Polycomb-mediated repression compensates for loss of postnatal DNA methylation in excitatory neurons

Junhao Li[1]*, Antonio Pinto-Duarte[2]*, Mark Zander[3], Chi-Yu Lai[2], Julia Osteen[2], Linjing Fang[4], Chongyuan Luo[3,7], Jacinta D. Lucero[2], Rosa Gomez-Castanon[3], Joseph R. Nery[3], Isai Silva-Garcia[2], Yan Pang[2], Terrence J. Sejnowski[2], Susan B. Powell[5], Joseph R. Ecker[3,6†], Eran A. Mukamel[1†], M. Margarita Behrens[2,5†]

Affiliation:

1. Department of Cognitive Science, University of California San Diego, La Jolla, CA
2. Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA
3. Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA
4. Waitt Advanced Biophotonics Core, Salk Institute for Biological Studies, La Jolla, CA
5. Department of Psychiatry, University of California San Diego, La Jolla, CA
6. Howard Hughes Medical Institute, La Jolla, CA
7. Current address: Department of Human Genetics, University of California, Los Angeles, CA

Abbreviated title: Two mechanisms of epigenetic repression regulate neuronal maturation

Keywords: Epigenetics, methylome, RNAseq, excitatory synapse, working memory

*These authors contributed equally

[†]**Corresponding authors:** mbehrens@salk.edu; emukamel@ucsd.edu; ecker@salk.edu

Abstract

Epigenetic modifications of DNA regulate gene expression throughout the lifespan in neurons. Two major epigenetic pathways of repression, DNA methylation and Polycomb repressive complex 2 (PRC2) mediated gene silencing, regulate neuronal physiology and function, but their relative contributions are unknown. We found that conditional loss of the *de novo* DNA methyltransferase *Dnmt3a* in mouse excitatory neurons altered expression of synapse-related genes, impaired the maturation of postsynaptic dendritic spines and dampened neuronal excitability. These phenotypes were accompanied by working memory and social interest deficits. To elucidate the epigenetic mechanisms, we performed deep sequencing of DNA methylation, transcription, and chromatin modifications in cortical excitatory neurons. Loss of *Dnmt3a* abolished postnatal accumulation of CG and non-CG DNA methylation, leaving neurons with an unmethylated, fetal-like epigenomic pattern at ~140,000 genomic regions. The PRC2 associated histone modification H3K27me3 increased at many of these sites, partially compensating for the loss of DNA methylation. Our results suggest a complex interaction between two key modes of epigenetic repression of gene expression during brain development that supports cognitive function in adulthood.

Epigenetic modifications of DNA and chromatin-associated histone proteins establish and maintain the unique patterns of gene expression in maturing and adult neurons. Neuronal development requires the reconfiguration of epigenetic modifications, including methylation of genomic cytosine (DNA methylation, or mC)[1,2] as well as covalent histone modifications associated with active or repressed gene transcription. While mC primarily occurs at CG dinucleotides (mCG) in mammalian tissues, neurons also accumulate a substantial amount of non-CG methylation (mCH) during postnatal brain development[1,3,4].

Accumulation of mCH, and the gain of mCG at specific sites, depend on the activity of the *de novo* DNA methyltransferase DNMT3A[5]. In mice, the abundance of *Dnmt3a* mRNA and protein peaks during the second postnatal week, a time of intense synaptogenesis and neuronal maturation[2,6]. Despite evidence for a unique role of *Dnmt3a* and mCH in epigenetic regulation of developing neurons, the long-term consequences of *Dnmt3a*-mediated methylation on brain function remain largely unknown[2].

One challenge for investigating the developmental role of *Dnmt3a* has been the lack of adequate animal models. Deleting *Dnmt3a* around embryonic day 7.5 driven by the *Nestin* promoter dramatically impaired neuromuscular and cognitive development and led to early death[7]. This early loss of *Dnmt3a* specifically affects the expression of long genes with high levels of gene body mCA[8,9]. By contrast, deletion of *Dnmt3a* starting around postnatal day 14 by using the *Camk2a* promoter caused few behavioral or electrophysiological phenotypes[10], with only subtle alterations in learning and memory depending upon the genetic background[11]. These results suggest that *Dnmt3a* has a critical role during a specific time window between late gestation and early postnatal life. During these developmental stages, regulated gains and losses of DNA methylation throughout the genome establish unique epigenomic signatures of neuronal cell types[12,13].

To address the role of *Dnmt3a*-dependent epigenetic regulation in the functional maturation of cortical excitatory neurons, we created a mouse line using the *Neurod6* promoter[14] (Nex-Cre) to delete exon 19 of *Dnmt3a*[15]. In this conditional knockout (cKO), *Dnmt3a* is functionally ablated in excitatory neurons in the neocortex and hippocampus starting in mid- to-late gestation (embryonic day E13-15; *Dnmt3a* cKO)[16]. In *Dnmt3a* cKO animals, DNA methylation was substantially disrupted in excitatory neurons, with consequences for behavior and synaptic physiology, but without early life lethality or overt brain morphological alterations. We generated deep DNA methylome, transcriptome and histone modification data in *Dnmt3a* cKO and control pyramidal cells of the frontal cortex, enabling detailed

assessment of the molecular basis of neurophysiological and behavioral phenotypes. We found that Polycomb repressive complex 2 (PRC2) associated chromatin modifications increase following loss of DNA methylation in cKO neurons. Our data support a dynamic interaction between two fundamental modes of epigenetic repression in developing brain cells, which together confer robustness on neuronal regulation and prevent more extensive consequences for the loss of DNA methylation.

Results

Dnmt3a conditional knockout in pyramidal neurons during mid-gestation specifically impairs working memory, social interest, and acoustic startle

Previous studies of *Dnmt3a* KO mice yielded results ranging from little cognitive or health effects[10,11] to profound impairment and lethality[7]. These studies suggest the developmental timing of *Dnmt3a* loss in neurons may determine the extent of subsequent phenotypes. Here, we took a targeted approach by functionally ablating *Dnmt3a* in cortical pyramidal cells starting during mid-gestation (Fig. 1a). We took advantage of the developmental onset of *Neurod6* expression between embryonic day E11 and E13, after the onset of *Nestin* expression[17] but well before the major postnatal wave of reprogramming of the neuronal DNA methylome (Supplementary Fig. 1a)[1]. We confirmed the deletion of *Dnmt3a* exon 19 in cortical excitatory neurons of cKO animals (Fig. 1b), and the reduction in DNMT3A protein in whole tissue extracts at early postnatal time points (P5 and P13) when *Dnmt3a* mediated accumulation of mCH normally begins in the frontal cortex[6] (Fig. 1b, Supplementary Fig. 1b,c). *Dnmt3a* cKO animals survived and bred normally, without overt morphological alterations in the brain (Supplementary Fig. 1d). We found no impairments in gross motor function in an open field test: cKO mice traveled more during the first 5 minutes of testing (Supplementary Fig. 2a) while performing fewer rearings associated with exploratory interest (Supplementary Fig. 2b). Moreover, cKO mice had no signs of increased anxiety-like

behavior on three separate behavioral tests (Supplementary Fig. 2c-e), in contrast with the reported anxiogenic effects of *Dnmt3a* knockdown in the mPFC of adult mice[18]. The absence of major impairments in overall health, motor function, or anxiety-like behavior established an appropriate baseline for investigating the role of *Dnmt3a* in specific cognitive and social behaviors.

Neurodevelopmental illnesses are associated with altered cognitive function, including working memory and sensorimotor gating[19–21], as well as social interest[22,23]. Unlike control mice, *Dnmt3a* cKO mice did not alternate spontaneously between the arms of a Y-maze ($p = 0.0079$ for males, $p = 0.011$ females, Fig. 1c), indicating impaired spatial working memory. Moreover, male *Dnmt3a* cKO animals showed less preference for interacting with a novel mouse in the three-chamber paradigm, indicating reduced social interest (Fig. 1d, left panel; $p = 0.01048$). Male *Dnmt3a* cKO mice also had significantly attenuated acoustic startle reflex ($p = 0.0019$, Fig. 1e and Supplementary Fig. 3b). We observed increased prepulse inhibition (PPI) in male *Dnmt3a* cKO mice, but this may be driven by the reduced startle reflex (Supplementary Fig. 3a, $p < 0.05$). It is noteworthy that the observed deficits in startle response were not due to hearing deficits, since *Dnmt3a* cKO mice displayed intact hearing in tests of prepulse inhibition and fear conditioning.

To test whether these deficits in specific neuro-cognitive domains reflect generalized impairment in brain function, we assessed long-term memory using a fear conditioning paradigm. There were no significant differences between *Dnmt3a* cKO and control male mice in acquisition or recall of fear memory following re-exposure to the context or conditioned stimulus after 24-48 h (Supplementary Fig. 3c-f). Altogether, these behavioral results indicate that *Dnmt3a* cKO in excitatory neurons specifically impairs working memory, social interest and acoustic startle.

Increased and decreased expression of hundreds of genes in *Dnmt3a* cKO excitatory neurons

To investigate the impact of epigenetic disruption on gene expression, we compared the transcriptomes of cKO and control excitatory neurons using nuclear RNA-seq (Supplementary Table 1). We isolated nuclei from excitatory neurons by backcrossing the *Dnmt3a* cKO animals into the INTACT mouse background[12] followed by fluorescence-activated nuclei sorting (FANS). Although sorted nuclei contain only a subset of the cell's total mRNA and are enriched in immature transcripts, nuclear RNA-seq is nevertheless a quantitatively accurate assay of gene expression that is robust with respect to neural activity-induced transcription[24,25]. Nuclear RNA abundance was highly consistent across independent replicates within the same group (Spearman correlation r = 0.92-0.95, Fig. 2a and Supplementary Fig. 4a). Given the repressive role of DNA methylation (mCG and mCH) in regulating gene expression in neurons[26,27], we expected to find increased gene expression in the cKO. Consistent with this, we detected 932 differentially expressed (DE) genes with higher expression in the cKO (FDR < 0.05, Fig. 2a, Supplementary Fig. 4b, Supplementary Table 2). To our surprise, we also detected significantly lower expression of 788 genes in cKO neurons. These DE genes cannot be explained by a direct repressive effect of DNA methylation, suggesting a more complex cellular response to the loss of mC.

Many key regulators of synaptic function and dendrite structure were differentially expressed in *Dnmt3a* cKO neurons. Both up- and down-regulated genes were significantly enriched for functions related to synaptic physiology and dendrite structure (Fig. 2b, FDR<10^{-3}). Indeed, "dendrite morphogenesis" was the most enriched biological process for down-regulated genes and was also highly enriched among up-regulated genes (Fig. 2b, Supplementary Fig. 4c, Supplementary Table 3). Several DE genes with annotated roles in dendrite morphogenesis were specifically related to glutamatergic synapses and voltage-gated channels, including *Camk2a, Camk2b, Cit, Cacna1a, Shank1* and *Shank2* (Fig. 2c).

These results suggest that mC accumulation in pyramidal neurons during the first two postnatal weeks may contribute to excitatory neuron synapse formation and maturation.

Loss of *Dnmt3a* impairs synaptic maturation and attenuates neuronal excitability

To directly test the impact of *Dnmt3a* cKO on dendritic morphology, we quantified the number and structure of 1,278 DiI-labeled dendritic spines (NexCre/C57: n = 701 from 5 mice; cKO: n = 577 from 4 mice) of layer 2 pyramidal neurons of the mouse prelimbic region (~2 mm anterior to Bregma) (Fig. 3a; Methods). This brain region is critical for working memory[28] and social approach behavior[29]. We reasoned that the prolonged period of neuronal maturation in mPFC may make it particularly vulnerable to developmental disruption of epigenetic regulation. While the overall density of dendritic spines was equivalent in control and *Dnmt3a* cKO neurons (Fig. 3b), the spines were significantly longer (mean length 2.219±0.052 μm in cKO, 1.852±0.034 μm in control) and narrower (mean width 0.453±0.008 μm in cKO, 0.519±0.008 μm in control) in *Dnmt3a* cKO neurons (Fig. 3c, KS test $p < 0.001$; Supplementary Fig. 5). Consistent with this, a larger proportion of spines in *Dnmt3a* cKO mice were classified as immature filopodia, and fewer were mushroom-shaped mature spines (Fig. 3d) according to pre-established morphometric criteria (see Methods). The proportion of spines with other morphologies, including branched spines with more than one neck (not shown), were not significantly different between genotypes (Fig. 3d). These data indicate a role for *Dnmt3a* in promoting the maturation of dendritic spines.

To test the impact of altered spine morphology on synapse sensitivity, we performed patch-clamp experiments in visually identified layer 2 pyramidal neurons from the prelimbic region (Fig. 3e). Whole-cell current clamp recordings showed that *Dnmt3a* cKO neurons required greater current injections to trigger an action potential (higher rheobase, $p = 0.0042$, Fig. 3f), though there was no difference in membrane potential at the firing threshold (Supplementary Fig. 6a). *Dnmt3a* cKO neurons

also produced fewer spikes in response to injected current (Fig. 3g). These neurons were slightly hyperpolarized at rest (mean -70.90 mV in cKO vs. -67.22 mV in control, p = 0.049, Fig. 3h), which could reflect differential expression of ion channels at the plasma membrane. Consistent with this, *Dnmt3a* cKO neurons were leakier (lower input resistance, Fig. 3i), suggesting increased expression of functional transmembrane ion channels. Indeed, *Dnmt3a* cKO mice expressed significantly more *Kcnq3* mRNA (Supplementary Table 2), a critical subunit of potassium channels that mediate hyperpolarizing M-currents in the sub-threshold voltage range for action potential generation[30] and thus regulate membrane potential and neuronal excitability[31]. Whole-cell voltage-clamp recordings of miniature excitatory postsynaptic currents (mEPSCs) showed slight, yet significant, increased amplitude variability in *Dnmt3a* cKO mice (Supplementary Fig. 6b, variance 0.820 pA in cKO vs. 0.063 pA in control, F-test p = 0.0032), consistent with a disruption at postsynaptic sites. However, we found no alteration in the mean amplitude (Supplementary Fig. 6b) or frequency (Supplementary Fig. 6c) of mEPSCs recorded at the soma.

***Dnmt3a* cKO abolishes postnatal DNA methylation**

Deletion of *Dnmt3a* during mid-gestation should disrupt the subsequent gain of DNA methylation at specific genomic sites during development[1,32], without affecting sites that maintain or lose methylation after E14.5. Using single base-resolution, whole-genome methylC-seq[33], we found that non-CG DNA (mCH) methylation was all but eliminated (0.065% of all CH sites in cKO vs. 2.11% in control) in excitatory neurons of the frontal cortex, validating the necessity of *Dnmt3a* for the postnatal accumulation of mCH[1,2] (Fig. 4a, Supplementary Fig. 7a). In contrast, mCG is present throughout the genome in neural progenitors and the genome-wide level of mCG remains constant throughout the lifespan[6]. We found that mCG is substantially reduced by ~11.8% throughout the genome in cKO neurons (Fig. 4a). mCG was reduced in 61% of all genomic bins (10kb resolution), and was significantly lower in all genomic

compartments (promoters, gene bodies, exons, introns, and intergenic regions) (Supplementary Fig. 7b). The reduction in mCG was strongly correlated with reduced mCH (Spearman correlation 0.677, $p < 10^{-3}$, Supplementary Fig. 7c). These data support a role for *Dnmt3a* in *de novo* CG DNA methylation across genomic compartments[1,2].

Reduced DNA methylation cannot fully explain altered transcription in *Dnmt3a* cKO

The alterations in gene expression after *Dnmt3a* cKO led us to investigate whether they are correlated with loss of DNA methylation at specific sites. We first analyzed DNA methylation around DE genes. The simple model of DNA methylation as a repressive regulator of gene expression implies that cKO neurons would have lower levels of mC at promoters and gene bodies of up-regulated genes, and higher mC levels at down-regulated genes. Instead, we found the mCG level was lower in cKO animals around both up- and down-regulated genes (Fig. 4b), consistent with the global reduction in mCG in the cKO animals (Fig. 4a). The difference in gene body methylation (mCG and mCH) was negatively correlated with gene expression changes[5,34] (Fig. 4c). However, this correlation accounted for <1% of the variance of differential gene expression (Supplementary Fig. 7e).

We next sought to determine if up- and down-regulated genes differ from each other in ways that could explain their different responses to the loss of *Dnmt3a*. Up-regulated genes were on average longer than down-regulated genes (Supplementary Fig. 4d, Wilcoxon, $p<10^{-5}$), consistent with the reported enrichment of mCA and MeCP2-dependent gene repression in long genes[5,8,9]. However, there was a broad distribution of gene lengths for both up- and down-regulated genes, and up-regulated genes were not significantly longer than unaffected, non-DE genes. We compared DNA methylation at up- vs. down-regulated genes to test whether these gene sets have different epigenetic profiles. In both cKO and control animals, the amount of mCG was similar around up-regulated genes and non-DE genes with comparable expression levels. By contrast, the mCG levels around down-regulated genes were lower than

around up-regulated genes and non-DE genes with matched expression (Fig. 4b). This suggests that down-regulated genes may be less susceptible to direct repression by DNA methylation, and that their downregulation is an indirect consequence of the loss of mC.

Whereas *Dnmt3a* cKO neurons lose 11.8% of the normal CG methylation, non-CG methylation is entirely abolished and hence may be more strongly correlated with differential gene expression. Indeed, we found that mCH is more abundant in up-regulated than in down-regulated genes in control neurons (Fig. 4b, Supplementary Fig. 7d). However, we also observed lower mCH in both up- and down-regulated genes than in non-DE genes with equivalent expression levels in control animals. These results suggest that some genes with relatively high levels of mCH and mCG can maintain their expression despite the loss of DNA methylation, and that de-repression is not a universal outcome for strongly methylated genes. The relatively lower mCH level in down-regulated genes, by contrast, could make them less sensitive to the loss of *Dnmt3a*. The dysregulation of their expression may be due to secondary effects subsequent to the direct loss of DNA methylation.

In addition to promoters and gene bodies, distal regulatory elements such as enhancers are major sites of dynamic DNA methylation where epigenetic regulation can activate or repress expression of genes over long genomic distances through 3D chromatin interactions[12,35,36]. We identified 141,633 differentially methylated regions (DMRs) where cKO excitatory neurons had substantially lower mCG compared with controls (at least 30% mCG difference; FDR<0.01; Supplementary Fig. 7f; Supplementary Table 4). For example, we found five DMRs in a ~40 kb region around the promoter of the differentially expressed gene *Nedd4,* several of which were also unmethylated in fetal cortex[1] (Fig. 4d). The majority of DMRs (67.8%) were distal from annotated transcription start sites (at least 10kb). DMRs were significantly enriched in both active enhancers and in repressed chromatin, suggesting that they have a regulatory role (Supplementary Fig. 7g; see also below). DMRs overlapped strongly with regions that

gain methylation during normal development[6] (6.9% of DMRs, overlap odds ratio 78.4; Fig. 4e). Moreover, the DMRs overlapped DNA sequence motifs of multiple transcription factors associated with neuronal differentiation, including *Rest*, *Lhx6*, *Pou3f2* and *Pax6* (FDR<0.05, Supplementary Fig. 7h, Supplementary Table 5). These results suggest that *Dnmt3a* is essential for the methylation and subsequent repression of neuronal enhancers that are active during prenatal brain development.

Although the loss of methylation at DMRs was substantial, it is only a part of the global reduction in mCG that we observed throughout the genome. Indeed, we found that mCG is reduced by ~10% in all genomic compartments, even after excluding DMR regions (Supplementary Fig. 7b, c). Moreover, the density of DMRs around the DE genes was not significantly different in the up- and down-regulated genes and the non-DE genes (Supplementary Fig. 7i). This further supports a complex link between reduced DNA methylation and the transcriptomic changes we observed in the *Dnmt3a* cKO animals.

Genome-wide gain of repressive histone modification H3K27me3 in *Dnmt3a* cKO

Given the weak correlation between changes in DNA methylation and gene expression, we explored other potential regulators of the DE genes. To identify transcription factors (TFs) and chromatin regulators with experimental evidence of binding at cis-regulatory regions of the DE genes, we performed Binding Analysis for Regulation of Transcription (BART)[37]. Chromatin regulators associated with Polycomb repressive complex 2 (PRC2), including *Ezh2*, *Suz12*, *Eed* and *Jarid2*, were among the top DNA binding proteins enriched near the promoters of both up- and down-regulated DE genes (Fig. 5a). Several TFs associated with chromatin organization, including the histone deacetylase (*Hdac*) and demethylase (*Kdm*) families, *Cxxc1* and *Ctcf*, were also enriched (Supplementary Fig. 8a, Supplementary Table 6). These results suggest *Dnmt3a* cKO impacts the chromatin landscape in excitatory neurons, potentially via altered PRC2 activity.

To experimentally address this, we performed ChIP-seq to measure trimethylation of histone H3 lysine 27 (H3K27me3), a repressive mark catalyzed by PRC2 that is important for transcriptional silencing of developmental genes. We also measured H3K4me3 (trimethylation of histone H3 lysine 4, associated with promoters) and H3K27ac (acetylation of histone 3 lysine 27, associated with active promoters and enhancers)[38–41]. In both *Dnmt3a* cKO and control animals, the active marks H3K4me3 and H3K27ac were positively correlated with gene expression, while the PRC2-associated mark H3K27me3, mCG and mCH levels were negatively correlated (Supplementary Fig. 8b). Comparing ChIP-seq data with DNA methylation, we noted regions, such as the *Mab21l2* locus (Fig. 5b), where increased H3K27me3 in *Dnmt3a* cKO neurons coincides with DMRs. Using a conservative strategy to call ChIP-seq peaks[42], we found strong conservation of peaks of H3K4me3 and H3K27ac in the cKO and the control (Fig. 5c). By contrast, cKO neurons had 51.9% more H3K27me3 peaks than control (Fig. 5c, Supplementary Table 7). When we directly identified differentially modified (DM) regions, we found no DM for H3K4me3 and H3K27ac between the cKO and control. By contrast, we found 4,040 regions with significantly enriched H3K27me3 in the cKO, covering ~31.05 MB of the genome (FDR<0.05, Supplementary Table 8).

The H3K27me3 DM regions were strongly associated with genes involved in the development of the nervous system and other tissues (Supplementary Fig. 8c, Supplementary Table 9), suggesting PRC2 may step in to silence developmental genes when DNA methylation is removed. Consistent with this, most (75.0%) of the regions that gain H3K27me3 in the cKO overlap with DMRs (Fig. 5d-f). Likewise, DMRs were significantly enriched in peaks and DM regions of H3K27me3 (28.1% overlapped H3K27me3 peaks in cKO), but not H3K27ac (13.5% overlapped H3K27ac peaks in cKO) or H3K4me3 (1.1% overlapped H3K4me3 peaks in cKO) (Fig. 5d-f). Moreover, H3K27me3 was more abundant at the center of DMRs (Fig. 5g). These results suggest that PRC2-mediated repression may compensate for the

loss of mCG and/or mCH, acting as an alternative repressive mechanism when DNA methylation is disrupted.

Bivalent promoters with both active and repressive marks are targets of developmental methylation

Epigenetic modifications of DNA and chromatin are particularly powerful regulators of expression when they occur in regions flanking CG-rich promoters, i.e. the shores of so-called CpG islands (CGIs)[43,44]. The impact of *Dnmt3a*-dependent DNA methylation may be different at active CGI promoters (marked by H3K4me3, but not H3K27me3) compared with bivalent promoters with both active and repressive marks (H3K4me3+/H3K27me3+). Bivalent chromatin modifications indicate a poised state that can permit transcription following an activating signal. Such regions are specifically targeted by *Dnmt3a1*, the longer of the gene's two isoforms[45]. Using the ChIP-seq data from control neurons, we identified 889 bivalent and 9,203 active CGI promoters (Supplementary Table 10). We found that *Dnmt3a* cKO reduced methylation in regions flanking bivalent, but not active, CGI promoters (Fig. 5h; Supplementary Fig. 9a). Even so, the loss of DNA methylation at bivalent promoters had little effect on gene expression (2.2% of genes were DE). Instead, we found that active CGI genes were more likely to be differentially expressed (10.9% of genes, including both up- and down-regulated genes; binomial test $p < 0.001$, Supplementary Fig. 9c). These results suggest that the shores of bivalent CGI promoters are key targets of *Dnmt3a*, although the loss of DNA methylation does not directly control gene expression at these loci.

Novel DNA methylation valleys with stronger H3K27me3 signal in the *Dnmt3a* cKO

The observed compensation of lower *Dnmt3a*-mediated DNA methylation by H3K27me3 prompted us to further investigate the interaction between these two epigenetic features. DNA methylation and H2K27me3 have complementary roles at DNA methylation valleys (DMVs), i.e. large regions (≥5 kb) with low mCG (≤15%) that occur around key transcriptional regulators of development in human and

mouse tissues[46,47]. Previous studies comparing the epigenetic profile of DMVs across tissues identified multiple categories, including constitutive DMVs present in all tissues as well as tissue-specific DMVs[47]. We found more than twice as many DMVs in cKO (2,969) compared with control (1,351) neurons (Fig. 6d, Supplementary Table 11), covering a greater genomic territory (25.29 MB in cKO, 11.93 MB in control). By clustering the DMVs using their pattern of DNA methylation, chromatin modifications, and gene expression, we found six distinct categories (Fig. 6b-d). Whereas most DMVs lack H3K27me3 (clusters C1, 2, 4 and 5), we found two clusters associated with moderate (C3) or high (C6) levels of PRC2 repression. Cluster C6 DMVs, such as the promoter of *Foxp2* (Fig. 6a), had high H3K27me3 and low mCG in both control and cKO neurons, and were not strongly affected by loss of mC in the *Dnmt3a* cKO. By contrast, cluster C3 DMVs, including the second promoter of *Foxp2* and the promoter of *Slc17a6* (encoding vesicular glutamate transporter, Vglut2), gain mC during normal development (Fig. 6a). These DMVs were notably enriched for regions that gain H3K27me3 in the cKO compared to control animals. The loss of mC in these regions in the *Dnmt3a* cKO did not lead to strong activation of gene expression, potentially due to compensatory PRC2-mediated repression.

The remaining clusters lack H3K27me3 and are instead marked by low mCG and either high H3K27ac (cluster C5) or both H3K27ac and H3K4me3 (clusters C1, C2). Cluster C5 DMVs, such as the *Foxp1* promoter (Fig. 6a), have high mCG in the fetal brain and lose methylation in excitatory neurons during brain development. This demethylation is not affected in the *Dnmt3a* cKO, and we found little difference between the control and cKO neurons at these sites. Finally, clusters C1 and C2 were enriched for up- and down-regulated genes, respectively. Like C5 DMVs, they lose methylation during excitatory neuron development and are expanded in the cKO compared with control animals at 6 weeks of age.

Discussion

We developed a new mouse model for the loss of DNA methylation in excitatory neurons at a critical mid-embryonic stage of neurodevelopment, prior to the postnatal increase in *Dnmt3a* expression and non-CG methylation[1]. Conditional deletion of *Dnmt3a* in cortical excitatory neurons using the Nex-Cre line[16] abolished non-CG methylation and reduced CG methylation throughout the genome. DNA methylation in both CG and non-CG contexts is negatively correlated with gene expression and thus considered a repressive epigenetic mark. Here, we manipulated DNA methylation in a specific cell type to test the causal role of this epigenetic modification *in vivo*. In contrast with the assumed repressive role of DNA methylation, we found that cKO neurons had both up-regulation and down-regulation of hundreds of genes. Complex patterns of both activation and repression have recently been reported in other *Dnmt3a* cKO models[34] and in manipulations of the DNA methylation reader MeCP2[8,48]. These gene expression changes may partly reflect the direct effect of lower gene body CH methylation and of loss of CG and CH methylation at gene promoters and distal enhancers, potentially via increased transcriptional initiation due to reduced MeCP2 binding across broad domains[8,49]. In addition, some expression changes could be a consequence of disruption of other regulatory processes such as transcription factor expression or chromatin modification. Indeed, we found that the *Dnmt3a* cKO DMRs overlapped with regions that gain methylation during normal development[1]. This suggests an essential role for *Dnmt3a* and DNA methylation in shaping the transcriptome during development, with potentially long lasting effects for the gene expression pattern of mature neurons.

Accumulating evidence indicates that DNA methylation can influence the deposition of histone modifications. In particular, there is a strong antagonistic relationship between DNA methylation and the PRC2 associated mark H3K27me3[50–53]. Switching between Polycomb- and DNA methylation-mediated repression has been observed during development and in cancer[54–56]. Severe depletion of mCG can lead to

redistribution of H3K27me3, causing derepression of developmental regulators such as the *Hox* gene clusters[52]. We did not observe ectopic expression of these genes, possibly due to the relatively modest reduction in mCG in our model compared with cells lacking *Dnmt1*[52]. Instead, we found that in cortical excitatory neurons, the global reduction of DNA methylation increased PRC2-mediated repression in three genomic contexts. First, we found increased H3K27me3 at thousands of DMRs which gain CG methylation during normal development but are left unmethylated in the cKO. Second, we found more loss of mCG in the flanking regions of the bivalent (i.e. PRC2 repressed) CGI promoters compared to those of active CGI promoters after the *Dnmt3a* cKO. Third, we found a group of large DNA methylation valleys (DMVs) spanning key regulatory regions that are demethylated in the *Dnmt3a* cKO and gain H3K27me3.

We found no differential expression in adult *Dnmt3a* cKO animals of the four core components of PRC2 (*Ezh2*, *Suz12*, *Eed* and *Rbbp4*). However, the expression of *Mtf2* (also known as *Pcl2*), whose protein product is a Polycomb-like protein reported to recruit PRC2 to its target genes[57], was ~47% higher in the *Dnmt3a* cKO (FDR = 0.016, Supplementary Table 2). *Mtf2* was also one of the top predicted regulators of DE genes based on comparison with public ChIP-Seq datasets (BART). These findings could indicate a potential mechanism for the increased H3K27me3 in *Dnmt3a* cKO neurons.

Given that the DNA 5-methylcytosine oxidase *Tet2* was also up-regulated in the *Dnmt3a* cKO, our results support the dynamic balance between *Dnmt3a*-dependent DNA methylation (mCG) and *Tet*-dependent hydroxymethylation (hmCG) that is recruited by PRC2 at the shores of bivalent CGIs and the boundaries of DMVs[45,47,58]. Since methylC-seq cannot distinguish mCG and hmCG, further experiments using OxBS-seq, TAB-seq are needed to address this. Overall, our results suggest that when DNA methylation is disrupted, PRC2-mediated repression may partially compensate for the loss of mCG and/or mCH and act as an alternative mode of epigenetic repression.

Mid-gestation deletion of *Dnmt3a* in pyramidal neurons caused the differential expression of 1,720 genes in the frontal cortex of animals analyzed at 5 weeks of age. A substantial fraction of these genes had annotated roles in dendrite morphology and synapse function. Previous studies reported gross motor deficits and a shortened lifespan following early embryonic deletion of *Dnmt3a*[7], or only slight alterations when the deletion occurred past the second postnatal week[10,11]. In contrast, our mid-embryonic *Dnmt3a* cKO mice had no overt motor deficits, but had immature synaptic phenotypes and were impaired in specific cognitive domains. Layer 2 pyramidal neurons in the prelimbic region of *Dnmt3a* cKO animals had a larger number of immature dendritic spines and were significantly less sensitive to somatic injections of depolarizing current compared to control neurons. In the mPFC (which includes the prelimbic region in mice), axons from other brain regions arborize in unique patterns throughout superficial layers 1 and 2/3[59]. The dendrites of layer 2 pyramidal neurons extend through these layers, and are well-positioned to sample diverse excitatory inputs[60] carrying functional signals related to attention, cognition, emotion and memory[61]. Our data showing impaired maturation of dendritic spines and reduced excitability in the prelimbic region suggest that the developmental disruption of DNA methylation may affect behaviors linked to those circuits. Indeed, we observed that *Dnmt3a* cKO mice had significant impairments in working memory and social interest. These data suggest that the disruption of methylation patterns established by *Dnmt3a* during neural maturation could have mechanistic relevance in the context of neurodevelopmental disorders, complementing recent genetic risk association studies which linked *Dnmt3a* with autism[62].

Our findings highlight the critical and interconnected roles in brain development and cognitive function of two major modes of epigenetic repression of gene expression: DNA methylation and PRC2-mediated repression. The loss of DNA methylation in excitatory neurons has far-reaching effects on gene expression, synaptic function, and cognitive behavior, and triggers a compensatory gain of PRC2-associated repressive marks. Although PRC2-associated repression partially compensates for the

loss of DNA methylation, the developmental programs are nevertheless disrupted leading to altered circuit formation and behavioral dysfunction. Our cKO is a restricted manipulation of one neuron type, yet it directly impacts DNA methylation throughout the genome at millions of sites. This complicates the interpretation of the effects observed in mature neurons in the adult. Future work focusing on earlier developmental stages, and using targeted methods to manipulate epigenetic marks in local genomic regions[63], may help elucidate the causal interactions among epigenetic modifications that are critical for neuronal maturation and function.

METHODS

Methods and associated references are available in the online version of the paper.

Acknowledgments

This work was supported by R01MH112763 to MMB and JRE, and a Kavli Foundation award to MMB, APD and SBP. JRE is an Investigator of the Howard Hughes Medical Institute. We acknowledge stimulating discussions with Huda Zoghbi and Laura Lavery. We thank the members of the Salk Biophotonics Core Dr. Uri Manor, Sammy Weiser Novak, and Dr. Tong Zhang for insightful suggestions. We also thank Joseph Chambers and Colleen Heller for technical assistance in animal handling, and Ms. Faith Zhang for her involvement in the morphometric analysis of dendritic spines. The Waitt Advanced Biophotonics Core Facility at the Salk Institute receives funding from NIH-NCI CCSG: P30 014195 and the Waitt Foundation. The Flow Cytometry Core Facility of the Salk Institute receives funding from NIH-NCI CCSG: P30 014195. The authors have no conflict of interest in relation to the work described here.

Data access. All sequencing data are available in the Gene Expression Omnibus under accession GSE141587 (to be released upon publication). A genome browser displaying the sequencing data is available at https://brainome.ucsd.edu/annoj/mm_dnmt3a_ko/.

Author contributions. MMB, JRE, EAM designed and supervised the research. MMB, YP, JDL created transgenic mouse models and prepared samples for sequencing. APD, SBP performed and interpreted behavioral studies. APD performed slice electrophysiology studies. APD, JO, CYL, LF, ISG analyzed spine morphology. CL, JRN, RGC, performed methylC-seq, RNA-seq. MZ, JRN and RGC performed ChIP-seq. JL, CL, EAM performed bioinformatic and computational analysis of sequencing data. JL, APD, MZ, CL, JRE, EAM, MMB interpreted the findings. JL, APD, EAM, MMB wrote and edited the manuscript in consultation with all authors.

Figure legends:

Fig. 1 | *Dnmt3a* conditional knockout (cKO) in cortical pyramidal neurons during mid-gestation impaired working memory, social interest and acoustic startle responses. **a,** Experimental model of the conditional loss of *Dnmt3a* in excitatory neurons. P39, postnatal day 39. FANS, fluorescence-activated nuclei sorting. **b,** RNA-seq confirmation of the deletion of *Dnmt3a* exon 19 in excitatory neurons. RPKM, reads per kilobase per million. R1/2, replicate 1/2. T-test p = 0.08. **c**, *Dnmt3a* cKO mice made fewer spontaneous alternations in the Y-maze test of working memory (Wilcoxon test, **, p = 0.0079; *, p = 0.011; n =15 male control, 15 male cKO, 11 female control, 10 female cKO). **d**, Male *Dnmt3a* cKO mice spent less time interacting with an unfamiliar mouse, indicating reduced social interest (Wilcoxon test; * p=0.01048; n =14 male control, 15 male cKO, 11 female control, 10 female cKO) **e**, Male *Dnmt3a* cKO mice had larger startle response to a 120 dB acoustic pulse (Wilcoxon test, **, p = 0.0019; n.s., not significant).

Fig. 2 | *Dnmt3a* cKO disrupted expression of genes associated with dendritic morphology in cortical pyramidal neurons. **a,** Gene expression in control vs. *Dnmt3a* cKO excitatory neurons, showing significant up- (red) and down-regulated (blue) differentially expressed (DE) genes (false discovery rate FDR < 0.05). Examples of genes involved in dendrite morphogenesis are highlighted. Inset: Consistent mRNA-seq across two replicates of *Dnmt3a* cKO. TPM, transcripts per million; r, Spearman correlation coefficient. **b,** Top 10 significantly enriched Gene Ontology Biological Process annotations for DE genes (dashed line: FDR = 0.05). **c,** Fold-change for DE genes with an annotated role in dendrite morphogenesis.

Fig. 3 | Immature spine morphology and reduced excitability of layer 2 excitatory neurons following *Dnmt3a* cKO. a, Example dendritic segments of layer 2 pyramidal neurons in the prelimbic region labeled with DiI and visualized using a 63x objective coupled to an Airyscan confocal microscope. Arrowheads show filopodia, which were more abundant in *Dnmt3a* cKO mice. **b**, The density of membrane protrusions was unchanged in the *Dnmt3a* cKO (Wilcoxon test, n.s., not significant). **c,** Membrane protrusions were significantly longer and narrower in the *Dnmt3a* cKO (K-S test, p < 0.001). **d,** More spines were classified as immature filopodia, and fewer as mature mushroom shaped spines with large postsynaptic densities (Wilcoxon test, **, p = 0.0015; *, p = 0.046 and 0.011 for Thin and Mushroom, respectively). **e**, Example whole-cell patch clamp recordings from prelimbic layer 2 pyramidal neurons following 60 pA current injections, the minimal current necessary to trigger an action potential (rheobase) in *Dnmt3a* cKO. **f**. The median rheobase was significantly higher in the *Dnmt3a* cKO (t-test, *, p = 0.0042). **g**. Action potential frequency vs. injected current (mean ± s.e.m) showed reduced excitability in *Dnmt3a* cKO (Wilcoxon test, *, p < 0.05). **h,i** *Dnmt3a* cKO neurons were slightly hyperpolarized at V_{rest} when compared to control (Wilcoxon test, *, p = 0.049) and had lower membrane resistance (Wilcoxon test, *, p = 0.023).

Fig. 4 | Loss of *Dnmt3a* leaves thousands of genomic regions in a fetal-like demethylated state. a, Non-CG DNA methylation (mCH) is eliminated, and mCG is reduced, in *Dnmt3a* cKO. (T-test: *, p = 0.025; **, p = 0.0020). **b,** DNA methylation (mCG, mCH; mean ± SEM) in 1 kb bins in the region around the transcription start (TSS) and end site (TES) of DE genes and non-DE genes with matched expression level. **c,** Difference in gene body methylation vs. fold-change of gene expression between *Dnmt3a* cKO and control. The plots show mean ± SEM gene

expression fold-change for genes in each of 10 non-overlapping bins (deciles of mC difference). **d,** Example locus containing five differentially methylated regions (DMRs) with naive, fetal-like mCG in *Dnmt3a* cKO near the *Nedd4* promoter. The height of ticks shows the level of mCG at individual CG sites. *Dnmt3a* cKO DMRs are shaded in blue and marked with arrows. CGI, CpG island. Fetal and Adult cortex data are from [6]. **e**, *Dnmt3a* cKO hypo-DMRs are significantly enriched (red) in DMRs that normally gain methylation during development and depleted (blue) in DMRs that normally lose methylation during development (Fisher test, $p < 0.05$).

Fig. 5 | Polycomb repressive complex 2 (PRC2) associated histone modification H3K27me3 is upregulated following loss of DNA methylation. a, Transcription factors (TFs) predicted to regulate differentially expressed genes include many proteins associated with PRC2. The functional TF rank score was assigned by Binding Analysis of Regulation of Transcription (BART) [37]. **b,** Browser view of the *Mab21l2* locus, where increased H3K27me3 coincides with loss of DNA methylation in *Dnmt3a* cKO. DMR, differentially methylated region; DMV, DNA methylation valley. **c,** Histone modification ChIP-seq peaks for active marks (H3K4me3, H3K27ac) are largely preserved in the *Dnmt3a* cKO, while repressive H3K27me3 peaks expand. **d,** Most (92.9%) of the H3K27me3 peaks, but only some (28.0%) of the H3K27ac peaks, overlap with DMRs. **e-f,** Significant enrichment (red) or depletion (blue) of DMRs in the histone modification ChIP-seq peaks (**e**) and differentially modified regions (**f**) (Fisher test, $p < 0.05$). **g,** Histone modification ChIP-seq signal around the center of DMRs. Shaded ribbon indicates standard error of the mean. RPKM, reads per kilobase per million. ***, Wilcoxon rank-sum test $p < 0.001$. **h,** *Dnmt3a* cKO DMRs are enriched around bivalent, but not active CGI promoters. Z-scores show the observed number of overlaps between DMRs and CGI promoters, normalized by a shuffled background; dashed lines indicate significant enrichment/depletion ($p<0.05$).

Fig. 6 | Distinct clusters of DNA methylation valleys were associated with the increased H3K27me3 signal in the *Dnmt3a* cKO. a, Representative browser tracks show examples of unique DNA methylation valleys (DMVs) in the *Dnmt3a* cKO samples and the increased H3K27me3 signal in their flanking regions. For each type of signal, the tracks for the Control sample are labeled black and the tracks for the *Dnmt3a* cKO samples are labeled orange. In the mCG panel, the grey boxes indicate DMVs, and the orange boxes indicate cKO hypo DMRs. In the histone ChIP-seq panels, the black and orange boxes indicate peaks called in the Control and the cKO samples, respectively, and the red boxes indicate up-regulated differentially modified regions in the cKO samples. **b**. Heatmap of the DMVs clustered by their methylation levels and histone modification marks signals. The last two columns show the enrichments of differentially modified peaks of H3K27me3 and differentially expressed genes in these DMVs. R1/2, replicate 1/2. RPKM, reads per kilobase per million. **c**. Boxplots show distributions of methylation levels, histone markers signals, associated gene expression, DE genes enrichment, enrichment of differentially modified peaks of H3K27me3, and the length of the DMVs shown in (**b**). **d**. Number of DMVs identified in the *Dnmt3a* cKO and the Control samples. Left panel, DMVs are merged and grouped by whether they are identified in the cKO and/or the control samples. Most DMVs are unique to the cKO samples, or expanded in the cKO samples. Right panel, the stack bars are colored by the overlaps of the DMVs with the H3K4me4 and/or the H3K27me3 ChIP-seq peaks.

Supplementary Fig. 1 | *Dnmt3a* was disrupted on both the mRNA and protein levels in the *Dnmt3a* cKO excitatory neurons. a, Expression quantification of *in situ* hybridization data of gene *Nestin* and *Neurod6* in telencephalic vesicle from the Allen Developing Brain Atlas (http://developingmouse.brain-map.org/). Left panel, heatmap of the gene expression across ages during development. Right panel, example images of the gene expression in E11.5. Image credit: Allen Institute. E11.5-18.5, embryonic days; P4-28, postnatal days. **b,** Genome browser tracks of mRNA-seq data show confirmation of the deletion of *Dnmt3a* exon 19 in *Dnmt3a* cKO excitatory neurons. The targeted exon region is highlighted in the light blue shaded box with an asterisk. R1/2, replicate 1/2. **c,** The protein product of the *Dnmt3a* gene is disrupted in the cKO sample. Top panel, Western blot; Bottom panel, quantification of the protein abundance. P5 and P13, postnatal days 5 and 13. **, T-test p = 0.0017. **d,** Nissl stained slices show no morphological alterations in the brain of the *Dnmt3a* cKO animals. mPFC, medial prefrontal cortex; CPu, caudate putamen.

Supplementary Fig. 2 | The conditional ablation of *Dnmt3a* in pyramidal neurons did not significantly impair motor activity nor increased anxiety levels. a and **d,** *Dnmt3a cKO* mice displayed normal behavior in the open field test, traveling a similar distance **(a)** as control mice, and also showing a similar degree of center activity **(d)**. **b**, The exploratory activity was slightly decreased in *Dnmt3a* cKO animals, as suggested by an attenuated rearing behavior. **c,** The time spent in light in the dark-light transfer test was not significantly affected by the lack of *Dnmt3a.* **e,** The female, but not the male, cohort of *Dnmt3a* cKO mice spent significantly more time than control mice in the open arms of the elevated plus maze, consistent with lower anxiety levels (Wilcoxon test, **, p = 0.0048; n.s., not significant.). In the line plots (**a-b**), data were presented as mean ± s.e.m. In all boxplots (**c-e**), the middle horizontal bar represents the median; the lower and upper hinges correspond to the first and third quartiles, and the whisker

extends from the hinge to the value no further than 1.5 * IQR from the hinge, where IQR is the interquartile range. The values of individual experiments are represented by dots superimposed on the boxplots. Wilcoxon test significance: *, $p < 0.05$; **, $p < 0.01$; n.s., not significant.

Supplementary Fig. 3 | *Dnmt3a* cKO-induced impairment of startle response was accompanied by increased prepulse inhibition, and the cKO did not affect fear memory. **a,** The percentage of PPI at prepulse intensity of 69, 73 and 81 dB (4, 8 and 16 dB above the 65 dB background, respectively) was increased in male, but not female mice (Wilconxon test, ***, $p = 0.00076$; *, $p = 0.016$; n.s, not significant). **b,** The increased PPI accompanied the impairment in startle responses to a 120 dB tone played at three time points during the recording session (HAB1 - beginning of session; HAB2 - middle of session; HAB3 - end of session) (Wilconxon test $p = 0.0027$, 0.0019 and 0.0035 in male HAB1, HAB2, HAB3, respectively, and not significant in female). The habituation to the 120 dB auditory tone (i.e. the relative reduction in startle response throughout the experiment) was not significantly different between genotypes. **c-f**, Fear learning and extinction were tested over four consecutive days. *N* = 14-15 per group. **c**, Fear acquisition to three tone-shock pairings occurred on day 1; **d**, contextual fear in relation to the acquisition context (8 min, Block = 2 min) was measured on day 2; **e**, cued fear recall and extinction training occurred on day 3 (Block = four tone trials), and **f**, extinction recall (Block = four tone trials) occurred on day 4. Wilcoxon test reported no significant changes between the *Dnmt3a* cKO and control. In boxplots (**a**), the middle horizontal bar represents the median; the lower and upper hinges correspond to the first and third quartiles, and the whisker extends from the hinge to the value no further than 1.5 * IQR from the hinge, where IQR is the interquartile range. The values of individual experiments are represented by dots superimposed on the boxplots. In the line plots (**b-f**), data were presented as mean ± s.e.m.

Supplementary Fig. 4 | RNA-seq data showed transcriptomic disruption, especially in genes associated with dendritic morphogenesis, in *Dnmt3a* cKO pyramidal neurons. a, Correlation of gene expression in the two biological replicates of the control mouse excitatory neurons. TPM, transcripts per million. r, Spearman correlation coefficient. **b,** Volcano plot shows the gene expression fold-change of *Dnmt3a* cKO vs. Control samples and their significances. Significant up- and down-regulated differentially expressed genes (DE genes, FDR < 0.05) are colored in red and blue, respectively. **c,** Comparison of fold enrichment of GO terms significantly enriched (FDR < 0.05) in up- and down-regulated genes. Terms significantly enriched only in up-regulated genes, enriched only in down-regulated genes and enriched in both the up- and down-regulated genes are colored in red, blue and purple, respectively. Representative terms related to dendrite and spine development are labeled. **d,** Gene length distribution of DE genes. As comparison, non-DE genes were selected with FDR ≥ 0.05 and fold-change < 1.1 (see Supplementary Table 2). The down-regulated genes are generally shorter than the up-regulated genes or the non-DE genes. kb, kilobases. Wilcoxon test, ****, $p < 10^{-5}$, n.s. not significant.

Supplementary Fig. 5 | The membrane protrusions in the *Dnmt3a* cKO showed longer dendritic spines and narrower heads. Related to Fig. 3c. Here each line/box represents data from one dendrite fragment. **a** and **c,** The cumulative distribution of the length (**a**) and width (**c**) of membrane protrusions in the *Dnmt3a* cKO (orange), as compared to control mice (black). **b** and **d**, distribution of the length (**b**) and width (**d**) of the protrusions. The shaded colors in the box represent the individual animals from which the dendrite fragment originated (5 control mice and 4 cKO mice). The middle horizontal bar represents the median; the lower and upper hinges

correspond to the first and third quartiles, and the whisker extends from the hinge to the value no further than 1.5 * IQR from the hinge, where IQR is the interquartile range. The Wilcoxon test was done with the medians of the boxes (16 control vs. 15 cKO).

Supplementary Fig. 6 | The conditional ablation of *Dnmt3a* in pyramidal neurons did not significantly alter the membrane potential threshold for action potential generation, the mean amplitude or frequency of miniature excitatory postsynaptic events. a, Action potentials were initiated at a similar membrane potential in both genotypes (Wilcoxon test, n.s., not significant). **b**, While mEPSCs amplitude was not significantly different between genotypes (Wilcoxon test, n.s., not significant), it was slightly, yet significantly, more variable in the *Dnmt3a* cKO (F-test p = 0.0032). **c,** mEPSCs frequency was not significantly changed. In all boxplots, the middle horizontal bar represents the median; the lower and upper hinges correspond to the first and third quartiles, and the whisker extends from the hinge to the value no further than 1.5 * IQR from the hinge, where IQR is the interquartile range. The values of individual experiments are represented by dots superimposed on the boxplots.

Supplementary Fig. 7 | Genome-wide reduction of DNA methylation was observed in *Dnmt3a* cKO, but this cannot fully explain the disruption in the transcriptome. a, The reduction of genome-wide DNA methylation level is observed in all three non-CG contexts (CA, CC and CT). t-test significance, ***, p = 0.00087; **, p = 0.0050. **b,** DNA methylation at CG sites is reduced across all functional genomic compartments. UTR, untranslated region. t-test significance, ***, p < 0.001; *, p < 0.05. **c,** Reduced mCG is strongly correlated with the reduction in mCH in 10-kb tiling genomic bins (258,012 bins with at least 10 reads covered in each sample). r, Spearman correlation coefficient. **d,** Correlation of gene expression and gene body mCH level for up-regulated genes (red), down-regulated genes (blue) and non-DE genes (black) in the Control samples. For each gene group, genes are stratified by their expression in the control sample by 15 bins, and the mean gene body mCH levels are plotted. The shaded ribbon areas indicate the standard error of the mean. TPM, transcripts per million. **e,** Density scatter plots showing the relationship between changes of gene body methylation (delta mCG or mCH) and the gene expression fold changes for expressed genes (14,534 genes) between *Dnmt3a* cKO and control samples. The linear regression fits, p-values and variances explained by Δ%mC (R^2) are shown. **f,** 2-D distribution of mCG levels in Control vs. *Dnmt3a* cKO samples at DMRs. DMR density is estimated through a Gaussian smoothed kernel. **g,** Enrichment (red) or depletion (blue) of DMRs in GENCODE annotated gene features (top), in CpG island related features (middle), and in the chromatin states map in mouse embryonic stem cell (bottom). All enrichments and depletions shown are significant (fisher test p < 0.05). **h,** Number of known transcription factor binding motifs within *Dnmt3a* cKO hypo-DMRs and their fold enrichment. Significant motifs (FDR < 0.05) are colored in red. **i,** The density of *Dnmt3a* cKO hypo-DMRs around differentially expressed genes. TSS, Transcription Start Site; TES, Transcription End Site.

Supplementary Fig. 8 | H3K27me3 signal was increased in *Dnmt3a* cKO and highly overlapped with DMRs. a, Transcription factors (TFs) associated with chromatin organization were predicted to regulate DE genes in the *Dnmt3a* cKO. The functional TF rank score was given by Binding Analysis of Regulation of Transcription (BART)[37]. TFs predicted to regulate only up-regulated genes, to regulate only down-regulated genes and to regulate both the up- and down-regulated genes are colored in red, blue and black, respectively. **b,** Correlation

heatmap of epigenetic marks and RNA reads in 10 kb tiling genomic bins. For mCG and mCH, the mean methylation in each bin is used. For RNA-seq and histone modification ChIP-seq, the RPKM (reads per kilobase per million) value in each bin is used. The dendrogram is calculated by hierarchical clustering with complete linkage using the euclidean distances of the Spearman correlation across the samples. Rep1/2, replicate 1/2. **c,** Top 10 Gene Ontology terms of the Biological Process ontology enriched in the genes associated with regions with up-regulated H3K27me3 signal in the *Dnmt3a* cKO samples. Gene-region association was achieved by the GREAT tool. FDR, false discovery rate; the vertical dashed line shows the threshold of FDR = 0.05. **d,** Venn diagrams show overlaps of *Dnmt3a* cKO DMRs and H3K27me3 and H3K27ac peaks in Control samples. 64.6% of the H3K27me3 peaks overlap with DMRs and 22.4% of the H3K27ac peaks overlap with DMRs.

Supplementary Fig. 9 | The shores of bivalent CGI promoters were targets of *Dnmt3a*. a, Bivalent CGI promoters had lower levels of the active chromatin marks, H3K4me3 and H3K27ac , consistent with the concept that bivalent promoters are poised but not active[38]. Histone modifications and DNA methylation around bivalent (H3K4me3+/H3K27me3+) and active (H3K4me3+/H3K27me3–) CGI promoters. Left panel, lines represent mean values in 100-bp bins of the 5 kb upstream and downstream regions flanking the CGI promoter centers. Bins whose signal values are significantly (t-test FDR < 0.05) different between the cKO and control samples are highlighted in light-blue bands. Right panel, lines represent the differences of the signals between the cKO and control samples for bivalent CGI promoter and active CGI promoters, respectively. CGI, CpG island. RPKM, reads per kilobase per million. kb, kilobase pairs. **b,** Expression of genes with bivalent CGI promoters and genes with active CGI promoters. Genes with bivalent CGI promoters are mostly not expressed. R1/2, replicate 1/2. TPM, transcript per million. **c,** The percentages of genes that were differentially expressed (DE) between cKO and control were higher in the group of genes associated with active CGI promoters compared to the group of genes associated with bivalent CGI promoters.

References

1. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
2. Stroud, H. *et al.* Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* **171**, 1151–1164.e16 (2017).
3. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320–331 (2010).
4. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
5. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015).
6. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905–1237905 (2013).
7. Nguyen, S., Meletis, K., Fu, D., Jhaveri, S. & Jaenisch, R. Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.* **236**, 1663–1676 (2007).
8. Boxer, L. D. *et al.* MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes. *Mol. Cell* (2019) doi:10.1016/j.molcel.2019.10.032.
9. Kinde, B., Wu, D. Y., Greenberg, M. E. & Gabel, H. W. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 15114–15119 (2016).
10. Feng, J. *et al.* Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010).

11. Morris, M. J., Adachi, M., Na, E. S. & Monteggia, L. M. Selective role for DNMT3a in learning and memory. *Neurobiol. Learn. Mem.* **115**, 30–37 (2014).

12. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).

13. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).

14. Schwab, M. H. *et al.* Neuronal basic helix-loop-helix proteins (NEX and BETA2/Neuro D) regulate terminal granule cell differentiation in the hippocampus. *J. Neurosci.* **20**, 3714–3724 (2000).

15. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).

16. Goebbels, S. *et al.* Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *Genesis* **44**, 611–621 (2006).

17. Thompson, C. L. *et al.* A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* **83**, 309–323 (2014).

18. Elliott, E. *et al.* Dnmt3a in the Medial Prefrontal Cortex Regulates Anxiety-Like Behavior in Adult Mice. *J. Neurosci.* **36**, 730–740 (2016).

19. Barendse, E. M. *et al.* Working memory deficits in high-functioning adolescents with autism spectrum disorders: neuropsychological and neuroimaging correlates. *J. Neurodev. Disord.* **5**, 14 (2013).

20. Habib, A., Harris, L., Pollick, F. & Melville, C. A meta-analysis of working memory in individuals with autism spectrum disorders. *PLoS One* **14**, e0216198 (2019).

21. Adler, L. E. *et al.* Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biol. Psychiatry* **17**, 639–654 (1982).

22. Frye, R. E. Social Skills Deficits in Autism Spectrum Disorder: Potential Biological Origins and

Progress in Developing Therapeutic Agents. *CNS Drugs* **32**, 713–734 (2018).

23. Dodell-Feder, D., Tully, L. M. & Hooker, C. I. Social impairment in schizophrenia: new approaches for treating a persistent problem. *Curr. Opin. Psychiatry* **28**, 236–242 (2015).
24. Lacar, B. *et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7**, 11022 (2016).
25. Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).
26. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (PMC2857523) (2009).
27. Lister, R. & Mukamel, E. A. Turning over DNA methylation in the mind. *Front. Neurosci.* **9**, 1–11 (2015).
28. Yang, S.-T., Shi, Y., Wang, Q., Peng, J.-Y. & Li, B.-M. Neuronal representation of working memory in the medial prefrontal cortex of rats. *Mol. Brain* **7**, 61 (2014).
29. Lee, E. *et al.* Enhanced Neuronal Activity in the Medial Prefrontal Cortex during Social Approach Behavior. *J. Neurosci.* **36**, 6926–6936 (2016).
30. Brown, B. S. & Yu, S. P. Modulation and genetic identification of the M channel. *Prog. Biophys. Mol. Biol.* **73**, 135–166 (2000).
31. Schwake, M., Pusch, M., Kharkovets, T. & Jentsch, T. J. Surface expression and single channel properties of KCNQ2/KCNQ3, M-type K+ channels involved in epilepsy. *J. Biol. Chem.* **275**, 13343–13348 (2000).
32. He, Y. *et al.* Spatiotemporal DNA Methylome Dynamics of the Developing Mammalian Fetus. *bioRxiv* 166744 (2017) doi:10.1101/166744.
33. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536 (2008).

34. Lavery, L. A. *et al.* Loss of Dnmt3a dependent methylation in inhibitory neurons impairs neural function through a mechanism that impacts Rett syndrome. *bioRxiv* 815639 (2019) doi:10.1101/815639.

35. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* (2014).

36. Visel, A. *et al.* A High-Resolution Enhancer Atlas of the Developing Telencephalon. *Cell* (2013).

37. Wang, Z. *et al.* BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).

38. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).

39. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).

40. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).

41. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).

42. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).

43. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).

44. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).

45. Manzo, M. *et al.* Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* **36**, 3421–3434 (2017).

46. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).

47. Li, Y. *et al.* Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* **19**, 18 (2018).

48. Johnson, B. S. *et al.* Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nat. Med.* **23**, 1203–1214 (2017).

49. Clemens, A. W. *et al.* MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. *Mol. Cell* (2019) doi:10.1016/j.molcel.2019.10.033.

50. Lynch, M. D. *et al.* An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* **31**, 317–329 (2012).

51. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).

52. Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).

53. Jermann, P., Hoerner, L., Burger, L. & Schübeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3415–21 (2014).

54. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* **39**, 232–236 (2007).

55. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39**, 157–158 (2007).

56. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).

57. Perino, M. *et al.* MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* **50**, 1002–1010 (2018).

58. Luo, C. *et al.* Global DNA methylation remodeling during direct reprogramming of fibroblasts to neurons. *Elife* **8**, (2019).

59. Cenquizca, L. A. & Swanson, L. W. Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Res. Rev.* **56**, 1–26 (2007).

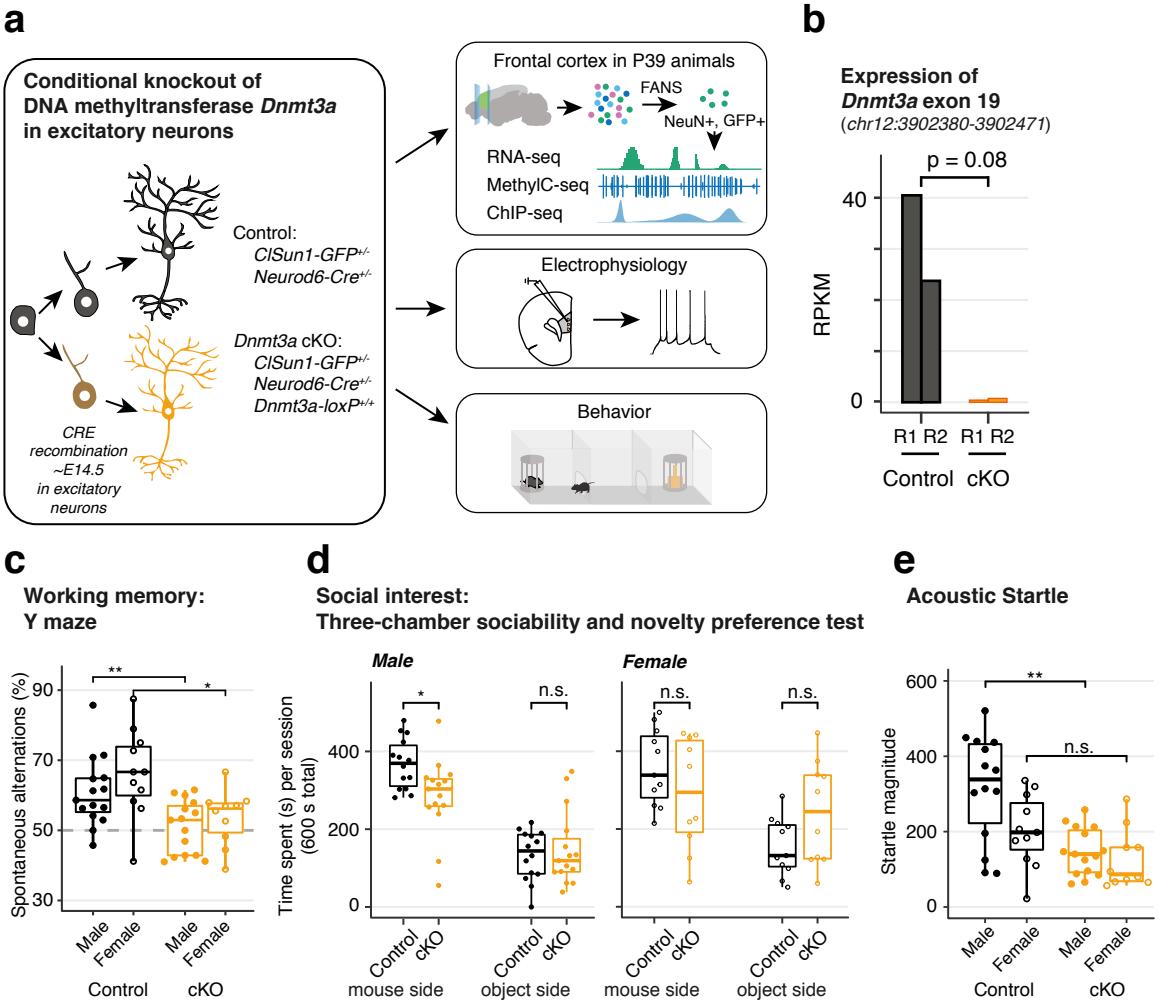
60. Spruston, N. Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* **9**, 206–221 (2008).

61. Vertes, R. P. Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. *Neuroscience* **142**, 1–20 (2006).

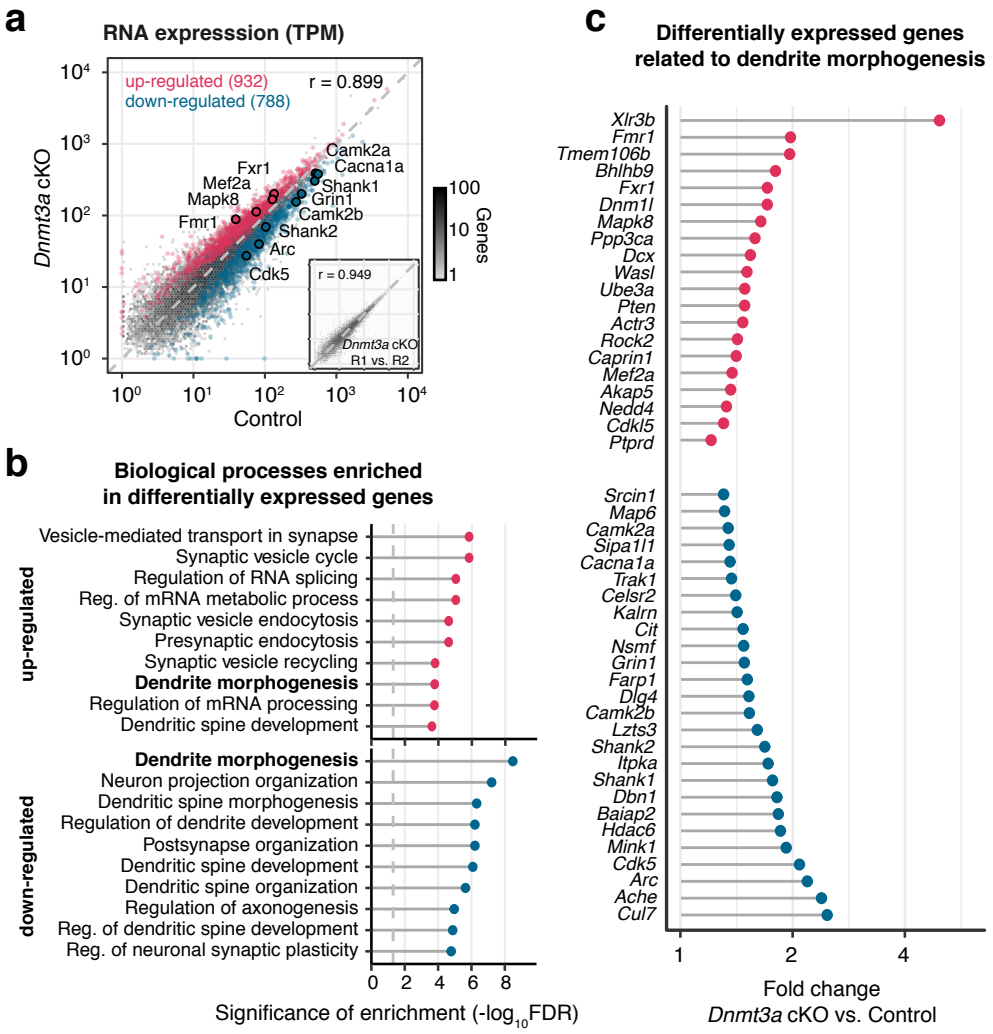
62. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).

63. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233–247.e17 (2016).

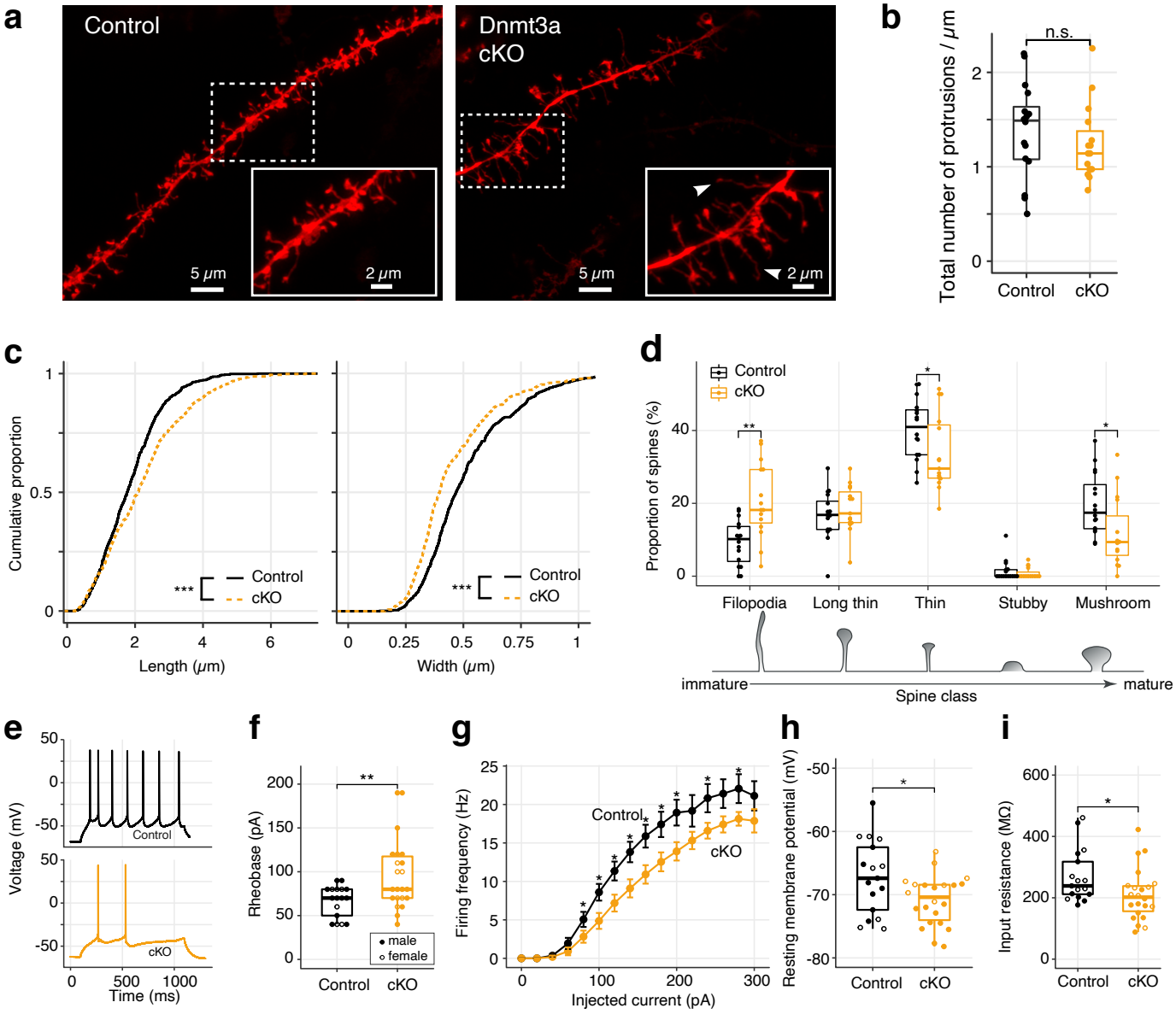
Li *et al.*, Figure 1



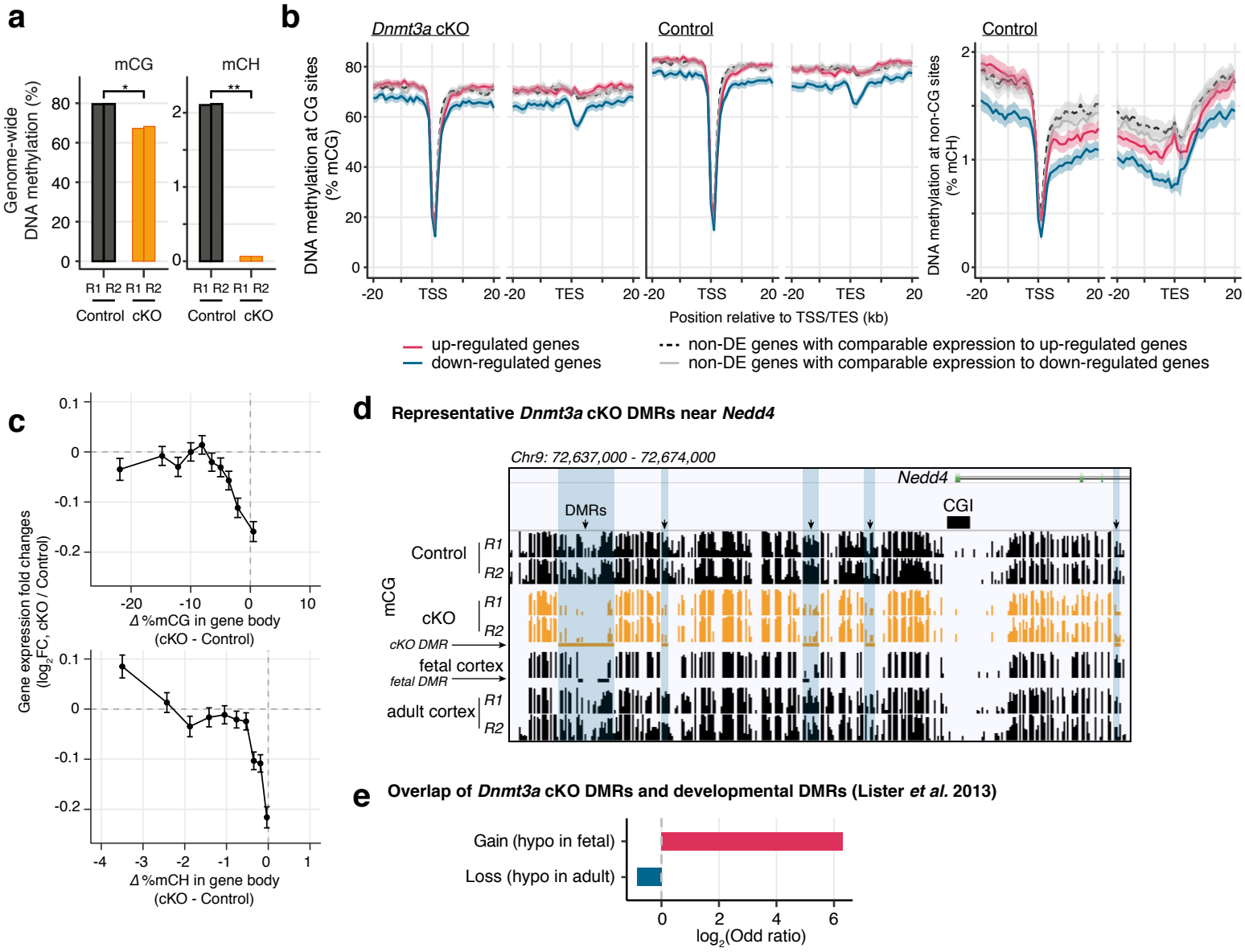
Li *et al.*, Figure 2



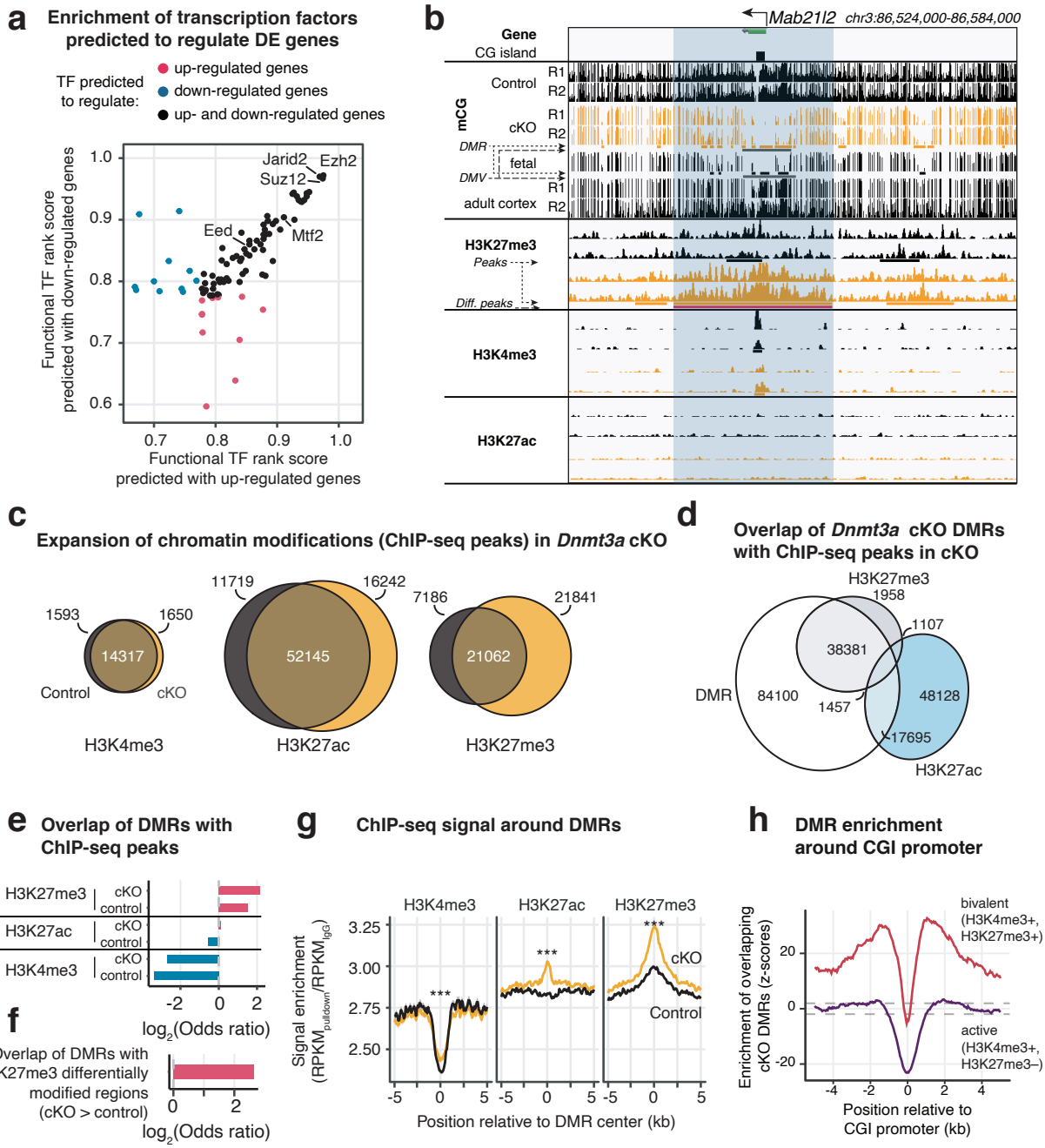
Li *et al.*, Figure 3



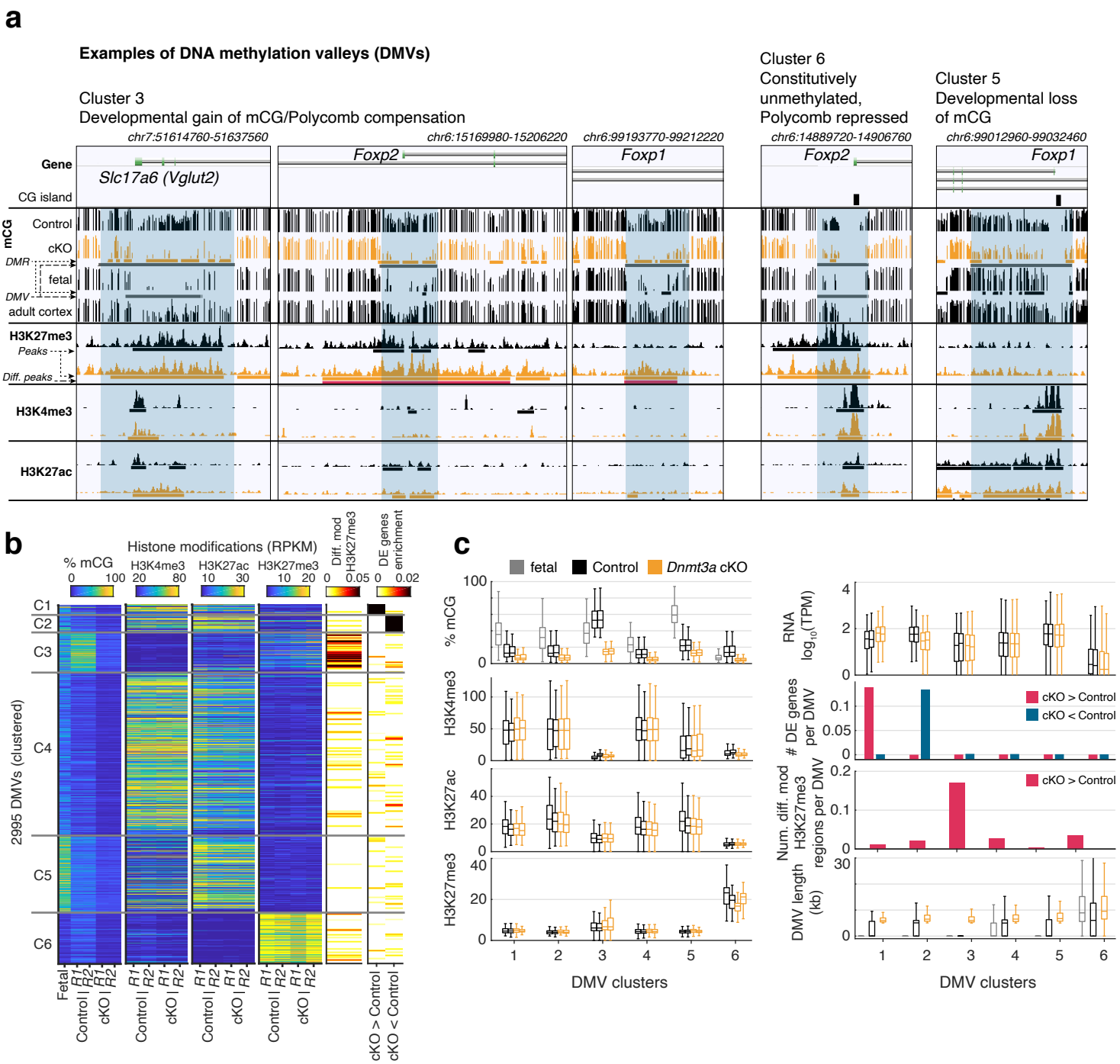
Li *et al.*, Figure 4



Li *et al.*, Figure 5



Li *et al.*, Figure 6



Supplemental methods for:

Polycomb-mediated repression compensates for loss of postnatal DNA methylation in excitatory neurons

Junhao Li[1]*, Antonio Pinto-Duarte[2]*, Mark Zander[3], Chi-Yu Lai[2], Julia Osteen[2], Linjing Fang[4], Chongyuan Luo[3,7], Jacinta D. Lucero[2], Rosa Gomez-Castanon[3], Joseph R. Nery[3], Isai Silva-Garcia[2], Yan Pang[2], Terrence J. Sejnowski[2], Susan B. Powell[5], Joseph R. Ecker[3,6†], Eran A. Mukamel[1†], M. Margarita Behrens[2,5†]

Supplemental Methods

Generation of the *Dnmt3a* cKO mice line

All animal procedures were conducted in accordance with the guidelines of the American Association for the Accreditation of Laboratory Animal Care and were approved by the Salk Institute for Biological Studies Institutional Animal Care and Use Committee. For behavior, slice physiology and spine analyses, *Dnmt3a*-floxed animals[1] (backcrossed to C57BL/6 for at least seven generations) were crossed to *Nex*-Cre (*Neurod6*-Cre)[2] mice to generate *Dnmt3a*-KO animals carrying the deletion only in pyramidal cells. To be able to isolate pyramidal neuron nuclei for DNA methylation, transcription, and ChIP analyses, the mouse lines (*Dnmt3a*-KO and *Nex*-Cre) were crossed to a mouse line carrying the INTACT background (B6.129-*Gt(ROSA)26Sor*$^{tm5(CAG-Sun1/sfGFP)Nat/MmbeJ}$). The deletion of *Dnmt3a* from pyramidal cells was

confirmed by RNA-seq (deletion of exon 19) and Western blot (Fig. 1b, Supplementary Fig. 1b, c). For both backgrounds, Nex-Cre hemizygous mice were used as controls.

Frontal cortex dissection, Nuclei isolation, and flow cytometry

Frontal cortex tissue was produced as described[3,4] from postnatal day 39 (P39) *Dnmt3a* cKO and control animals. The nuclei of GFP-expressing NeuN-positive excitatory neurons were isolated and collected using fluorescence activated nuclei sorting (FANS) as described[3,4] with the following modification: prior to FANS, nuclei were labeled with anti-NeuN-AlexaFluor647 and anti-GFP-AlexaFluor488. Nuclei were sorted as described[3]. Double positive nuclei were retained for RNA-seq, ChIP-seq and MethylC-seq library preparation and sequencing.

Western blot

Frontal cortex proteins were obtained by homogenization in RIPA buffer of the following composition: 150 mM NaCl, 10 mM Na2HPO4, 1% NaDOC, 1% NP-40, 0.5% SDS, 1 mM DTT, 1 mM PMSF in DMSO, supplemented with protease inhibitor (Sigma-Aldrich #11836153001) and phosphatase inhibitor (Pierce #A32957) cocktails. After centrifugation at 15000 x g, supernatants were preserved and protein concentration was determined by the BCA method (Pierce). Protein bands were separated in 8% PAGE-gels and transferred to nitrocellulose membranes. After blocking in TBS-tween with 5% milk, DNMT3A was detected by the use of anti-DNMT3A antibody (Abcam) and chemiluminescence. DNMT3A bands were normalized to ACTIN content in each sample.

RNA extraction, RNA-seq library construction and sequencing

Nuclei (between 50,000-60,000) were used to isolate RNA using the Single-Cell RNA Purification Kit (Norgen, catalog# 51800). In brief, aliquots of nuclei were resuspended in 350 µl of RL buffer (Norgen) and passed through an 18G syringe five times. RNA extraction, including DNase digestion, followed

manufacturers' instructions. RNA was eluted in 20 µl of Elution Solution A (Norgen). The nuclear RNA concentration was determined using TapeStation (Agilent). RNA was diluted to 1 ng/µl and a total of 5 ng processed for RNA-seq library preparation. RNA libraries were prepared using NuGen Ovation® RNA−Seq System V2 (#7102-32) for cDNA preparation following the product manual. cDNA purification was done using Zymo Research DNA Clean & Concentrator™-25 with modification from Ovation protocol. cDNA, eluted in 30-40 µl of TE (1 µg per sample), was fragmented at 300bp using Covaris S2 (Sonolab S-series V2), followed by library preparation according to KAPA LTP Library Preparation Kit (KK8232), using Illumina indexed adapters.

DNA extraction

DNA extraction was performed using the Qiagen DNeasy Blood and Tissue kit (catalog #69504) and eluted into 50–100 µL AE.

Genomic DNA library construction and sequencing

1.5 µg of genomic DNA was fragmented with a Covaris S2 (Covaris, Woburn, MA) to 400 bp, followed by end repair and addition of a 3' A base. Cytosine-methylated adapters provided by Illumina (Illumina, San Diego, CA) were ligated to the sonicated DNA at 16˚C for 16 hours with T4 DNA ligase (New England Biolabs). Adapter-ligated DNA was isolated by two rounds of purification with AMPure XP beads (Beckman Coulter Genomics, Danvers, MA). Half of adapter-ligated DNA molecules were enriched by 6 cycles of PCR with the following reaction composition: 25 µL of Kapa HiFi Hotstart Readymix (Kapa Biosystems, Woburn, MA) and 5 µl TruSeq PCR Primer Mix (Illumina) (50 µl final). The thermocycling parameters were: 95˚C 2 min, 98˚C 30 sec, then 6 cycles of 98˚C 15 sec, 60˚C 30 sec and 72˚C 4 min, ending with one 72˚C 10 min step. The reaction products were purified using AMPure XP beads and size selection done from 400 – 600 bp. Libraries were sequenced on Illumina HiSeq 4000.

MethylC-seq library construction and sequencing

MethylC-seq libraries were prepared as previously described[5]. All DNA obtained from the extraction was spiked with 0.5% unmethylated Lambda DNA. The DNA was fragmented with a Covaris S2 (Covaris, Woburn, MA) to 200 bp, followed by end repair and addition of a 3' A base. Cytosine-methylated adapters provided by Illumina (Illumina, San Diego, CA) were ligated to the sonicated DNA at 16˚C for 16 hours with T4 DNA ligase (New England Biolabs). Adapter-ligated DNA was isolated by two rounds of purification with AMPure XP beads (Beckman Coulter Genomics, Danvers, MA). Adapter-ligated DNA (≤450 ng) was subjected to sodium bisulfite conversion using the MethylCode kit (Life Technologies, Carlsbad, CA) as per manufacturer's instructions. The bisulfite-converted, adapter-ligated DNA molecules were enriched by 8 cycles of PCR with the following reaction composition: 25 µL of Kapa HiFi Hotstart Uracil+ Readymix (Kapa Biosystems, Woburn, MA) and 5 µl TruSeq PCR Primer Mix (Illumina) (50 µl final). The thermocycling parameters were: 95˚C 2 min, 98˚C 30 sec, then 8 cycles of 98˚C 15 sec, 60˚C 30 sec and 72˚C 4 min, ending with one 72˚C 10 min step. The reaction products were purified using AMPure XP beads. Up to two separate PCR reactions were performed on subsets of the adapter-ligated, bisulfite-converted DNA, yielding up to two independent libraries from the same biological sample. MethylC-seq libraries were sequenced on Illumina HiSeq 4000.

ChIP-seq library construction and sequencing

Sorted nuclei were crosslinked for 15min in 1% formaldehyde solution and quenched afterward with glycine at a final concentration of 0.125 M. After crosslinking, nuclei were sonicated in Lysis buffer (50 mM Tris HCl pH 8, 20 mM EDTA, 1% SDS, 1X EDTA free protease inhibitor cocktail). ChIP assays were conducted with antibodies against H3K27me3 (39156, Active Motif), H3K27ac (39133, Active Motif) and H3K4me3 (04-745, Millipore Sigma). Mouse IgG (015-000-003, Jackson ImmunoResearch) served as negative control. H3K4me3 ChIP-seq assays were conducted with 100K nuclei and 500K nuclei

were used for H3K27me3 and H3K27ac ChIP-seq assays. The respective antibodies and IgG were coupled for 4-6 hours to Protein G Dynabeads (50 µl, 10004D, Thermo Fisher Scientific). Equal amounts of sonicated chromatin were diluted with 9 volumes of Binding buffer (1% Triton X-100, 0.1% Sodium Deoxycholate, 2X EDTA free protease inhibitor cocktail) and subsequently incubated overnight with the respective antibody-coupled Protein G beads. Beads were washed successively with low salt buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.5% Triton X-100), high salt buffer (50 mM Tris HCl pH 7.4, 500 mM NaCl, 2 mM EDTA, 0.5% Triton X-100) and wash buffer (50 mM Tris HCl pH 7.4, 50 mM NaCl, 2 mM EDTA) before de-crosslinking, proteinase K digestion and DNA precipitation. Libraries were generated with the Accel-NGS 2S Plus DNA Library Kit (21024, Swift Biosciences) and sequenced on the Illumina HiSeq 4000 Sequencing system.

RNA-seq data processing

RNA-seq reads were first undergone quality control using FastQC[6] (v0.11.8, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and then were trimmed to remove sequencing adapters and low-quality sequences (minimum Phred score 20) using Trim Galore (v0.5.0, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, a wrapper tool powered by Cutadapt[7] v1.16) in the paired-end mode. Clean reads were then mapped to the mouse mm10 (GRCm38) genome and the GENCODE annotated transcriptome (release M10) with STAR[8] (Spliced Transcripts Alignment to a Reference, v2.5.1b). Gene expression was estimated using RSEM[9] (RNA-Seq by Expectation Maximization, v1.2.30). Gene-level "expected count" from the RSEM results were rounded and fed into edgeR[10] (v3.24.1) to call differentially expressed genes. Only genes that were expressed (with counts per million > 2) in at least two samples were kept. These counts were then normalized using the TMM

method[11], and DE genes were then called in the quasi-likelihood F-test mode, requiring FDR < 0.05 and FC > 20% (Supplementary Table 2).

Enrichment test of Gene Ontology (GO) terms in differentially expressed genes

GO enrichment analysis was performed using clusterProfiler[12] (v3.10.0). Only "Biological Process" terms with no less than 10 genes and no more than 250 genes were considered. Terms with FDR < 0.05 were considered significantly enriched.

Genomic DNA sequencing data processing and SNP calling

To estimate the completeness of the inbreeding of the mouse strains, and to avoid incorrect cytosine context assignment in the following methylC-seq data processing, we used the genomic DNA sequencing data of both the *Dnmt3a* cKO and the control animals to call SNPs against the mouse mm10 genome. We followed the GATK "best practices for germline SNPs and indels in whole genomes and exomes" pipeline[13–15]. Briefly, raw data were first trimmed to remove sequencing adapters and low-quality sequences (minimum Phred score 20) using Trim Galore in the paired-end mode. Clean data were then mapped to the mm10 genome using BWA[16] (v0.7.13-r1126). Duplicates reads were marked with Picard[17]. Then the analysis-ready reads were fed into GATK (v3.7) to perform two rounds of joint genotyping and base recalibration. Variants were then filtered using the following criteria: QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 4.0 . By that, we identified 548,530 and 507,669 SNPs (relative to mm10) in the *Dnmt3a* cKO and the control animals, respectively. At last, we created a substituted genome to mask out all these SNPs (replaced with Ns) with the maskfasta tool in the

BEDTools suite[18] (v2.27.1). This substituted genome was used in the following methylC-seq data processing pipeline.

MethylC-seq data processing

MethylC-seq reads were processed using the methylpy pipeline (v1.3.2, https://github.com/yupenghe/methylpy) as previously described[3,19]. Briefly, computationally bisulfite-converted genome index was built using the aforementioned substituted genome file appended with the lambda phage genomic sequence. MethylC-seq raw reads were first trimmed to remove sequencing adapters and low-quality sequences (minimum Phred score 10) using Cutadapt in paired-end mode. To acquire higher mappability, we treated the two ends of the clean reads as they were sequenced in single-end mode, and mapped them to the converted genome index with bowtie2[20] (v2.3.0) as aligner in the single-end pipeline of methylpy. Only reads uniquely mapped were kept, and clonal reads were removed. The bisulfite non-conversion rate was estimated using the spiked-in unmethylated lambda phage DNA. For each cytosine, a binomial test was performed to test whether the methylation levels are significantly greater than 0 with an FDR threshold of 0.01.

For a particular genomic region, the raw methylation level for a given cytosine context (CG or CH) was defined as:

$$\%mC = 100 \times \frac{m}{h},$$

where m is the total number of methylated based calls within the region, and h is the total number of covered based calls within the region. Methylation levels were then corrected for non-conversion rate (NCR) using the following maximum likelihood formula:

$$\%mC_adj = \frac{\%mC - \%NCR}{100 - \%NCR}, \text{ where } \%mC_adj \in [0, 100].$$

When profiling the methylation landscapes around differentially expressed genes, we selected control genes with comparable gene expression using the R package MatchIt[21] (v3.0.2). These control genes were defined with nearest neighbor matching of the expression using logistic link propensity score as a distance measure, requiring the standard deviation of the distance to be less than 0.01.

Differentially methylated regions (DMRs) calling

CG DMRs were identified using a previously reported method[22–24], which is implemented in the DMRfind function in methylpy. We required at least 3 differentially methylated sites within a particular DMR, and the maximum distance two significant sites can be to be included in the same DMR to be 500 nucleotides. With an FDR cutoff of 0.01 and a post-filtering cutoff of methylation levels change greater than 30%, we found 141,633 *Dnmt3a* cKO hypo-DMRs (Supplementary Table 4). Note that with these criteria we also found 19 *Dnmt3a* cKO hyper-DMRs, which we thought were noise and/or SNPs failed to be detected by the masking pipeline described earlier. Therefore we removed these hyper-DMRs from further consideration.

Enrichment test of DMRs and other genomic regions

To test whether DMRs were significantly enriched in certain genomic features, we use methods adapted from a recent report[25]. Briefly, for each genomic feature, we constructed a 2 by 2 contingency table of (n_{11}, n_{12}, n_{21}, n_{22}), where:

n_{11} is the number of CG in DMRs that were inside the feature;

n_{12} is the number of CG in DMRs that were outside of the feature;

n_{21} is the number of CG not in DMRs that were inside in feature;

n_{22} is the number of CG not in DMRs that were outside of the feature.

The total number of CG sites in consideration was the number of autosomal CG in the reference genome. Counting number of CG rather than the number of DMRs or bases accounts for the non-uniform distribution of CG along the genome and avoids double-counting DMRs that are both inside and outside of the feature.

With this contingency table, we estimated the enrichment log odd ratio (OR) along with its standard error (se) and 95% confidence interval (ci) with the following formulas:

$$\log_2 OR = \log_2 n_{11} + \log_2 n_{12} - \log_2 n_{21} - \log_2 n_{22}$$

$$se(\log_2 OR) = \sqrt{1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}}$$

$$ci(\log_2 OR) = [\log_2 OR - 2\times se(\log_2 OR),\ \log_2 OR + 2\times se(\log_2 OR)]$$

P-value from performing Fisher's exact test for testing the null of independence of rows and columns in the contingency table (the null of no enrichment or depletion) was computed using the fisher.test() function in R.

The genomic regions/features used in these enrichment tests includes: a list of developmental DMRs that gain or lose methylation during development[3]; gene features (genic, exonic, intronic, promoter, 5'UTR, 3'UTR, intergenic) based on the Gencode vM10 annotation (promoters were defined as the $\pm$2kb regions around transcription start sites); CpG island (CGI) related features based on the "cpgIslandExt" annotation from the UCSC genome browser[26,27] (http://genome.ucsc.edu/index.html), where CGI shores

were defined as CGI ±2kb, CGI shelves were defined as ±2-4kb of CGI and open seas were defined as regions that were at least 4kb away from any CGI; the 12 states of the chromatin states map in mouse embryonic stem cell[28] (https://github.com/guifengwei/ChromHMM_mESC_mm10) generated by ChromHMM[29,30] using ChIP-seq data from the ENCODE project[31,32]; the H3K4me3, H3K27ac and H3K27me3 peaks and the H3K27me3 differentially binding regions generated with our ChIP-seq data (see the "ChIP-seq data processing" section).

Predicting functional transcription factors regulating the differentially expressed genes with BART

The lists of up-regulated and down-regulated genes were fed into BART[33,34] (Binding Analysis for Regulation of Transcription) separately to predict functional transcription factors (TFs) and chromatin regulators that bind at cis-regulatory regions of the DE genes. To make a better visualization we transformed the relative rank (a metric generated by BART to represent the average rank of Wilcoxon P-value, Z-score and max AUC for each factor divided by the total number of factors) into the functional TF rank score (which is simply 1 minus the relative rank) so that the higher the rank score the more possible the TF regulates the DE genes. The integrative rank significance was estimated with the Irwin-Hall p-value. TFs with Irwin-Hall p-value < 0.05 were considered significant.

Enrichment test of known transcription factor binding motifs in DMRs

We used the "findMotifsGenome.pl" tool in HOMER[35] (Hypergeometric Optimization of Motif EnRichment, v4.8.3) to find known transcription factor binding motif in the *Dnmt3a* cKO DMRs. The parameters used are as follows: "-size 500 -len 8,10,12 -S 25 -fdr 100 -p 10 -mset vertebrates -bits -gc -nlen 3 -nomotif". A set of non-neural DMRs[36] was used as the background.

ChIP-seq data processing

ChIP-seq reads were pre-processed with the ENCODE Transcription Factor and Histone ChIP-Seq processing pipeline (https://github.com/ENCODE-DCC/chip-seq-pipeline2, v1.1.6). Briefly, paired-end reads were mapped to the mm10 genome with BWA[16] (v0.7.13-r1126). Reads were then filtered using samtools[37] (v1.2) to remove unmapped, mate unmapped, not primary alignment and duplicates reads (-F 1804). Properly paired reads were retained (-f 2). Multi-mapped reads (MAPQ < 30) were removed. PCR duplicates were removed using the MarkDuplicates tool in Picard[17] (v2.10.6). Reads mapped to the blacklist regions[31] in mouse mm10 genome (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were also removed.

Peak calling was performed using epic2[38] (v0.0.16), a reimplementation of SICER[39]. For H3K4me3 and H3K27ac, we used the following parameters: "--bin-size 200 --gaps-allowed 1". For H3K27me3, we used the following parameters: "--bin-size 200 --gaps-allowed 3". The IgG sample was used as a control.

Differentially modified regions of the histone modification ChIP-seq data was called using DiffBind[40] (v2.10.0) in DESeq2[41] (v1.22.1) mode. Regions with FDR < 0.05 were considered significant. Genes associated with these regions were identified using GREAT[42] (v3.0.0) with the "Basal plus extension" association rule with default parameters. GO enrichment analysis of the associated genes was performed with GREAT, and we considered the GO biological process terms with hypergeometric test FDR < 0.05 as significant.

Definition of bivalent and active CGI promoters

CGI promoters were defined as CpG islands (downloaded from the UCSC genome browser) that overlapping with promoters (±2kb regions around transcription start sites annotated in Gencode vM10). These CGI promoters were further tested to see whether they overlapped with the ChIP-seq peaks of H3K4me3 and H3K27me3. Bivalent CGI-promoters were defined as CGI-promoters that overlapped with

both the H3K4me3 and H3K27me3 peaks, whereas active CGI-promoters were defined as CGI-promoters that overlapped with only the H3K4me3 peaks but not the H3K27me3 peaks.

Identification of DNA methylation valleys (DMVs)

To find DMVs, we first identified UMRs (undermethylated regions) and LMRs (low methylated regions) using MethylSeekR[43] (v1.22.0) with m = 0.5, n = 7 and FDR < 0.05. Noted that MethylSeekR did not recover any PMDs (partially methylated domains). DMVs were then defined as UMRs that with length ≥ 5 kb and mean methylation level ≤ 15%. To compare DMVs identified in the *Dnmt3a* cKO and the control samples, we further grouped these DMVs into 5 categories, namely consistent (exact same DMV in the two conditions), cKO unique, control unique, expanded (wider in the cKO) and shrunken (wider in the control).

DMR enrichment around CGI promoter

We used regioneR[44] (v1.14.0) to test whether two sets of genomic regions had significantly higher numbers of overlaps compared to expected by chance. We used permTest() to perform the permutation test, and used the randomizeRegions() function to generate the shuffled control for 5,000 times, where the query regions were randomly placed along the genome independently while maintaining their size. The strength of the association of the two sets of regions was estimated using z-score, the distance (measured in standard deviation) between the expected overlaps in the shuffled control and the observed overlaps, and the p-value was reported. To check if the association was specifically linked to the exact position of the query regions, we used the localZscore() function with window = 5000 and step = 50, which shifted the query regions and estimated how the value of the z-score changed when moving the regions.

Other tools used in the data analysis

Browser representations were created using AnnoJ[45]. All analyses were conducted in R (v3.5.0) or Matlab 2017a. Genomic ranges manipulation was done either with bedtools[18] or GenomicRanges[46]. Multiple comparison correction for p-values was performed with the Benjamini-Hochberg FDR method[47]. Results with FDR < 0.05 were considered significant except where stated otherwise.

Patch-clamp electrophysiology

Male and female mice (6-9 weeks) were anesthetized with isoflurane and decapitated. The brains were quickly removed and coronal slices of the frontal cortex containing the prelimbic region (~2 mm anterior to bregma) were cut in ice-cold slicing medium of the following composition (in mM): 110 sucrose, 2.5 KCl, 0.5 $CaCl_2$, 7 $MgCl_2$, 25 $NaHCO_3$, 1.25 NaH_2PO_4, and 10 glucose (bubbled with 95% O_2 and 5% CO_2). The slices were then transferred to artificial CSF (aCSF) containing (in mM) 130 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 23 $NaHCO_3$, 1.3 $MgCl_2$, 2 $CaCl_2$ and 10 glucose, equilibrated with 95% O_2 and 5% CO_2 at 35°C for 30 minutes and afterward maintained at room temperature (22-24 °C) for at least 1 hour (patch-clamp recording) before use. Brain slices were then transferred to a recording chamber and kept minimally submerged under continuous superfusion with aCSF at a flow rate of ~2 ml/min. Whole-cell recordings were obtained from putative prelimbic layer 2 (L2) pyramidal cells (identified by their pyramidal-shaped cell bodies and long apical dendrite using an upright microscope equipped with differential interference contrast optics). In acute mPFC slices, the prelimbic L2 is clearly distinguishable from L1 and L3 as a thin dark band that is densely packed with neuron somata. Pipettes had a tip resistance of 4-8 MΩ when filled with an internal solution of the following composition (in mM): 125 K-gluconate, 15 KCl, 8 NaCl, 10 HEPES, 2 EGTA, 10 Na_2Phosphocreatine, 4 MgATP, 0.3 NaGTP (pH 7.25 adjusted with KOH, 290-300 mOsm). Access resistance (typically 15-35 MΩ) was monitored throughout the experiment to ensure stable recordings.

After obtaining the whole-cell configuration in voltage clamp mode, cells were switched from a holding potential of -70 mV to current-clamp mode and the bridge-balance adjustment was performed. Passive electrical properties were quantified from recordings with hyperpolarizing current injections that evoked small ~5 mV deflections in membrane potential from resting. Responses to stepwise current injections (10-300 pA in increments of 10 pA; duration, 1 s) were recorded at 20 kHz in order to calculate input-output curves and rheobase - the minimal current necessary to trigger the first action potential. Miniature excitatory postsynaptic currents (mEPSCs) were recorded for 5 minutes in voltage-clamp mode (Vh=-70 mV) in the presence of the Na+ channel blocker, TTX (0.5 μM), to prevent the generation of action potentials, and picrotoxin (50 μM), an antagonist of $GABA_A$ receptors, to minimize inhibitory responses. In these conditions, mEPSCs could be blocked by the AMPA receptor antagonist, CNQX (25 μM). Single events larger than 6 pA were detected off-line using the Minianalysis program (Synaptosoft Inc. Decatur, GA). All data were acquired using a Multiclamp 700B amplifier and pCLAMP 9 software (Molecular Devices, LLC. San Jose, CA).

Fluorescent labeling of dendritic spines

Coronal brain slices containing the mPFC of 10-13 weeks female mice were prepared as for electrophysiological recordings and placed in a beaker for 3 hours at room temperature (24° C) to allow functional and morphological recovery. One slice was then transferred to a recording chamber and kept minimally submerged under continuous superfusion with aCSF bubbled with carbogen (95% O_2 / 5% CO_2) at a flow rate of ~2 ml/min. Previously sonicated crystals of the fluorescent marker DiI were placed next to the somata of layer 2 neurons in the prelimbic cortex, identified with the aid of an upright microscope equipped with differential interference contrast optics. The mice in these experiments had a Thy1-YFP background to help rule out non specific labeling of deeper layer neurons. The neurons were exposed to the DiI crystals for 60 minutes. The slices were then gently removed from the incubation

chamber with a transfer pipette and immersed in fixative (4% PFA) for 30 minutes. Then, the slices were rinsed three times with PBS for 5–10 minutes each, after which they were mounted on slides with prolong gold antifade mounting medium (Life Technologies – Molecular Probes). The slides were kept in a dark box for 24 h at room temperature to allow the liquid medium to form a semi-rigid gel. Imaging took place 24-48 h from the time of the initial staining.

Confocal imaging

Dendritic spines were imaged by an investigator blind to the genotype using a Zeiss AiryScan confocal laser scanning microscope. All images were taken using the Zeiss Plan-APOCHROMAT 63× oil-immersion lens (N/A 1.4). A 543nm laser was used to visualize the fluorescence emitted by DiI. Serial stack images with a 0.2 μm step size were collected, and then projected to reconstruct a three dimensional image that was post-processed by the AiryScan software. Dendritic segments in layer 1, which were derived from layer 2 pyramidal neurons retrogradely labeled with DiI and that were well separated from neighboring neural processes, were randomly sampled and imaged. Each dendritic segment imaged for quantification belonged to a different neuron.

Dendritic spine quantification

The z-stack series were imported into the Reconstruct software (http://synapseweb.clm.utexas.edu/software-0/), with which a second investigator also blind to the genotype performed the identification of dendritic spines and their morphometric analysis. By scrolling through the stack of different optical sections, individual spine heads could be identified with greater certainty. All dendritic protrusions with a clearly recognizable stalk were counted as spines. Spine density was determined by summing the total number of spines per dendritic segment length (30-40 μm) and then calculating the average number of spines per μm. Individual dendritic spines were classified in the following order according to pre-established criteria: protrusion longer than 3 μm, filopodia; head wider

than 0.6 μm, mushroom; protrusion longer than 2 μm and head narrower than 0.6 μm, long-thin; protrusion longer than 1 μm and head narrower than 0.6 μm, long-thin; the remaining spines were labeled stuby). Branched spines (with more than one neck) were counted separately.

BEHAVIORAL TESTING

Phenotypic characterization was initiated when the animals reached 9 weeks of age using cohorts of 10-15 male or female mice per genotype, according to the order described below.

Open field test

The open field test was performed using MED Associates hardware and the Activity Monitor software according to the manufacturer's instructions (MED Associates Inc, St. Albans, VT). Animals were individually placed into clear Plexiglas boxes (43.38 x 43.38 x 30.28 cm) surrounded by multiple bands of photo beams and optical sensors that measure horizontal and vertical activity. Movement was detected as breaks within the beam matrices and automatically recorded for 60 minutes.

Light/dark transfer test

The light/dark transfer procedure was used to assess anxiety-like behavior in mice by capitalizing on the conflict between exploration of a novel environment and the avoidance of a brightly lit open field (150-200 lux in our experiments). The apparatus were Plexiglas boxes as for the open field test (43.38 x 43.38 x 30.28 cm) containing dark box inserts (43.38 x 12.8 x 30.28 cm). The compartments were connected by an opening (5.00 x 5.00 cm^2) located at floor level in the center of the partition. The time spent in the light compartment was used as a predictor of anxiety-like behavior, i.e. a greater amount of time in the light compartment was indicative of decreased anxiety-like behavior. Mice were placed in the dark compartment (4-7 lux) at the beginning of the 15 minute test.

Elevated plus maze

The maze consisted of four arms (two open without walls and two with enclosed walls) 30 cm long and 5 cm wide in the shape of a plus sign. The apparatus was elevated approximately 33 cm over a table. At the beginning of each trial, one animal was placed inside a cylinder located at the center of the maze for 1 min. The mouse was then allowed to explore the maze for 5 minutes. The session was video-recorded by an overhead camera and subjected to automated analysis using ANY-maze software. The apparatus was wiped down with sani-wipes between trials to remove traces of odor cues. The percentage of time spent in open or closed arms was scored and used for analysis.

Y maze test for spontaneous alternations

Spontaneous alternations between three 38 cm long arms of a Y-maze were taken as a measure of working memory. Single 6 minute trials were initiated by placing each mouse in the center of the Y maze. Arm entries were recorded with a video camera and the total number of arm entries, as well as the order of entries, was determined. The apparatus was wiped down with sani-wipes between trials to remove traces of odor cues. Spontaneous alternations were defined as consecutive triplets of different arm choices and % spontaneous alternation defined as number of spontaneous alternations divided by the total number of arm entries minus 2.

Social Approach

The apparatus consisted of a Plexiglas box (60 x 38 x 23.5 cm) divided into 3 compartments by Plexiglas partitions containing openings through which the mice could freely enter the 3 chambers. The test was conducted in two 10-minute phases. In phase I, the test mouse is first allowed to explore the chambers for 10 minutes. Each of the two outer chambers contained an empty, inverted stainless steel wire cup. In phase II, the test mouse is briefly removed, and a sex-matched unfamiliar mouse, was placed under one of the wire cups and plastic blocks were placed under the other wire cup. The test mouse was then gently

placed back in the arena and given an additional 10 minutes to explore. An overhead camera and video tracking software (ANY-maze, Wood Dale, IL) were used to record the amount of time spent in each chamber. The location (left or right) of the novel object and novel mouse alternates across subjects.

Acoustic startle responses and prepulse inhibition of the acoustic startle response

Acoustic startle responses were tested inside SR-LAB startle apparatus (San Diego Instruments, San Diego, CA), consisting of an inner chamber with a speaker mounted to the wall and a cylinder mounted on a piezoelectric sensing platform on the floor. At the beginning of testing, mice were placed inside the cylinder and then were subjected to background 65 dB white noise during a 5-min acclimation period. The PPI session began with the presentation of six pulse-alone trials of 120 dB, 40 ms. Then, a series of pulse-alone trials and prepulse trials (69, 73, or 81 dB, 20 ms followed by 100 ms pulse trial, 120 dB) were each presented 12 times in a pseudorandom order. The session concluded with the presentation of six pulse-alone trials. The apparatus was wiped down with sani-wipes between trials to remove traces of odor cues. The startle amplitude was calculated using arbitrary units, and the acoustic startle response was the average startle amplitude of pulse-alone trials. The percent PPI was calculated as follows: [100-(mean prepulse response/mean pulse alone response) x 100].

Cued and contextual fear conditioning

Fear conditioning experiments were performed using automated fear conditioning chambers (San Diego Instruments, San Diego, CA), similarly to previous studies[48,49]. On day 1, after a 2-min acclimation period, mice were presented with a tone conditioned stimulus (CS: 75 dB, 4 kHz) for 20 s that co-terminated with a foot shock unconditioned stimulus (1 s, 0.5 mA). A total of three tone-shock pairings were presented with an inter-trial interval of 40 s. To assess acquisition, freezing was quantified during foot shock presentations. Mice were returned to their home cages 2 min after the final shock. These moderate shock parameters were previously found suitable to detect both increases and decreases in

fear-conditioned behavior[48]. 24 h later, on day 2, mice were re-exposed to the conditioning chamber to assess context-dependent fear retention. This test lasted 8 min during which time no shocks or tones were presented and freezing was scored for the duration of the session. Time freezing was quantified across four 2-min blocks. Day 3: 24 h after the context fear-retention test, mice were tested for CS-induced fear retention and extinction. The context of the chambers was altered across several dimensions (tactile, odor, visual) for this test in order to minimize generalization from the conditioning context. After a 2 min acclimation period, during which time no tones were presented ('pre-tone'), 32 tones were presented for 20 s with an inter-trial interval of 5 s. Freezing was scored during each tone presentation and quantifications were done in eight blocks of four tones. Mice were returned to their home cage immediately after termination of the last tone. On day 4, after a 2 min acclimation period, during which time no tones were presented ('pre-tone'), a shorter session of 16 tones was used to assess extinction. Time freezing was quantified across four blocks of four tones.

References

1. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
2. Goebbels, S. *et al.* Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *Genesis* **44**, 611–621 (2006).
3. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
4. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
5. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).

6. Andrews, S. *et al.* FastQC. (2012).

7. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

8. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

9. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

10. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

11. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

12. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

13. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

14. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

15. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–33 (2013).

16. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

17. Broad Institute. Picard Tools. http://broadinstitute.github.io/picard/ (2018).

18. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

19. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**,

1369–1384 (2015).

20. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

21. Ho, D., Imai, K., King, G. & Stuart, E. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software, Articles* **42**, 1–28 (2011).

22. Ma, H. *et al.* Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511**, 177–183 (2014).

23. Schmitz, R. J. *et al.* Patterns of population epigenomic diversity. *Nature* **495**, 193–198 (2013).

24. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).

25. Rizzardi, L. F. *et al.* Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* (2019) doi:10.1038/s41593-018-0297-8.

26. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

27. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–6 (2004).

28. Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* **68**, 955–969.e10 (2017).

29. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

30. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).

31. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

32. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

33. Wang, Z. *et al.* BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).

34. Wang, S. *et al.* Modeling cis-regulation with a compendium of genome-wide histone H3K27ac profiles. *Genome Res.* **26**, 1417–1429 (2016).

35. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

36. Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).

37. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

38. Stovner, E. B. & Sætrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz232.

39. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).

40. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).

41. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

42. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

43. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).

44. Gel, B. *et al.* regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).

45. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

46. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

47. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

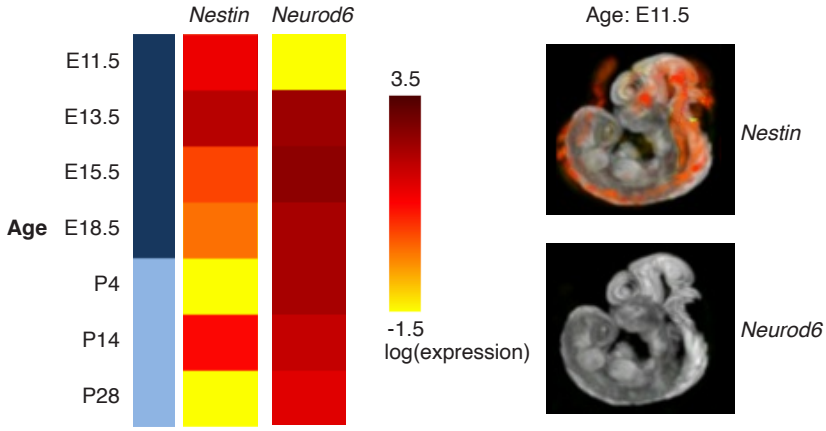
48. Risbrough, V., Ji, B., Hauger, R. & Zhou, X. Generation and characterization of humanized mice carrying COMT158 Met/Val alleles. *Neuropsychopharmacology* **39**, 1823–1832 (2014).

49. Gresack, J. E. *et al.* Isolation rearing-induced deficits in contextual fear learning do not require CRF(2) receptors. *Behav. Brain Res.* **209**, 80–84 (2010).

Li *et al.*, Supplementary Figure 1

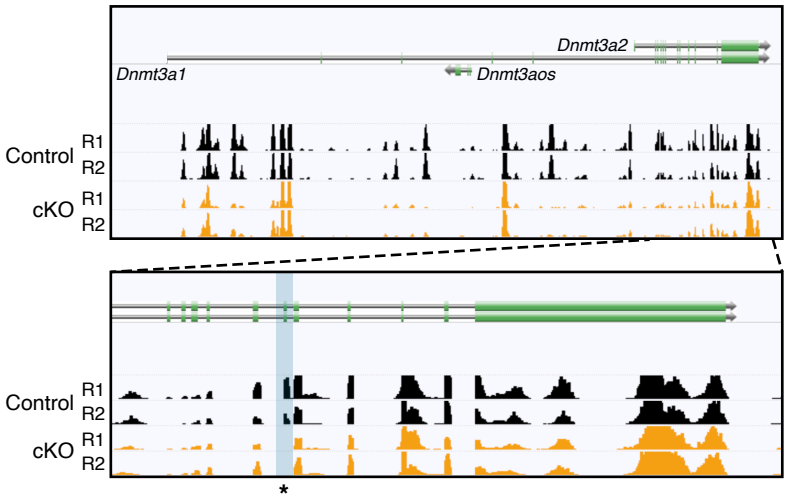
a

Expression quantification of *in situ* hybridization data in telencephalic vesicle from the Allen Developing Brain Atlas (http://developingmouse.brain-map.org/)



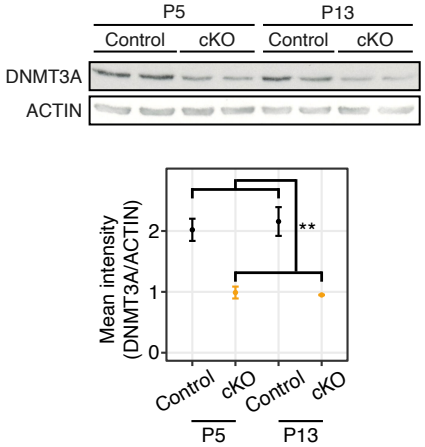
b

RNA-seq reads at the *Dnmt3a* locus (chr12:3806006-3914443)



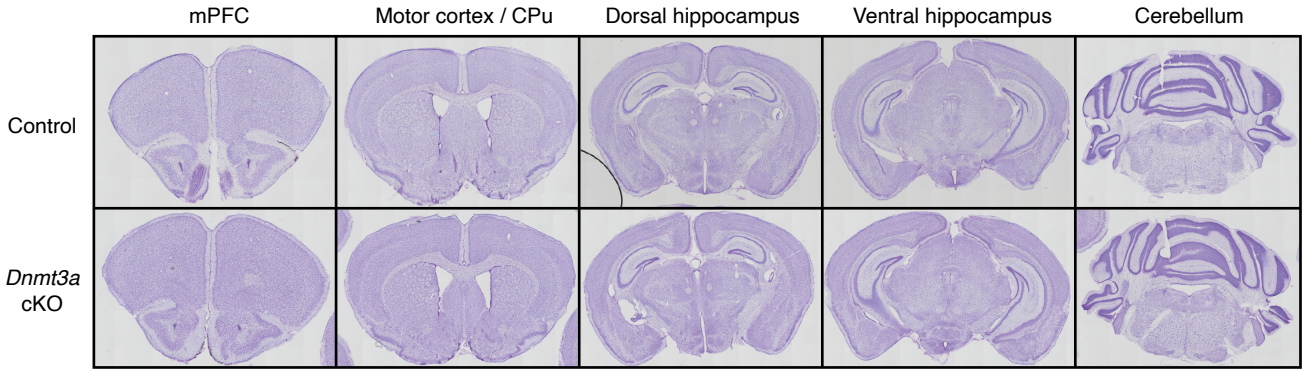
c

***Dnmt3a* protein abundance**

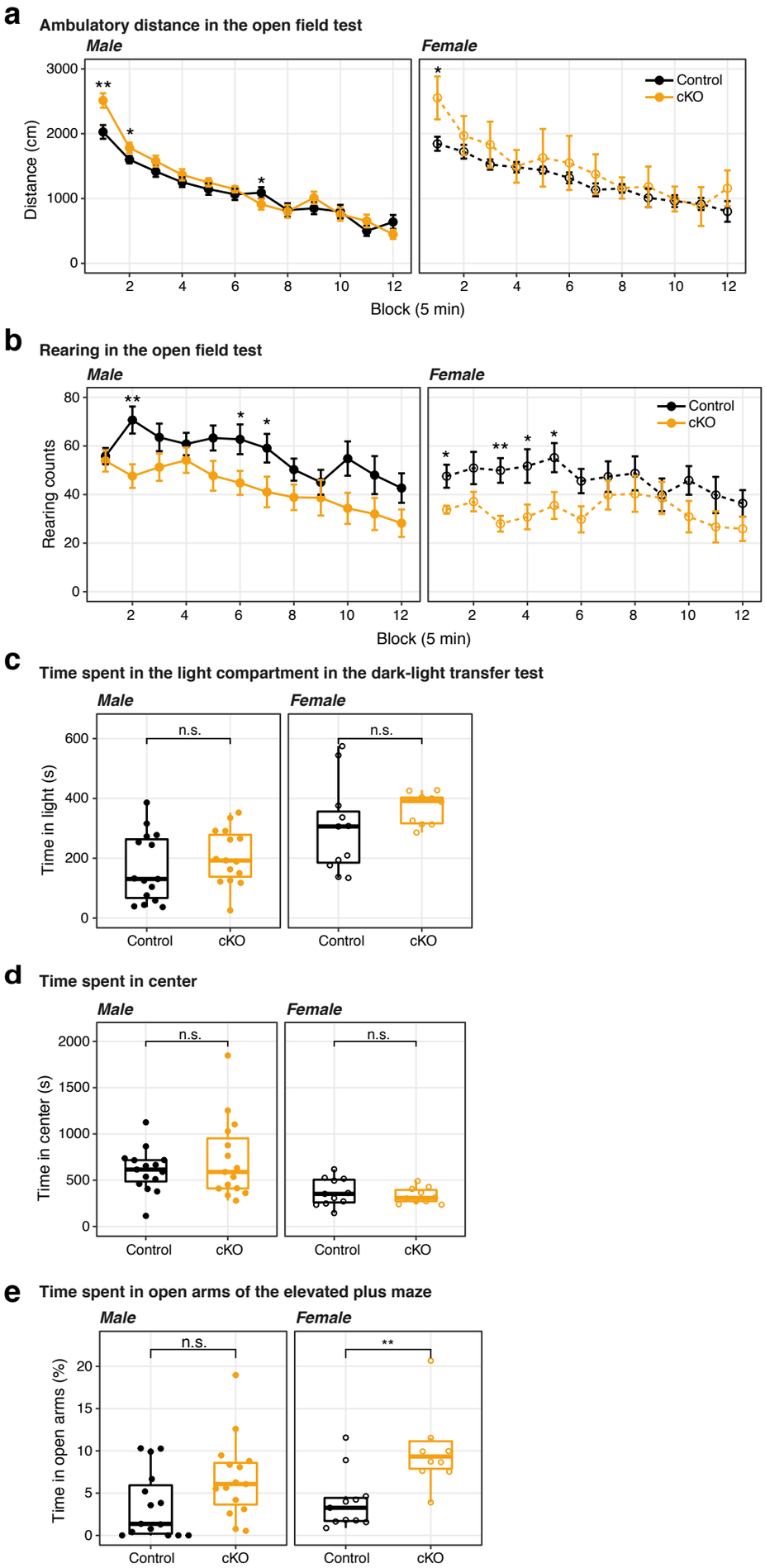


d

Nissl stained slices of brain in Control and *Dnmt3a* cKO animials



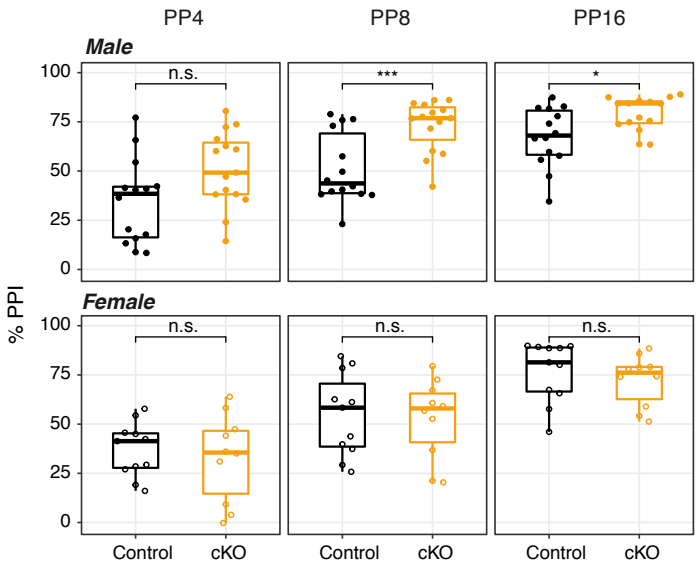
Li *et al.*, Supplementary Figure 2



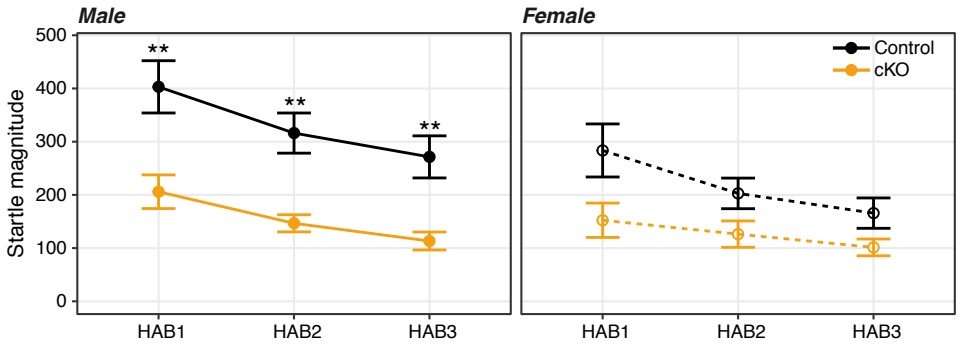
Li *et al.*, Supplementary Figure 3

Sensory gating

a **Pre-pulse inhibition (PPI) of the startle response**

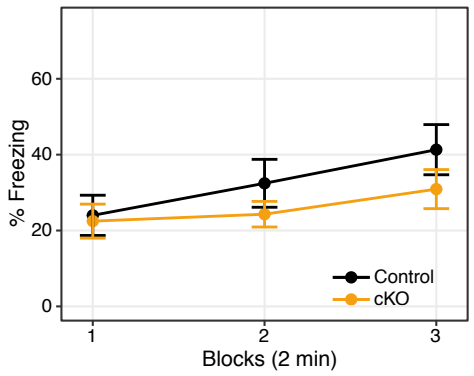


b **Habituation**

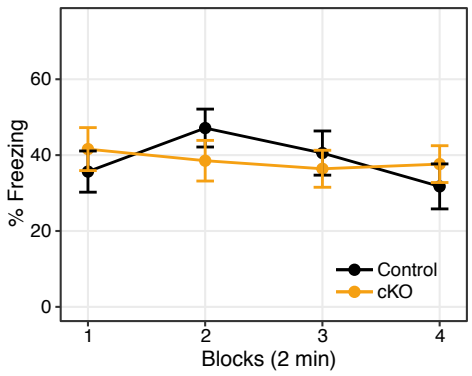


Contextual fear conditioning

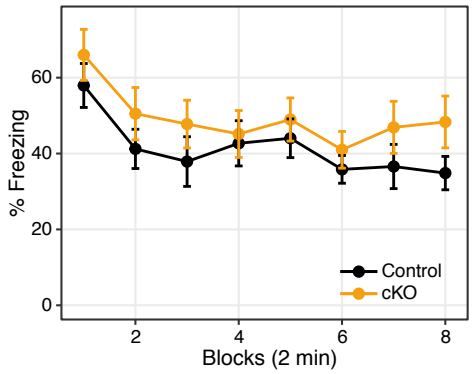
c **Acquisition**



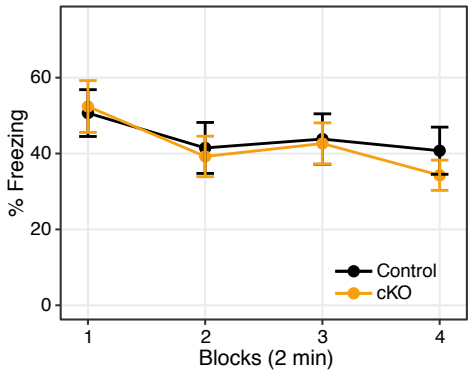
d **Context**



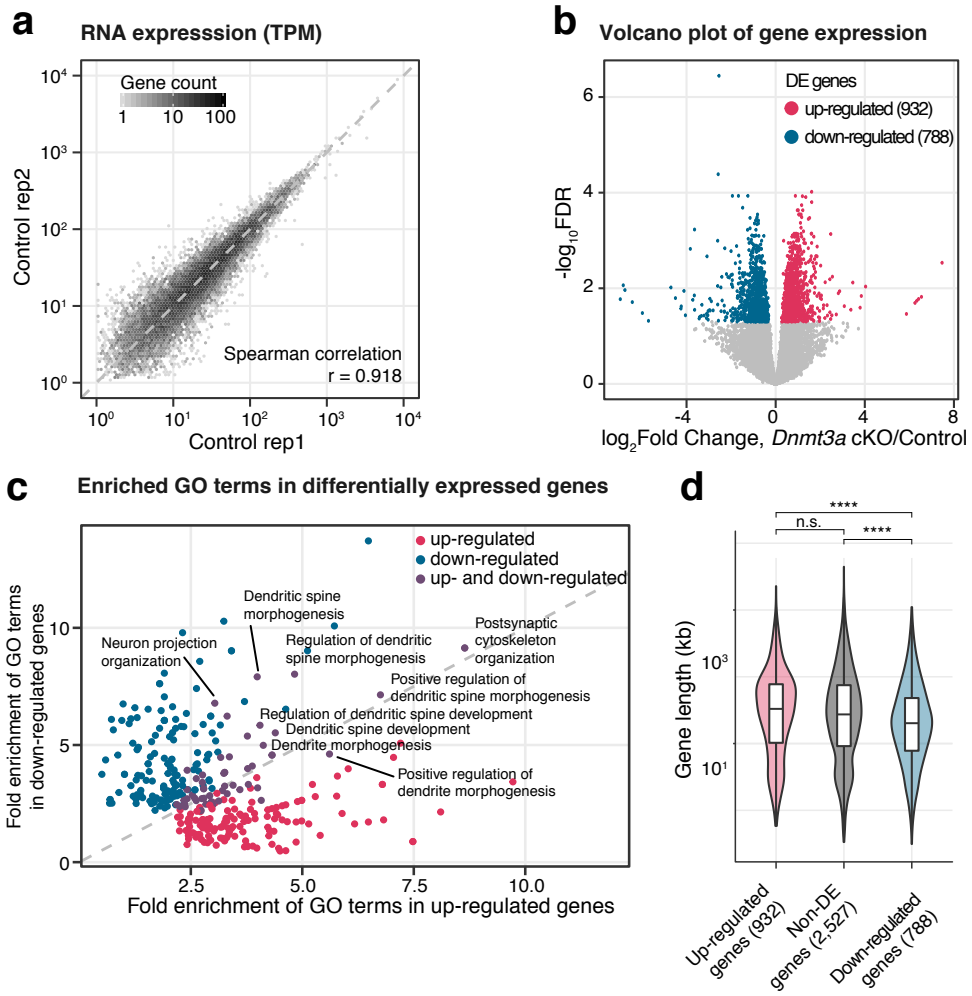
e **CS+**



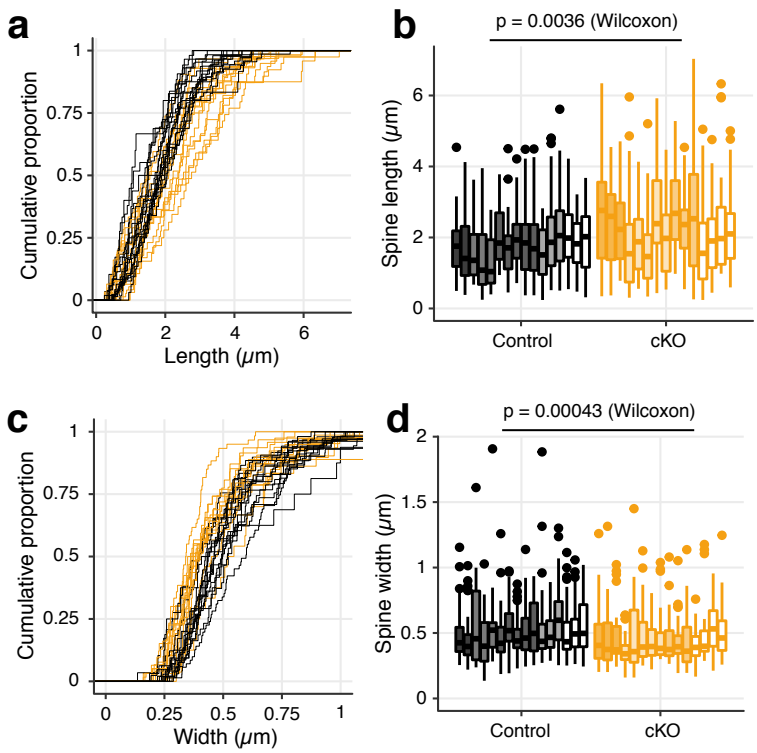
f **Extinction**



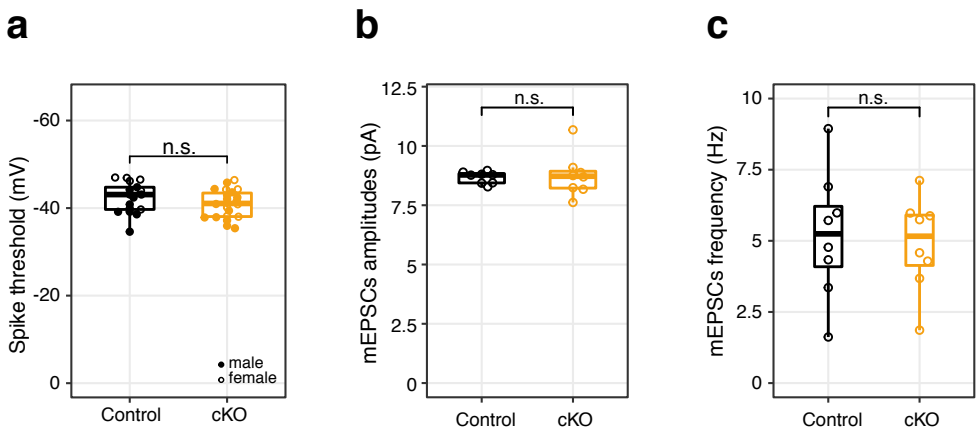
Li *et al.*, Supplementary Figure 4



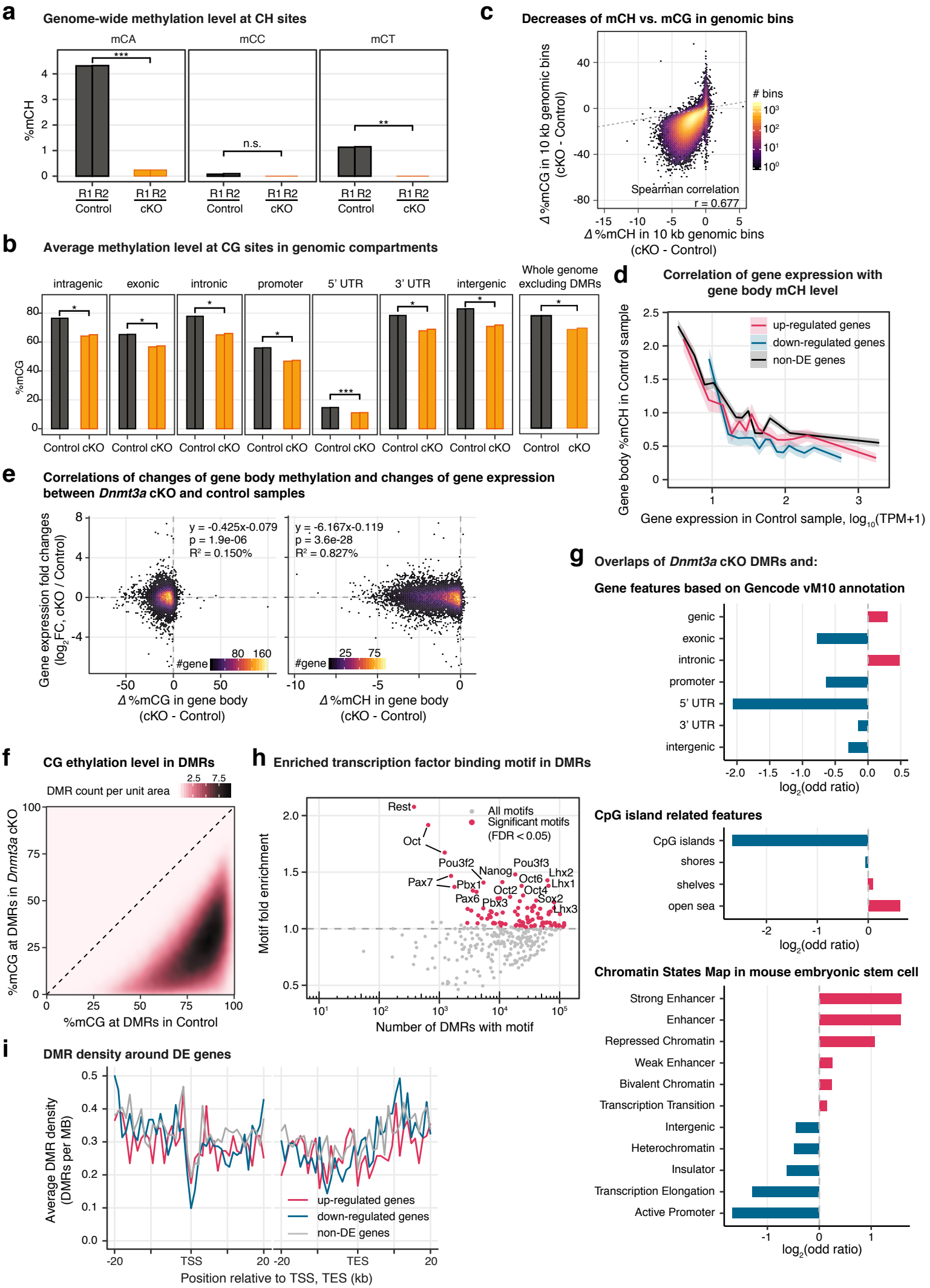
Li *et al.*, Supplementary Figure 5



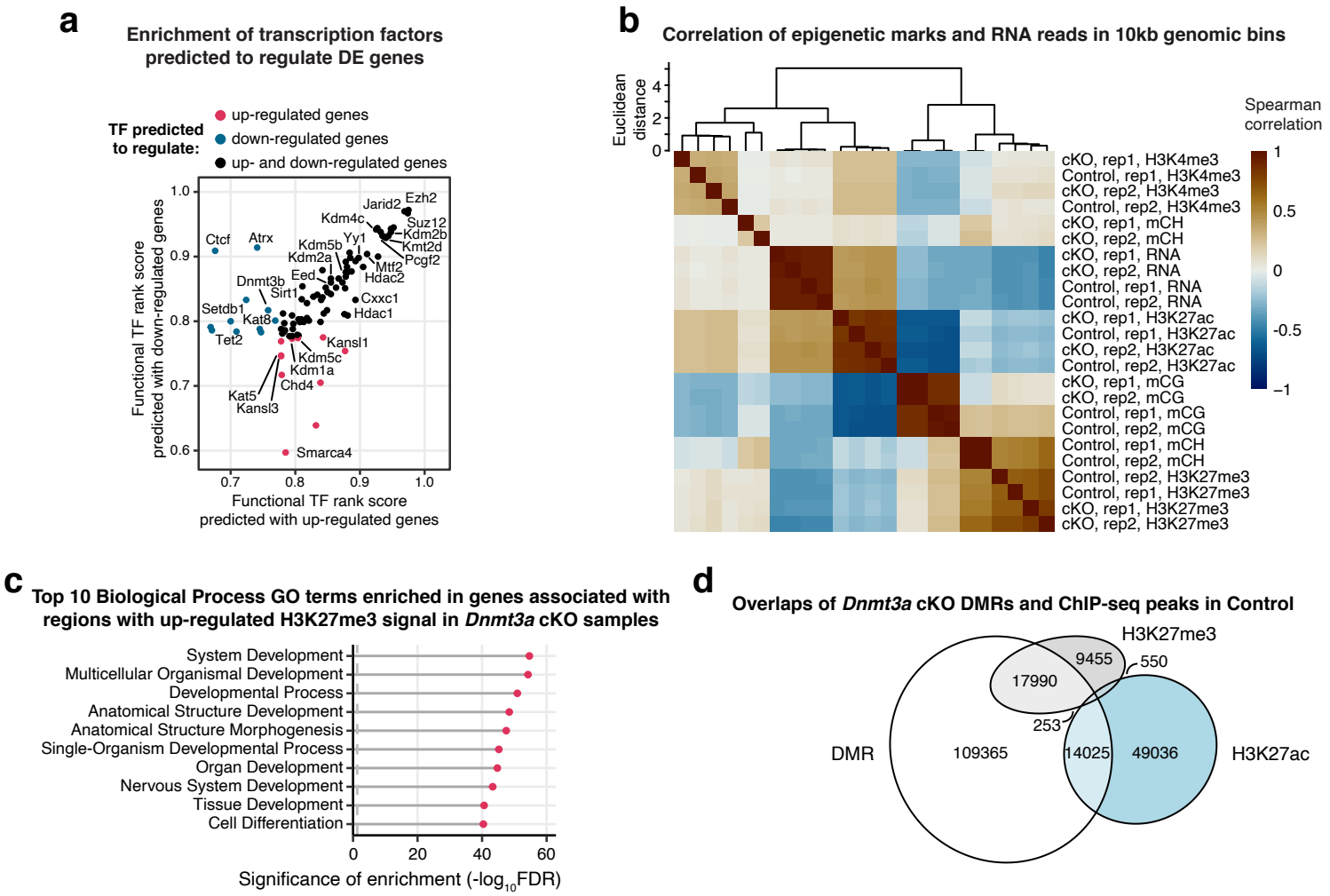
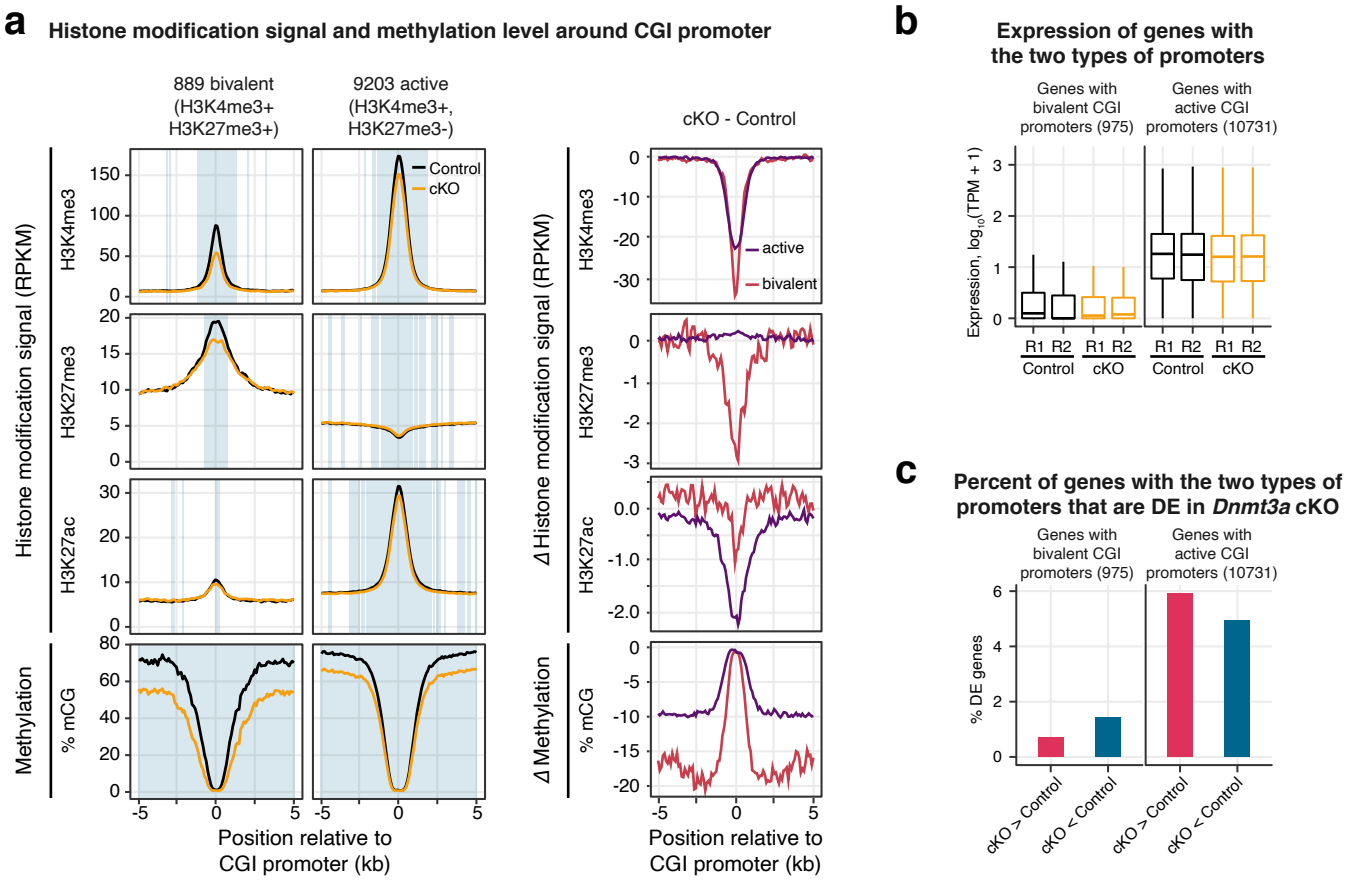
Li *et al.*, Supplementary Figure 6



Li *et al.*, Supplementary Figure 7



Li *et al.*, Supplementary Figure 8



Li *et al.*, Supplementary Figure 9