wild type (N=11,375). Bottom, average DNMT3A distribution across defined H3K27me3-enriched regions in wild type.
(B) Average distribution of DNMT3A ChIP-seq reads across all genes and genes up-regulated in *Ezh2* cKO (N=11).
(C) Genome-wide DNA methylation levels in *Ezh2* cKO and wild type.
(D) Average distribution of mCA across all genes and genes up-regulated in *Ezh2* cKO.
(E) Average distribution of mCG across all genes and genes up-regulated in *Ezh2* cKO.
(F) Boxplots of the difference between mCG in *Ezh2* cKO and wild type control corticies within all genes and genes up-regulated in *Ezh2* cKO.

Figure S5. DNA methylome and transcriptome data in two neuron subtypes at three developmental time points, Related to Figure 5.

(A) Comparison of INTACT-purified nuclear RNA-seq with ribosome-bound mRNA-seq approach (Mardinly et al., 2016). Genes were classified by the INTACT-purified RNA-seq dataset (*Pv* high, *Vip* low genes: N=39; *Vip* high, *Pv* low genes: N=72; *Pv* high, *Vip* high genes: N=105), and gene expression levels were plotted. Data for two biological

replicates for the INTACT-purified nuclear RNA-seq and three biological replicates for the ribosome-bound mRNA-seq is shown. Genome wide correlation between the different datasets: *Pv*: ρ=0.73, p<1e-4; *Vip* ρ=0.67, p<1e-4.

(B) Genome-wide DNA methylation levels in *Pv* and *Vip* neurons.

(C) Overlaps of genomic 100 bp tiles with high levels of mCG (right) or mCA (left) between *Pv* and *Vip* neurons. Tiles were selected if they were greater than 70% for mCG, and greater than 10% for mCA (see STAR Methods). Gene ontology (GO) analyses of genes with high mCA densities within the gene bodies (top 5% of all genes) for both *Pv* and *Vip* neurons suggested that, in addition to annotated neuronal functions such as axon guidance, regulation of neuronal development, genes involved in various biological processes such as transcriptional regulation, response to stimuli, and cellular differentiation were also over-represented, consistent with the finding that mCA coats the neuronal genome.

(D) Expression levels of genes that are differentially expressed between *Pv* and *Vip* neurons during the early postnatal development. As a control for potential genetic differences between the two mouse strains that may potentially lead to differential gene expression, we performed RNA-seq on unpurified nuclei (i.e. before immunoprecipitation of the nuclei) from each mouse strain and observed no significant differences in gene transcription between these two mouse strains (right), confirming that the cell type-specific differences in gene expression we observe are not driven by genetic differences between the two strains of mice. Examples of neuronal subtype-specific genes at one week are also shown (below). Error bars represent s.d. for two biological replicates for *Pv* neurons and three biological replicates for *Vip* neurons.

(E) Average distribution of mCA in *Pv* and *Vip* neurons over genes of different expression levels at 3 weeks.

(F) Genome browser views of DNA methylation and transcription data in cortical *Pv* and *Vip* neurons at indicated developmental time points.

(G) Quantifications of intragenic mCA and gene expression levels of the genes shown in the browser tracks in Figures 5C and S5F. Error bars represent s.d. for two to three biological replicates.

(H) Average distribution of mCG over genes differentially expressed in *Pv* and *Vip* neurons at one week.

(I) Average distribution of mCG and mCA over genes differentially expressed in *Pv* and *Vip* neurons at three weeks.

Figure S6. High-throughput single-nuclei profiling of neurons deficient in DNMT3A, Related to Figure 6.

(A) Seurat t-SNE output of distinct cell populations.

(B) Number of post-filtered nuclei analyzed (top). Percentage of identified *Pv* and *Vip*-expressing nuclei among all nuclei (bottom). Error bars represent s.d. between three biological replicates. DNMT3A disruption did not result in strong changes in the proportions of *Pv* and *Vip*-expressing neurons in the *Dnmt3a* cKO cortex compared to wild type, suggesting that DNMT3A likely does not play a major role in specification of neuronal subtypes, consistent with our finding that mCA is deposited after specification. Having similar numbers of *Pv* and *Vip*-expressing neurons enabled us to carefully compare gene expression profiles between the two subtypes in the *Dnmt3a* cKO and wild type mice.

(C) Comparison between single-nuclei RNA-seq and INTACT-purified nuclear RNA-seq data. Genes were classified by the INTACT-purified RNA-seq dataset (*Pv* high, *Vip* low genes: N=34; *Vip* high, *Pv* low genes: N=56; *Pv* high, *Vip* high genes: N=102), and gene expression levels determined by single-nuclei RNA-seq were plotted.

(D) Association between mCA densities within gene bodies and gene expression levels in wild type and *Dnmt3a* KO *Pv* and *Vip* neurons (average single-nuclei RNA-seq data). Genes were separated based on the mCA densities within gene bodies: High-mCA: >9.55% (*Pv*) (N=1,716) or >6.67% (*Vip*) (N=1,716); Low-mCA: <5% (*Pv:* N=9,586 and *Vip:* N=12,878). P<0.05, two-tailed t-test. Error bars represent s.d. between three biological replicates.

Figure S7. Cell type-specific MECP2 binding, and single-nuclei RNA-seq in *Mecp2* KO brains, Related to Figure 7.

(A) Genome browser views of MECP2 and mCA data in 8-week cortical *Pv* and *Vip* neurons. Gene expression level determined by INTACT-purified RNA-seq is shown on right. Error bars represent s.d. between two biological replicates. Genome wide correlation between the different datasets: *Pv* data: ρ=0.79, p<1e-4; *Vip* data: ρ=0.75, p<1e-4.

(B) Boxplots of MECP2 densities within genes with differential CA methylation levels between *Pv* and *Vip* neurons (see STAR Methods). (mCA *Pv* > *Vip* genes: N=872; *Vip* > *Pv* genes: N=84). *P=1.6e-100, **P=4.4e-27 Wilcoxon rank-sum test. Consistent with the fact that mCA marks a wide variety of genes involved in range of biological processes, GO analyses on genes enriched for MECP2 binding within the gene bodies for both *Pv* and *Vip* neurons suggested that MECP2 binding is not restricted to genes involved in neuronal functions, but also binds across genes involved in other biological processes, such as development, cell proliferation, and transcriptional regulation (see STAR Methods).

(C) Seurat t-SNE plots of all nuclei in samples. Nuclei expressing indicated marker genes are depicted in purple. Number of *Pv* and *Vip*-expressing nuclei analyzed

(bottom), and percentage of identified *Pv* and *Vip*-expressing nuclei among all nuclei (right). Error bars represent s.d. between six biological replicates.

(D) Overlap between genes mis-regulated in *Dnmt3a* and *Mecp2* KO *Pv* and *Vip* neurons using a threshold approach. Genes with mCA levels greater than 5% mCA within gene bodies in wild type that became significantly up-regulated in *Dnmt3a* KO ($p<0.05$ cutoff) were defined to be a small subset of genes that are putatively directly repressed by DNMT3A (*Pv*: N=93; *Vip*: N=25). The expression levels of this small subset of DNMT3A-target genes in *Mecp2* KO and littermate wild type neurons were calculated. The expression levels of all genes were also plotted. *P=0.001, **P=0.022, Wilcoxon rank-sum test. The genes that are up-regulated in *Dnmt3a* or *Mecp2* KO neurons are partially overlapping between the two KO strains compared to wild type mice but they can differ between *Pv* and *Vip* neurons, and they include genes that are involved in specific neuronal processes that presumably display some neuronal subtype-specificity with regard to their function. For example, one gene that is selectively mis-regulated in *Pv*, *Sez6* modulates synaptic connectivity and is associated with febrile seizures in humans (Gunnersen et al., 2007; Yu et al., 2007). A gene, *Epha5* that is selectively up-regulated in *Vip* neurons encodes a protein that regulates axon guidance and mediates aggressive behavior in mice possibly via its selective expression in *Vip* neurons (Mamiya et al., 2008; Wilkinson, 2001).

(E) Correlation between gene length and gene expression levels in wild type and *Mecp2* KO neurons (average single-nuclei RNA-seq data) (500 genes per bin). Long genes (>100 kb) were separated based on the density of mCA within the gene bodies. High mCA: >5% (*Pv*: N=1,294; *Vip*: N=948); Low mCA: <5% (*Pv*: N=859; *Vip*: N=1,205). *P<0.001, two-tailed t-test. Error bars represent s.d. between six biological replicates.

(F) Gene expression levels of *Pex5l* (highly CA-methylated in *Pv* neurons) and *Stx16* (highly CA-methylated in *Vip* neurons) at three postnatal time points. Error bars

represent s.d. between two to three biological replicates. The reverse correlation between level of gene expression and extent of mCA and MECP2 binding holds true for the majority of genes but not all genes. Intriguingly, we find that a small subset of high-mCA genes is highly expressed. Inspection of the highly-expressed, high-mCA genes indicated that a subset of these genes including *Pex5l* and *Stx16* are moderately expressed during early life and subsequently become highly expressed in adults. The moderate expression of these genes in early life may facilitate methylation across their gene bodies, and these genes may then be up-regulated by other transcriptional control mechanisms. On the other hand, a small subset of genes that are highly expressed during early life nevertheless become highly CA-methylated, suggesting that there may be unknown gene-specific DNMT3A-targeting mechanisms that remain to be identified.

(G) Model of how early-life transcription modulates lasting mCA and MECP2 patterns across genes. Once mCA is deposited it is bound by MECP2 to fine-tune the expression of genes.

(H) Hind limb clasping in adult *Dnmt3a* cKO mice.

(I) Developmental weight loss in *Dnmt3a* cKO mice. Young (two weeks) and adult (ten weeks) *Dnmt3a* cKO (*Nestin-cre*; *Dnmt3a*$^{fl/fl}$) were compared to littermate wild type mice (*Dnmt3a* $^{fl/fl}$). Error bars represent s.d. between two to three biological replicates. *P=3.2e-5, **P=3.2e-4, two-tailed t-test.