Recapitulation and reversal of schizophrenia-related phenotypes in *Setd1a*-deficient mice

Jun Mukai[1*], Enrico Cannavò[2*], Ziyi Sun[1,#], Gregg Crabtree[1], Anastasia Diamantopoulou[1], Pratibha Thakur, Chia-Yuan Chang[1], Yifei Cai[1], Stavros Lomvardas[2,3,4], Atsushi Takata[5], Bin Xu[6], Joseph A. Gogos[1,2,4*]

Affiliations:

1. Department of Physiology and Cellular Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
2. Mortimer B. Zuckerman Mind Brain and Behavior Institute Columbia University, New York, NY 10027 USA
3. Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY 10032 USA
4. Department of Neuroscience, Columbia University, New York, NY 10032 USA
5. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Fukuura 3-9, Kanazawa-ku, Yokohama 236-0004, Japan
6. Department of Psychiatry, Columbia University Medical Center, New York, NY 10032, USA

*Equal Contribution

[#]Present Address: State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, and Collaborative Innovation Center for Biotherapy, Chengdu, Sichuan 610041, China

Corresponding author

*Correspondence and requests for materials should be addressed to: Joseph A. Gogos

(jag90@columbia.edu)

Summary

***SETD1A*, a histone methyltransferase, is a key schizophrenia susceptibility gene. Mutant mice carrying a heterozygous loss-of-function mutation of the orthologous gene exhibit alterations in axonal branching and cortical synaptic dynamics, accompanied by specific deficits in working memory that recapitulates SCZ-related alterations. We show that Setd1a targets mostly enhancers and reveal a striking overlap between Setd1a and Mef2 chromatin targets. Setd1a targets are highly expressed in pyramidal neurons and enriched for genes with postnatally-biased expression involved in synaptic structure and function. Notably, evolutionary conserved Setd1a binding sites and target genes are strongly associated with neuropsychiatric genetic risk burden. Reinstating *Setd1a* expression in adulthood rescues working memory deficits. We identify LSD1 as a major demethylase counteracting the effects of Setd1a methyl transferase activity and show that LSD1 antagonism in adult *Setd1a*-deficient mice results in a full rescue of the behavioral abnormalities and axonal branching deficits. Our findings advance our understanding of how *SETD1A* mutations predispose to SCZ and point to therapeutic interventions.**

Keywords

SETD1A, MEF2, LSD1, histone methyltransferase, enhancer, loss-of-function (LoF) mutation, schizophrenia, axonal branching, working memory, pharmacological reversal

Introduction

Schizophrenia (SCZ) is a disabling neuropsychiatric disorder defined by positive and negative symptoms as well as cognitive impairment. Over the past decade, human genetic studies have revealed an important role of individually rare (but collectively common) *de novo* (DNMs) and inherited mutations in genetically complex neuropsychiatric diseases (Flint, 2016; Fromer et al., 2014; Genovese et al., 2016; Gulsuner et al., 2013; Rodriguez-Murillo et al., 2012). We have

previously linked loss-of-function (LoF) mutations in *SETD1A*, a histone methyltransferase, to risk for SCZ, a finding confirmed by large scale exome sequencing studies (Singh et al., 2016; Takata et al., 2016; Takata et al., 2014). SETD1A is part of the Set/COMPASS complex, which mediates mono-, di-, and tri-methylation of the lysine 4 on the histone H3 protein (H3K4(Miller et al., 2001). In humans, Set/COMPASS is comprised of one of six catalytic subunits [SETD1A, SETD1B, KMT2A, KMT2D, KMT2C and KMT2B], four common subunits (WDR5, RBBP5, ASH2L, and DPY30), and several subtype-specific proteins (Shilatifard, 2012). In addition to *SETD1A*, mutations in subunits of Set/COMPASS have been reported in SCZ and other neurodevelopmental disorders (NDD) (Gulsuner et al., 2013; Neale et al., 2012; O'Roak et al., 2012) as well as dominant Mendelian syndromes (Fahrner and Bjornsson, 2014).

SETD1A is a confirmed SCZ risk gene and *SETD1A* mutations confer a large increase in disease risk (Singh et al., 2016), which provides an excellent genetic substrate for disease modeling. However, the impact of SETD1A deficiency on brain structure and function and the mechanisms by which a ubiquitous molecular process, such as histone methylation, leads into specific psychiatric disease symptoms remains unknown. Here, we performed a comprehensive analysis of mutant mice carrying a LoF allele in the endogenous *Setd1a* orthologue that models the various identified *SETD1A* SCZ LoF risk alleles to uncover the role of *SETD1A* in gene regulation, neuronal architecture, synaptic plasticity and behavioral paradigms dependent on

learning. We also used this mouse model to examine whether malfunction of neural circuits and neurocognitive deficits emerging due to *Setd1a*-deficiency can be reversed by pharmacological interventions during adulthood, thus paving the way for the development of new therapeutics.

Results

The Allen Brain Atlas (www.brain-map.org) predicts that *Setd1a* is expressed throughout the adult mouse brain, with higher expression in neocortex. Real-time Quantitative Reverse Transcription PCR (qRT-PCR) analysis on mouse prefrontal cortex (PFC) detected Setd1a mRNA at various developmental stages (Figure S1A). Immunostaining of the prelimbic area in medial PFC of 6-weeks old mice revealed that Setd1a positive cells are distributed in all cortical layers except layer 1 (L1) and merged with NeuN positive ($NeuN^+$) cells (Figure S1B). Setd1a protein exhibits mixtures of punctate and diffuse fibrillary nuclear staining in all $NeuN^+$ neuron and parvalbumin positive (PV^+) interneurons in the medial PFC superficial layers, but only a few small puncta in CNPase (oligodendrocytes) and GFAP (astrocytes) positive cells in the cortical plate and corpus callosum (Figure S1C). Setd1a localizes to euchromatin and does not overlap with heterochromatin regions in either pyramidal or inhibitory neurons (Figure S1C).

We employed a mouse model carrying a Setd1a LoF allele (Figures S1D and S1E). Reduction of *Setd1a* RNA and protein levels to approximately half of WT levels was confirmed by

qRT-PCR and immunoblots of adult frontal cortex (Figures S1F and S1G). Although *Setd1a*$^{-/-}$ is embryonic lethal (Bledau et al., 2014), *Setd1a*$^{+/-}$ mice were indistinguishable from WT littermates in terms of body size, weight and posture (not shown).

***Setd1a*$^{+/-}$ mice exhibit deficits in working memory (WM) and cortical synaptic plasticity.** Cognitive impairment is a core feature of SCZ (Barch and Ceaser, 2012). Histone methylation mechanisms play a role in the coordination of complex cognitive processes and in modulating the genetic risk and the phenotypic variation of cognitive disorders (Rudenko and Tsai, 2014). To evaluate the relative effects of *Setd1a* deficiency on different cognitive domains and underlying brain circuits, we first asked whether *Setd1a* deficiency in mice affects performance in a variety of cognitive tasks. First, we established that *Setd1a*$^{+/-}$ mice show normal activity in the open-field assay (Figures S1H and S1I). Assessment of spatial WM performance in a T-maze delayed non-match to place task revealed that *Setd1a*$^{+/-}$ mice learned the task and performed almost as well as WT littermates during training (Figure 1A and 1B). In the WM test, however, *Setd1a*$^{+/-}$ mice manifested a marked impairment (correct responses: 63.63 ± 3.03 %, P = 0.0038 at 10 sec; 64.58 ± 3.10 %, P = 0.0172 at 30 sec), compared to WT littermate controls (76.39 ± 2.68 % at 10 sec, 76.39 ± 3.38 % at 30 sec) (Figure 1C). In contrast, memory performance of *Setd1a*$^{+/-}$ mice was intact in a series of other cognitive tasks, including direct interaction test for social memory (Figure 1D), novel object recognition test (Figure 1E) and fear conditioning task

for associative memory (Figure 1F). Thus, *Setd1a* deficiency leads to a robust and specific deficit in spatial WM, a core cognitive phenotype of psychotic disorders (Arguello and Gogos, 2012).

WM is sensitive to PFC dysfunction and can be affected by the synaptic properties of PFC neurons (Barak and Tsodyks, 2014). The glutamatergic inputs from L2 neurons to the apical and basal dendrites of L5 pyramidal neurons form monosynaptic synapses in medial PFC, which play a key role in WM and goal directed behaviors (Goldman-Rakic, 1995). We performed field recordings in brain slices on both genotypes to examine the synaptic properties in L5 neurons in response to the stimulation of L2. The input-output relation across experiments was comparable between genotypes, which indicates normal basic synaptic transmission. Similarly, the fiber volley amplitude was unaltered in *Setd1a*$^{+/-}$ mice indicating that when the same number of afferent fibers is activated, basal transmission is unaffected (Figures S1J, S1K, and S1L). We next determined whether short-term synaptic plasticity was impaired in *Setd1a*$^{+/-}$ mice. To this end, we employed a paired-pulse ratio (PPR) protocol at various inter-stimulus intervals (ISIs) as previously described for two other genetic models (Crabtree et al., 2017; Fenelon et al., 2013). At ISIs of less than 200 ms, both genotypes exhibited a significant facilitation, but *Setd1a*$^{+/-}$ mice showed overall significantly decreased PPR values (Figure 1G, WT, N = 7, n = 23; *Setd1A*$^{+/-}$, N = 7, n = 23; two-way repeated-measures ANOVA, P = 0.04. Bonferroni post-tests further revealed a

difference at ISI of 50 ms, $P < 0.05$) indicating an increased initial probability of presynaptic neurotransmitter release.

During PFC-dependent WM tasks, PFC neurons in local ensembles receive trains of inputs from neighboring cells in the 20 to 60 Hz frequency range (Miller et al., 1996). We tested the effect of *Setd1a* deficiency on synaptic plasticity within this physiological frequency range by applying 40 stimuli at 5 Hz, 20 Hz and 50 Hz. In response to these stimulus trains, a rapid depression of synaptic strength was observed in both genotypes, attributable to depletion of the readily releasable vesicle pool (Zucker and Regehr, 2002). Compared to WT mice, *Setd1a*$^{+/-}$ mice show significantly greater depression of fEPSP responses at 5 Hz (Figure 1H, WT, N = 7, n = 23; *Setd1a*$^{+/-}$, N = 7, n = 22; two-way repeated-measures ANOVA, $P = 0.03$) and at 20 Hz (Figure 1H, WT, N = 7, n = 20; *Setd1a* $^{+/-}$, N = 7, n = 19; two-way repeated-measures ANOVA, $P = 0.04$). No genotypic difference was observed at 50 Hz (Figure 1H, WT, N = 7, n = 18; *Setd1A*$^{+/-}$, N = 7, n = 16; two-way repeated-measures ANOVA, $P = 0.95$). To determine the impact of *Setd1a* deficiency on the potentiation of synaptic efficacy, short-term potentiation (STP) and long-term potentiation (LTP) were induced as described previously (Crabtree et al., 2017; Fenelon et al., 2013). No differences were found in either one of these measures (Figure S1M, $P = 0.57$) suggesting normal synaptic potentiation of L5 cortical synapses.

In addition to synaptic function, we tested whether PFC neurons displayed alterations in intrinsic excitability. Neuronal excitability is known to critically regulate network information flow (Titley et al., 2017) and has previously been shown to be altered in animal models of SCZ (Benamer et al., 2018; Crabtree et al., 2017; Gruter et al., 2015; Hamm et al., 2017). To test whether *Setd1a*-deficient neurons displayed altered intrinsic excitability, we performed whole-cell current-clamp recordings in PFC L2/3 pyramidal neurons – a major source of excitatory input to L5 – and found significantly increased action potential (AP) responses to current steps (Figure 1I). In contrast, we found the minimal current required for AP generation (rheobase) was unaltered (Figure S1N). Notably, neuronal input resistance, membrane time constant, and resting membrane potential were all found to be unaltered and thus did not underlie the observed enhanced neuronal excitability (Figures S1O, S1P and S1Q). Subsequent whole-cell voltage-clamp studies employing voltage steps revealed that *Setd1a*$^{+/-}$ mice displayed reduced outward currents near AP threshold voltages (-40mV to -30mV, Figures S1R). These reductions in hyperpolarizing currents may in part explain the enhanced AP generation we observed in *Setd1a*$^{+/-}$ mice.

Laminar organization and cytoarchitecture in the cortex of *Setd1a*$^{+/-}$ mice. Adult 8-week-old *Setd1a*$^{+/-}$ mice manifest normal gross brain morphology and only a modestly diminished density of pyramidal neurons in L2 and L5, as well as of PV^{+} inhibitory neurons in L5 in the

prelimbic area of medial PFC (Figures 2A, 2B, and 2C). Further analysis showed that alterations in the laminar organization are, at least in part, due to alterations in the proliferation of neural progenitor cells (NPCs) during corticogenesis, as previously suggested (Bledau et al., 2014) (Figures S2A, S2B, S2C, and S2D).

We also investigated whether *Setd1a* deficiency affects development of cortical axons and dendrites. Plasmids encoding EGFP were electroporated *in utero* into the dorsolateral ventricular zone of the mouse forebrain at E15.5, and embryos were allowed to be born and develop to P8.5 for analysis. All cells labeled with EGFP completed their migration into L2/3 by P8.5. As shown previously (Mukai et al., 2015), callosal axons crossed the midline, coursed beneath the gray matter of the contralateral somatosensory cortex, penetrated the gray matter, and ascended to L1–3 and 5 (Figures 2D and 2E). The total number of callosal axon-terminal branches within the contralateral cortical L1–4 was decreased by disruption of *Setd1a* (7.86 ± 0.46, N = 10, n = 30, $P < 0.0001$), compared to WT brains (12.20 ± 0.44, N = 10, n = 30) (Figures 2F and 2G). Notably secondary branch number was robustly decreased in *Setd1a*$^{+/-}$ brains (1.95 ± 0.44, N = 10, n = 30 versus WT brains 3.63 ± 0.48, N = 10, n = 30; $P < 0.007$) (Figures 2F and 2G). Similar deficits in axonal branching were observed in cortical primary cultures from *Setd1a*$^{+/-}$ and WT mice (Figures 2H, 2I, 2J, 2K, and 2L). We also looked for deficits in dendrite and spine development in the basal dendrites of L5 neurons in the prelimbic area of medial PFC by intercrossing *Setd1a*$^{+/-}$

mice with a reporter Thy1-GFP/M line, where L5 pyramidal neurons are labeled sparsely (Figure S2E). In contrast to the observed effect in axonal growth, *Setd1a*$^{+/-}$ mice did not show a simplification in the dendritic arbors by Sholl analysis (Figures S2F, S2G, and S2H). Morphotypic analysis of spines in 8-week-old mice showed that mushroom spine density was modestly reduced in the basal dendrites of L5 *Setd1a*$^{+/-}$ (100.0 ± 3.4 / 100 μm, N = 10, n = 20, $P < 0.002$) compared to WT neurons (118.3 ± 3.6 / 100 μm, N = 10, n = 20; Figures 2M, 2N, and S2I). Mushroom spines showed a small decrease in the head width (16%; N = 10, n = 605, $P < 0.0001$) and length (6.6%; N = 10, n = 605, $P < 0.001$) in *Setd1a*$^{+/-}$ compared to WT neurons (N = 10, n = 644) (Figures 2O and 2P).

ChIP-Seq of adult *Setd1a*$^{+/-}$ cortex shows peak enrichment within enhancers. Changes in gene expression in complex adult tissues, consisting of heterogeneous cellular populations, represent a composite state of primary effects on gene expression in combination with secondary downstream transcriptional events (i.e. due changes in local neuronal activity and connectivity patterns) as well as with alterations in the transcriptional landscape due to earlier developmental insults. Given that Setd1a is a histone modifying factor, we sought to identify direct targets of Setd1a in the PFC by chromatin immunoprecipitation followed by sequencing (ChIP–Seq) using an antibody that was previously utilized in Setd1a ChIP experiments (Li et al., 2016). We performed ChIP-Seq analysis in chromatin prepared from micro-dissected PFCs of 6-week

old mice for histone modifications that mark active promoters (H3K4me2, H3K4me3 and H3K27ac) and active or poised enhancers (H3K4me1 and H3K27ac) using the same chromatin preparations. Using stringent criteria [Irreproducibility Discovery Rate (IDR) < 0.05], we identified 1178 Setd1a peaks genome-wide. Analysis of the genome-wide binding profile suggests that, despite the fact that Setd1a mediates the trimethylation of H3K4 (Shilatifard, 2012), only a minority of Setd1a peaks are located at promoter elements (6.8%, 74/1178) as observed by the overlap of Setd1a with H3K4me3 peaks (Figures 3A, 3B, S3A, and Data Table 1). The majority of Setd1a peaks (94%, 1104/1178) overlap with enhancer marks and not H3K4me3 (Figures 3B, S3A, Data Table 1). Further analysis of the genomic distribution of the non-promoter Setd1a peaks, revealed that 57% (627/1104) reside within intergenic regions while 41% (452/1104) reside in intronic sites with enhancer-like chromatin properties (Figures 3B, S3A). A minority resides within exons (0.45%, 5/1104), 3'UTRs (0.63%, 7/1104) or within 1 kb from the transcription end-point (1.1%).

The *Setd1a*$^{+/-}$ phenotypic profile suggest a prominent impact of *Setd1a* deficiency on cortical structure and function. To assess if Setd1a-bound regulatory elements have pan-neuronal or cortical-specific functions we compared our H3K27ac ChIP-Seq from the PFC with ENCODE H3K27ac ChIP-Seq from the olfactory bulb (OB) and the cerebellum (Shen et al., 2012). Setd1a bound PFC promoters, as a whole, have similar H3K27ac levels in the OB and less so in the cerebellum (Figure 3C and S3B). In contrast the majority of the enhancer peaks (66%, 788/1178)

are enriched for H3K27ac specifically in the PFC (Figures 3D, 3E and S3B). These observations suggest that cortex-specific transcription factors recruit Setd1a at enhancers whereas more generally expressed transcription factors recruit Setd1a to promoters.

Setd1a targets are extensively enriched for Mef2 binding sites. *De novo* motif analysis of the Setd1a peaks using HOMER revealed Mef2 consensus motifs as the most enriched DNA motif within the 1104 enhancer peaks (Figure 3F, Data Table 1) and Sp1 consensus motifs as the most enriched motifs on the promoters (Data Table 1). ChIP-Seq experiments with an antibody that recognizes both Mef2a and Mef2c isoforms, which are expressed in the PFC, showed a striking overlap between Setd1a and Mef2 targets, with all but 27 of the 1178 Setd1a peaks being occupied by Mef2 (Figure 3G, S3C and S3D). Co-immunoprecipitation experiments with nuclear extracts prepared from the cortical plate, revealed a strong interaction between Setd1a and Mef2 (Figures 3H). Despite this interaction of Mef2 and Setd1a in solution, only a small fraction of the Mef2 peaks are strongly bound by Setd1a genome-wide (Figures S3C and S3D), suggesting that Mef2 is required but not sufficient for Setd1a recruitment and additional factors contribute to Setd1a recruitment to Mef2 binding sites. To determine if this regulatory relationship is conserved in the human genome we mapped Setd1a peaks from the mouse genome (mm10) to the corresponding positions in the human genome (hg38). Using two blast algorithms with different conservation thresholds (blastn and megablast tools, see methods), we mapped 691 and 304 out

of 1178 peaks to the human genome, respectively (Data Table 1). Notably, motif analysis on the conserved Setd1a peaks with HOMER revealed MEF2 motifs as being highly represented in the orthologous human sequences (megablast set, *P* = 1e-25; blastn set, *P* = 1e-32, Data Table 1), suggesting a conserved mechanism of Setd1a recruitment. In fact, the coordinated function between the two proteins appears highly conserved among supra-primates. PhastCons (see methods) in the Euarchontoglire clade revealed Setd1a targets with high sequence conservation throughout the peak (329/1178 peaks), and targets with partial conservation limited to only half of the peak (386/1178 peaks) (Figures S3E and S3F). On these partially conserved peaks, there is a significant enrichment for MEF2 motifs only on the conserved half of the peak (Figures S3E and S3F), further supporting a functional role of Mef2 in Setd1a recruitment.

To evaluate the functional conservation of the Setd1a/Mef2 peaks we explored the chromatin accessibility of the most conserved mouse Setd1a ChIP-Seq peaks in the human brain. To accomplish this we analyzed a previously published ATAC-Seq dataset from 14 regions of human postmortem brains (Fullard et al., 2018) focusing our attention specifically on neuronal and non-neuronal cells of 5 brain areas: dorsal and ventral PFC, hippocampus, medial dorsal thalamus and putamen. Of the 304 most conserved Setd1a sites, we found that 47 sites show a pattern of open chromatin in neurons across the 5 brain regions we analyzed, whereas the remaining 257 sites are occupied only in cortical and hippocampal neurons (Figure S3G).

Altogether, these data suggest that the role of Mef2 and Setd1a in transcriptional regulation is conserved between mouse and human and that the intergenic Setd1a peaks likely originate from neuronal cortical populations (see also below).

Setd1a targets are enriched for orthologues of disease-risk genes. To elucidate whether Setd1a-dependent genome regulation in mice is linked to neuropsychiatric etiology in humans, we determined more directly if Setd1a genomic binding sites are enriched for genes and genomic intervals implicated in SCZ by human GWAS studies. We first performed an interval-based analysis to assess enrichment of Setd1a peaks within the 108 loci identified by SCZ-GWAS (Schizophrenia Working Group of the Psychiatric Genomics, 2014). We focused only on conserved peaks (see above) and applied a permutation test (N=20,000 permutations) implemented in the RegioneR program to determine if there is any significant overlap between the conserved Setd1a peak intervals and the 108 SCZ-GWAS associated intervals (Schizophrenia Working Group of the Psychiatric Genomics, 2014). In both cases of conserved peaks defined either by blastn or megablast, we found a significant overlap (P = 0.001 for both blastn and megablast) (Figures 4A and Data Table 2). This was true for both intronic and intergenic Setd1a peaks (see Figure S4A-S4L). To confirm the specificity of this finding, we conducted the same analysis on data from other GWAS of CNS and non-CNS diseases, with

similar sizes of genome-wide significant intervals (Welter et al., 2014), but we did not observe any overlap (Tables S1 and S2).

We followed this finding by a complementary gene-based analysis, which directly assess whether the observed enrichment of Setd1a regulatory elements within SCZ-GWAS loci extends to Setd1a target gene content, as well. We restricted our analysis on targets genes with Setd1a peaks within their introns that represent well-annotated, high-confidence targets (N = 277 genes) as opposed to multiple putative target genes abutting each intergenic peak (Data Table 2). We assessed enrichment of these high-confidence Setd1a targets within the 108 SCZ-GWAS loci, using INRICH, and found a significantly overlap (permutation $P = 0.007$). Similar analysis of data from other GWAS of CNS and non-CNS diseases did not show significant enrichment (Tables S3 and S4). Notably, in keeping with the hypothesis that Mef2 and Setd1a have a coordinated role in SCZ-related gene regulation we also observe an excess within SCZ-GWAS intervals of target genes with intronic peaks co-bound by Setd1a and Mef2 but not Mef2-only peaks ($P = 0.007$ and 0.271 respectively) (Table S5).

We also investigated whether the high-confidence Setd1a target genes are significantly overrepresented among genes implicated in SCZ and other NDDs by studies of rare LoF DNMs. Using DNENRICH, we observed that there is significant enrichment of the high-confidence Setd1a target genes (Figure 4B, $P = 7.0 \times 10^{-6}$, FC = 1.71) among the genes hit by LoF DNMs in the

combined group of NDD (SCZ, ID/DD and ASD). This was especially true when restricting to the LoF-intolerant genes (pLI > 0.9 or LoF-FDR < 0.01) hit by LoF DNMs (Figure 4B, $P = 5.0 \times 10^{-6}$, FC = 1.96 with pLI, $P < 1.0 \times 10^{-6}$, FC = 2.32 with LoF-FDR). Significant enrichment was not observed among the genes hit by LoF DNMs, irrespective of their selective pressure, in unaffected subjects (Figure 4B, smallest $P = 0.25$) or congenital heart disease patients (Figure 4B, smallest $P = 0.30$). Stratification of NDD by disease, revealed an enrichment in ID/DD, ASD and SCZ although enrichment was significant only for ID/DD and ASD probably due to the number of input DNMs and the consequent statistical power (Figure 4B).

Setd1a targets synaptic genes with postnatally-biased expression. In the hope of finding clues to their mechanism of action, we further searched the functional and expression properties of Setd1a targets. GO analysis of high-confidence Setd1a target genes revealed terms relevant to synaptic structure and function as well as significant enrichment for the “schizophrenia” term as defined in the DisGeNET database (umls:C0036341) (Figure 4C). We utilized the CSEA tool to map the orthologues of mouse target genes to their potential brain regions and developmental stages (see methods). We saw a significant enrichment (pSI threshold $P < 0.01$) in the cortex at five developmental stages: early mid fetal (FDR corrected pSI $P = 0.002$), early infancy (FDR corrected pSI $P = 2.130e-06$), childhood (FDR corrected pSI $P = 2.762e-06$), adolescence (FDR corrected pSI $P = 3.360e-14$), and young adulthood (FDR corrected pSI $P =$

3.515e-14) the evidence being the strongest for adolescence and young adulthood stages (Figures 4D). Notably, although mutation-intolerant genes hit by LoF DNMs in NDDs manifest an overall early fetal-biased expression (Figure S4M), we found that the subset of mutation-intolerant Setd1a target genes hit by LoF DNMs in NDDs show strong enrichment in the cortex during adolescence (FDR corrected pSI P = 0.001) to young adulthood (FDR corrected pSI P = 0.002) (Figures 4E). Providing further clues to their mechanism of action, the observed pattern of expression bias of Setd1a targets is in agreement with results from a study using CSEA on genes implicated in SCZ by GWAS, which also showed strong enrichment in the cortex from childhood to young adulthood of the most likely candidate genes (Figure S4N) (Shohat et al., 2017).

To obtain insights into the molecular alterations in the PFC of *Setd1a*$^{+/-}$ mice, we used RNA-Seq analysis to profile transcriptomes of micro-dissected PFCs from four 6-week-old *Setd1a*$^{+/-}$ mice and four WT littermates. We identified 342 genes that significantly change their expression in *Setd1a*$^{+/-}$ mice (FDR<5%) (Figure 5A, Data Table 1). Transcriptional alterations were modest with most of the genes (> 70%) displaying a fold change in the range of 0.7-1.3. Surprisingly, with few exceptions (28 genes, 16 up-regulated and 12 down-regulated), most of the genes proximal to Setd1a peaks do not show significant change in *Setd1a*$^{+/-}$ mice (Figure 5A). While bulk PFC RNA-seq analysis failed to detect robust statistically significant transcriptional changes of most genes proximal to Setd1a peaks, enrichment analysis indicated that reduction

of *Sedt1a* levels by half results for the most part in a collective upregulation of high-confidence Setd1a target genes (Figure 5A). This pattern of transcriptional dysregulation was corroborated by qRT-PCR analysis of selected targets (Figure S5A). One notable exception is the direct Setd1a target *Tcf4*, a gene implicated in SCZ and other NDD via rare and common variant studies (Deciphering Developmental Disorders, 2017; Stefansson et al., 2009), which was found to be downregulated (also corroborated using qRT-PCR, Figure S5A). Thus, reduction of Setd1a levels by half leads to a complex pattern of low magnitude changes in the expression of direct targets.

To examine the possibility that this pattern is caused by cellular heterogeneity of the PFC, we performed single cell RNA-Seq (sc-RNA-Seq) on dissociated cells from the PFC of WT and *Setd1a*$^{+/-}$ mice (see Methods). We obtained expression profiles for 5,944 cells from WT mice and 6,031 cells from *Setd1a*$^{+/-}$ mice (with a mean of 2262 genes per cells). We identified 8 main clusters that represent the major neuronal [pyramidal neurons (*Camk2a*) and interneurons (*Igfbpl1*)] and non-neuronal [mature and immature oligodendrocytes (*Olig2* and *Mbp*), microglia (*Aif1*), astrocytes (*Aldoc*) and cells of the blood brain barrier (*Cldn5*)] populations of PFC (Figures 5B, S5B, S5C, S5D and S5E). As an indicator for the quality of our scRNA-Seq data, we compared the level of expression, as a whole, of genes that we detected as significantly changing in bulk PFC RNA-Seq and confirmed the same trend of expression changes between bulk and scRNA-Seq (Figures S5F and S5G).

Analysis of Setd1a-proximal genes, revealed the highest expression levels in pyramidal neurons (Figure 5C). This finding is in line with follow-up ATAC-seq results from pyramidal neurons and interneurons (Figures 5D and 5E). Specifically, FAC-sorting of pyramidal and interneuronal populations from double transgenic progeny of Emx1-ires-Cre or Parvalbumin-ires-Cre lines crossed to tdTomato reporter mice, followed by ATAC-Seq, revealed that Sedt1a-bound intergenic peaks are accessible and occupied by transcription factors specifically at pyramidal neurons (Figures 5D). As an example, the *Lzts1* gene locus is shown in more details in Figure 5E. We identified two enhancer elements (regions marked by H3K27ac but not H3K4me3) that are bound by both Setd1a and Mef2c specifically in pyramidal neurons and not in interneurons, as shown by cell-type specific ATAC-Seq. Moreover, combined whole-PFC RNA-Seq and scRNA-Seq indicated that this gene is expressed in the PFC and in particular in the pyramidal neurons.

Similar to bulk tissue analysis, differential expression analysis failed to detect robust genotypic changes in the transcription of most genes proximal to Setd1a peaks. However, in keeping with our bulk tissue analysis, a modest but significant collective upregulation of high-confidence Setd1 target was observed in pyramidal neurons (*p-val*= 2.2^{-16}). Thus, heterozygous Setd1a mutations lead to modest, cell-type specific transcriptional dysregulation of Setd1a targets.

Reinstating *Setd1a* expression in adulthood rescues WM deficits. The observation that Setd1a targets genes with postnatally-biased expression and the lack of evidence of

irreversible neurodegenerative changes in *Setd1a*-deficient brains prompted us to investigate the possibility of reversing cognitive deficits in adulthood. To this end, we took advantage of the conditional nature of the targeted allele and adopted a genetic method that allows for inducible reactivation of Setd1a in *Setd1a*$^{+/-}$ adult mice (see Methods)

To achieve temporal control, we crossed a *Setd1a*$^{+/-}$ mice to an inducible *R26*^{FlpoER} mouse line (*R26FlpoER*) that activates global Flpo function upon tamoxifen treatment (Figures 6A), by using the *Flp/Frt*-dependent genetic switch strategy, which enables the conditional manipulation of the Setd1a gene at its endogenous genomic locus, thus preserving Setd1a expression within physiological levels. When the *Setd1a*$^{+/-}$*: R26FlpoER*$^{+/-}$ mice reached 8 weeks, we used oral gavage to deliver tamoxifen and implement Flpo-mediated gene restoration (Figures 6A and 6B).

Behavioral analysis was performed on the following 4 groups: *Setd1a*$^{+/+}$: *R26FlpoER*$^{+/-}$ mice treated with tamoxifen (WT + TAM), *Setd1a*$^{+/-}$:*R26FlpoER*$^{+/-}$ mice treated with tamoxifen (*Setd1a*$^{+/-}$ + TAM), *Setd1a*$^{+/+}$:*R26FlpoER*$^{+/-}$ mice treated with corn oil vehicle (WT + oil), and *Setd1a*$^{+/-}$:*R26FlpoER*$^{+/-}$ mice treated with corn oil vehicle (*Setd1a*$^{+/-}$+ oil). Investigation of the efficiency of Setd1a reactivation upon Flpo-mediated deletion of the *LacZ* cassette confirmed that upon tamoxifen treatment Setd1a mRNA and protein levels were reinstated in *Setd1a*$^{+/-}$ + TAM mice to ~75% of WT levels in the adult mouse brain (Figures 6C, 6D, S6A, S6B, and S6C). The effect of elevation of Setd1a levels on spatial WM performance was assessed in a T-maze

delayed non-match to place task. As expected, *Setd1a*$^{+/-}$+ oil mice showed impaired WM performance compared to WT + oil control littermates, which was partially rescued upon tamoxifen treatment (Figures 6E, S6D, and S6E). Rescue was particularly evident at 10s in *Setd1a*$^{+/-}$ + TAM mice (82.64 ± 2.29%, N = 24) compared to WT + TAM littermates (85.42 ± 1.83%, N = 24), or *Setd1a*$^{+/-}$+ oil (68.25 ± 3.38%, N = 21) (P = 0.35 versus WT + TAM; P = 0.0012 versus *Setd1a*$^{+/-}$ + oil), but only modest (and statistically non-significant) at longer delays.

Partial restoration of neurocognitive deficits might have resulted from the incomplete reinstatement of Setd1a expression (~75% of WT levels) or could reflect a partial developmental origin of the behavioral abnormalities requiring additional earlier interventions. These limitations notwithstanding our findings indicate that, consistent with continued plasticity in the adult brain (Mei et al., 2016), adult *Setd1a* function reinstatement has substantial effect on ameliorating neurocognitive deficits and suggest a broad window of therapeutic opportunity (Neale et al., 2012). We explored this possibility further by taking advantage of the more homogeneous biodistribution of systemically administered pharmacological compounds that overcome the limited efficiency of genetic manipulations.

Pharmacological Inhibition of LSD1 activity in adulthood counteracts the effects of *Setd1a* deficiency. Within the past few years, a number of histone demethylases have been discovered, as well as an ever increasing number of potent and selective histone demethylases

inhibitors that exhibit good brain penetration. We thus surmised that these compounds would most likely generate robust methylation changes, which could, in turn, counteract effects of *Setd1a* deficiency leading to restoration of Setd1a-linked deficits. The Setd1a-counteracting demethylase(s) are unknown. We hypothesized that LSD1 may be one such demethylase since both our whole tissue RNA-Seq and scRNA-Seq analysis identified LSD1 as the most abundant demethylase in the frontal cortex. To test whether LSD1 binds *in vivo* the same regulatory elements bound by Setd1a we performed LSD1 ChIP-Seq from the prefrontal cortex of 6 weeks old mice, identifying 43,917 LSD1 bound regions (IDR<0.05). Strikingly, as for Mef2, we observed a compelling overlap between LSD1 and Setd1a bound regions (1,137/1,178), both at promoter and enhancer elements (Figures 7A, 7B, and 7C, the *Neurod6* locus is shown as an example in Figure 7A).

To confirm the ChIP-Seq findings we undertook parallel phenotypic screens on primary cultures of *Setd1a*$^{+/-}$ cortical neurons using well-characterized demethylase inhibitors and determining their potential to rescue the robust axonal arborization deficit (Figures 7D and 7E). *Setd1a*$^{+/-}$ neurons were treated with inhibitors of LSD1/KDM1a [RN-1, tranylcypromine (TCP), ORY-1001, and SP2509], KDM5B (PBIT) and the pan-KDM inhibitor IOX-1 from DIV3 to DIV5, transfected with plasmids expressing EGFP at DIV4 and fixed at DIV5. TCP, ORY-1001, SP2509, and RN-1 (all LSD1 inhibitors), but not PBIT or IOX-1 rescued the axon branching deficits in

Setd1a$^{+/-}$ neurons, confirming an important role of LSD1 antagonism in counteracting the effects of *Setd1a* deficiency (Figures 7D and 7E).

We followed up the LSD1 findings by examining the effect of inhibition of LSD1 on cognitive deficits following systemic TCP administration (Figures 7H, S6F, S6G, S6H, S6I, and S6J). *Setd1a*$^{+/-}$ mice and WT littermates were injected with 3 mg/kg/day TCP or vehicle for 14 consecutive days followed by the non-match to sample T-maze task. During the task, mice were administered daily with the same dose of TCP or vehicle. Assessment of spatial WM performance revealed that, similar to our previous findings, there are only modest genotypic effects during training (Figures S6F and S6G), but a robust impairment in the performance of vehicle-treated mutant mice over vehicle treated WT control littermates during testing with increasing time delays (Vehicle-treated *Setd1a*$^{+/-}$ mice correct responses: 67.08 ± 3.23%, $P = 0.0016$ at 10 sec; 63.33 ± 2.37%, $P < 0.0001$ at 20 sec; 62.92 ± 2.93%, $P = 0.045$ at 40 sec, N = 20 compared to vehicle-treated WT control correct responses: 80.83 ± 2.23% at 10 sec; 78.75 ± 2.60% at 20 sec; 70.83 ± 2.46% at 40 sec, N = 20) (Figure 7H). This deficit was completely reversed by TCP treatment, such that TCP-treated *Setd1a*$^{+/-}$ mice were indistinguishable from TCP-treated WT littermates (TCP-treated *Setd1a*$^{+/-}$ mice correct responses: 81.25 ± 2.33%, $P = 0.900$ at 10 sec; 76.67 ± 3.06%, $P = 0.924$ at 20 sec; 72.5 ± 2.35 %, $P = 0.705$ at 40 sec, N = 20, compared to TCP-treated WT control correct responses: 81.67 ± 2.31% at 10 sec; 76.25 ± 3.04% at 20 sec; 72.75 ± 2.28%,

at 40 sec, N = 20) (Fig. 7H). TCP had no effect on WT mice. While vehicle treated mutant mice showed the expected deficits in days to criterion and correct percentages during training phase TCP treatment did not improve these deficits. Follow-up dose-response studies indicated that the effective TCP dose range is rather narrow: a lower dose (1 mg/kg/day, i.p.) was ineffective in improving the WM performance (Figures S6H,S6I and S6J) whereas a higher dose of TCP (10 mg/kg/day, i.p.) resulted in increased lethality precluding further cognitive assessment.

Pharmacological inhibition of LSD1 *in vitro* can repair axonal terminal branching but it is unclear whether *in vivo* post-developmental LSD1 antagonism by TCP has similar beneficial effects on the attenuated terminal arbor complexity, parallel to the rescue of the cognitive deficits. To address this question, we injected adeno-associated virus rAAV2/hSynapsin/EGFP into the frontal cortex of 12-14 week-old mice following the T-maze test to label cortical neurons with EGFP in TCP- or vehicle-treated mutant and WT mice. Mice were sacrificed after 2 weeks incubation. We analyzed EGFP intensity in contralateral axon target regions as a quantitative index of the complexity of axon branching. As expected mutant mice treated with vehicle showed significantly reduced callosal axon-terminal branching within the contralateral cortical L2/3 (Figures 7F and 7G). By contrast TCP treatment restored axon branching in contralateral L2/3 of *Setd1a*$^{+/-}$ to near WT level (Figures 7F and 7G).

While TCP is a prototype irreversible LSD1 inhibitor, it also inhibits two major isoforms of monoamine oxidases MAO-A and MAO-B. To control for effects of MAO inhibition on WM we also evaluated the effect of ORY-1001, a TCP-derivative that represents the most potent and selective LSD1 inhibitor described to date (Hosseini and Minucci, 2017; Maes et al., 2015; Maes et al., 2018). The compound has an IC50 of 18 nM and an over 1000- fold selectivity for LSD1 compared to MAOs (Maes et al., 2015; Maes et al., 2018). Indeed, in our in vitro axon branching rescue assays ORY-1001 was 1000-fold more potent than TCP and demonstrated sub-nanomolar cellular activity (Figure 7E). To examine whether ORY-1001 is effective in reversing WM deficits in *Setd1a*$^{+/-}$ mice, mutant mice and WT littermates were administered with ORY-1001, or vehicle (PBS) for 14 consecutive days prior to the T-maze task and daily throughout the duration of the task. As expected, we observed a robust impairment in the performance of vehicle-treated mutant mice over vehicle treated WT control littermates during testing with increasing time delays (Vehicle-treated *Setd1a*$^{+/-}$ mice correct responses: 69.84 ± 2.96%, P =0.00049 at 10 sec; 64.68 ± 2.63% P = 0.0048 at 20 sec; 59.92 ± 2.48% P = 0.00013 at 40 sec, N = 21 compared to vehicle-treated WT control correct responses: 82.97 ± 1.61% at 10 sec; 75.72 ± 2.62% at 20 sec; 74.28 ± 2.34% at 40 sec, N = 23). This WM deficit was reversed by 10 μg/kg/day ORY-1001 treatment of *Setd1a*$^{+/-}$ mice (ORY-1001-treated *Setd1a*$^{+/-}$ mice correct responses: 84.44 ± 2.42%, P = 0.767 at 10 sec; 81.67 ± 3.27%, P = 0.272 at 20 sec; 80.56 ± 3.01%, P = 0.628 at 40 sec, N = 15,

compared to TCP-treated WT control correct responses: 83.33 ± 2.82% at 10 sec; 85.56 ± 2.22 % at 20 sec; 82.78 ± 3.40%, at 40 sec, N =15) (Figure 7l).

Days to criterion and percentage of correct choices during training phase were also improved in both ORY-1001-treated *Setd1a*$^{+/-}$ and WT mice (Figures S6K and S6L). Notably, administration of lower of ORY-1001 (3 μg/kg/day) proved ineffective in improving the WM performance during testing with increasing time delays (Figures S6M, S6N, and S6O).

Discussion

The discovery of rare but highly penetrant genetic risk variants for SCZ and the generation of mouse models that faithfully recapitulate their effects has provided unprecedented insights into disease pathophysiology (Arguello and Gogos, 2012; Del Pino et al., 2018). A major remaining challenge is how to translate these mechanistic insights into effective therapies. Here we describe a mouse model of *SETD1A* LoF SCZ-risk mutations with remarkable face validity crossing multiple levels of analysis and offering a reliable entry point to translate histone methylation, a ubiquitous process, into psychiatric disease-specific mechanistic and therapeutic insights.

We describe a profile of robust but circumscribed cognitive impairments in spatial WM as well as cortical synaptic alterations due to *Setd1a*-happloinsufficiency, including sparser structural connectivity due to reduction in terminal arborization of cortical axons as well as alterations in

short-term synaptic plasticity and neuronal excitability, which collectively may lead to disruptions of network activities that support WM and related cognitive operations (Mi et al., 2017). Overall, *Setd1a*$^{+/-}$ phenotypes relates well to established evidence that SCZ patients have reduced WM capacity (Arguello and Gogos, 2012) as well as to cognitive and synaptic alterations described in two other models of highly penetrant SCZ-risk mutations (Crabtree et al., 2017; Koike et al., 2006). Overall, comparison among these strains, together with previous *in vivo* analysis of neuronal activity of locally (Hamm et al., 2017) or distally (Tamura et al., 2016) connected neocortical populations, suggests a pattern of dysconnectivity within and between brain regions, shaped by alterations in rapid synaptic dynamics, altered intrinsic excitability and sparser connections, which may affect the coherence of neuronal activity and may underlie the emergence of cognitive deficits (Arguello and Gogos, 2012).

Our ChIP-Seq analysis, revealed a large number of putative Setd1a direct chromatin targets, overwhelmingly within enhancer rather than promoter elements. Most of them appear to have a cortex- and neuronal-specific function both in mouse and human, suggesting that the molecular mechanisms that lead to the recruitment of Setd1a to regulatory elements is conserved between species. Interestingly, approximately one-third of them reside within introns, consistent with accumulating evidence that more than 50% of variants associated with human phenotypes in GWAS are within intronic space (Welter et al., 2014). Our ChIP-Seq analysis also reveals a

key role for Mef2 transcription factors in mediating the effect of Setd1a chromatin binding. Notably, Mef2 isoforms have been shown to modulate cortical structural and functional plasticity and have been previously implicated in SCZ, ASD and other neuropsychiatric disorders (Harrington et al., 2016; Mitchell et al., 2017). Thus, Setd1a and Mef2 may be working in concert to regulate expression of disease-relevant genes.

Setd1a targets are enriched for genes with postnatally-biased expression involved in synaptic structure and function. Moreover, there is a highly specific overlap between Setd1a targets in mice and lists of risk genes suggesting that *Setd1a* haploinsufficiency in both mice and humans may lead to behavioral and physiological deficits in part via coordinated mis-regulation of neuropsychiatric-risk genes. Indeed, interpreted in the context of our RNA-Seq analysis, reduction of Setd1a levels by half may lead to low magnitude changes in the expression of many downstream targets with likely cumulative effects on cellular structure and function, especially in pyramidal neurons where Setd1a targets show the highest levels of expression. This finding, taken together with our results from ChIP-Seq analysis indicating a cortex-specific signature of Setd1a chromatin binding in the mouse brain as well as with our results from analysis of ATAC-seq data indicating that Sedt1a intergenic peaks are accessible specifically at pyramidal neurons, strongly suggests the presence of auxiliary factors that may mediate the brain region-specific effects of this relatively ubiquitous chromatin regulator. Given the previously described role of the

COMPASS complex as transcriptional activator, one unexpected finding from our RNA-seq analysis, is that reduction of Sedt1a levels by half results to a large extent (albeit not exclusively) in upregulation of direct target genes. While this finding requires further investigation, it is possible that Setd1a may have other functions as a chromatin regulator, in addition to catalyzing methylation of H3K4 (Hoshii et al., 2018), which may be locus-dependent and differentially sensitive to Setd1a-dosage. For example, recent data suggest that Mef2 can be regulated by G9a-dependent lysine methylation, which inhibits its activity (Choi et al., 2014; Ow et al., 2016). In that vein, it is possible that Setd1a also methylates Mef2, acting as a specific inhibitor for some Mef2-bound enhancers.

Notably, while some of the alterations in *Setd1a*$^{+/-}$ mice may be caused by impaired prenatal development, one of the novel aspects of the current study was the demonstration that specific cognitive and neural circuitry pattern impairments in the Setd1a-deficient brain can be partially or fully corrected by delayed restoration of Setd1a function. Our results revealed a beneficial effect of post-developmental activation of *Setd1a* expression on cognitive dysfunction, highlighting a unique role of Setd1a expression after development and arguing that the deficiency of Setd1a function during early developmental stages does not irreversibly compromise the potential for neural circuits to achieve beneficial plasticity responses. Similar analysis in other models of monogenic brain disorders has provided mixed results. For example,

adult reactivation of *Ube3a* or the *Syngap1* gene in mouse models of autism and intellectual disability did not reverse any of the core behavioral deficits (Barnes et al., 2015; Rotaru et al., 2018), while adult reactivation of *Mecp2* could rescue behavioral alterations in a mouse model of Rett syndrome (Tillotson et al., 2017).

Guided by these findings, we employed a series of effective pharmacological interventions to directly target, with well-characterized and potent inhibitors, downstream demethylation processes. Our goal was to generate robust methylation changes which, in turn, could counteract effects of *Setd1a* deficiency leading to restoration of Setd1a-linked deficits of cell function and perhaps behavior. Combined results from ChIP-Seq analysis and *in vitro* phenotypic screens identified LSD1 as a major demethylase counteracting the effects of Setd1a methyl transferase activity. Further *in vivo* analysis showed that LSD1 antagonism in *Setd1a*-deficient mice results in a full rescue of the behavioral abnormalities and axonal arborization deficits. While it will be important to determine the specific methylation sites underlying the beneficial effects of LSD1 antagonism, our findings indicate that reactivating Setd1a function or counteracting downstream effects of *Setd1a* deficiency in the adult brain may be sufficient to reverse the manifestation of neurocognitive phenotypes highlighting the continued structural and functional plasticity in the adult brain. Thus, in addition to mechanistic insights, our findings may prove useful for development of new therapeutics. Therapies targeted to SETD1A may have wider implications

as suggested by our findings here as well by the fact that pathways related to histone methylation are strongly enriched among biological pathways in GWAS data of SCZ (Network and Pathway Analysis Subgroup of Psychiatric Genomics, 2015).

Methods

Animals. *Setd1a*$^{tm1a(EUCOMM)Wtsi}$ mouse (referred to as *Setd1a*$^{+/-}$) was obtained from EMMA (https://www.infrafrontier.eu/search). *Setd1a*$^{+/-}$ was backcrossed for 10-11 generations in the C57BL/6J (The Jackson Laboratory, Bar Harbor, ME) background. B6N.129S6(Cg)-*Gt(ROSA)26Sor*$^{tm3(CAG\text{-}flpo/ERT2)Alj}$/J (referred to as *R26*^{FlpoER}) was purchased from Jackson laboratory. To achieve temporal control of *Setd1a* expression for the *Setd1a* rescue experiments, we crossed the *Setd1a*$^{+/}$ mice to an inducible *R26*$^{FlpoER+/+}$ (Lao et al., 2012) mouse line that activates global *Flp* function after tamoxifen treatment. B6.Cg-*Gt(ROSA)*$^{26Sortm1.3(CAG\text{-}tdTomato,\text{-}EGFP)Pjen}$/J (referred to as *RC::FLTG*) (Plummer et al., 2015) was purchased from Jackson Labs. To evaluate the tamoxifen-inducible Flp recombination in neurons, we crossed the *RC::FLTG* mice to an inducible *R26*$^{FlpoER+/+}$ (Lao et al., 2012) mouse line that activates global *Flp* function after tamoxifen treatment.

Animals were housed in groups of five per cage, under a 12-h light/12-h dark cycle, with controlled room temperature and humidity, and then housed singly for behavioral experiments.

Food and water were provided *ad libitum*. All behavioral experiments were performed on adult male mice aged 8–14 weeks during the light cycle. All animal procedures were carried out in accordance with and approved by the Columbia University Institutional Animal Care and Use Committee.

Compounds. RN-1 (Tocris, 4977), PBIT (Sigma, SML1058), IOX-1 (Sigma, SML0067), tranylcypromine (Cayman Chemical, 10010494), ORY-1001 (Cayman Chemical, 19136), and SP2509 (Cayman Chemical, 19136).

Antibodies.

Western blots: The following antibodies were used: hSETD1A (rabbit, 1:500, A300-289, Bethyl), Histone H3 (rabbit, 1:30000, ab1791, abcam). Mef2c/a (rabbit: 1: 1000, ab64644, Abcam).

Immunocytochemistry and immunohistochemistry: The following antibodies were used: NeuN (mouse, 1:100, Millipore), Tau-1 (mouse,1:500, Millipore), Tbr2 (rabbit, 1:200, Abcam), EGFP (rabbit,1:200, Invitrogen), PV (mouse, 1:1000, Swant), GFAP (mouse, 1:300, Chemicon), PH-3 (rat, 1:500, Abcam). RFP (rabbit, 1:300, Rockland). All secondary antibodies (goat, Molecular Probes) were used at 1:200 and 1:500 dilution for immunocytochemistry and immunohistochemistry, respectively.

ChIP analysis: hSETD1A (rabbit, A300-289, Bethyl), Mef2 (rabbit, sc-313, Santa Cruz Biotechnology), H3K4me1 (rabbit, ab8895, Abcam), H3K4me2 (mouse, 05-1338, Millipore),

H3K4me3 (rabbit, 04-745, Millipore), H3K27ac (rabbit, 39133, Active Motif). LSD1 (rabbit, ab17721, Abcam).

Co-immunoprecipitation. Nuclear extract was prepared from fresh cortex of 8-wk old mice using Nuclear Complex Co-IP Kit (Active Motif. Inc.) according to the manufacturer's instructions. For co-immunoprecipitation analysis, 600 μg each of nuclear extracts were incubated with 10 μg of rabbit anti-Setd1a antibody (Bethyl) or 10 μg of normal rabbit IgG (Jackson ImmunoResearch Inc.) at 4°C, overnight. The immuno-complex was adsorbed onto protein A-agarose (Sigma) beads by incubating one hour at 4 °C. For immunoblot analysis we used rabbit anti-Mef2a/c antibody (1:1000, Abcam) or anti-Setd1a antibody (1:500) as a primary antibody and mouse monoclonal anti-rabbit IgG light chain (HRP) (1:5000 and 1:8000, clone SB62a, ab99697, Abcam) as a secondary antibody.

Immunocytochemistry. Primary neurons were fixed in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) (pH 7.4) for 10 min at room temperature (RT), washed 3 times for 5 min in 1×PBS, permeabilized in 0.1% Triton X-100 in PBS for 5 min and washed again in PBS (3 x 5 min). Coverslips were blocked for 30 min in 3% bovine serum albumin supplemented with normalized serum and incubated for 1 hr in primary antibodies diluted into blocking buffer and washed in PBS (3 x 5 min). Prior to application of secondary antibodies, cells were blocked again in blocking buffer supplemented with 0.5M NaCl for 30 min at RT. Secondary antibodies

were applied in blocking/0.5M NaCl for 30 min at RT. Following another round of washing, the coverslips were mounted onto SuperfrostPlus glass slides using Prolong Gold (Invitrogen).

Immunohistochemistry. All the following experiments utilized brains from P8.5 or 6-weeks old males, perfused with PBS and 4% PFA, and post-fixed in PFA overnight. A vibratome was used to cut 60 µm coronal sections. Free-floating sections were blocked and permeabilized overnight at 4°C in PBTX (PBS/Triton X-100) containing 5% normal goat serum (PBTX / 5% NGS). Primary antibodies were diluted in PBTX / 5% NGS and incubated with sections for 24 hrs at 4°C. Sections were washed overnight at 4°C in 1× PBS. Sections were incubated with secondary antibody (1:500, diluted in PBTX / 5% NGS) for 4 hrs at RT. Following another round of washing in 1× PBS, sections were incubated with TO-PRO3 (1:2500 in PBS) for 30 min, washed in 1× PBS (4 x 15 min), and mounted on slides using Prolong Gold (Invitrogen).

Primary neuronal culture and transfection. Primary cultures were prepared from cortices dissected bilaterally from E17 embryos. Cortices were individually dissociated using fire-polished Pasteur pipettes, counted, and plated in DMEM supplemented with 0.5% penicillin/streptomycin and 10% FBS. Neurons for immunocytochemical analyses were plated on poly D-lysine coated 12 mm coverslips (2 x 10^4 cells/coverslip). Following 1.5 hr incubation, the modified DMEM was replaced with neurobasal media supplemented with B27 and Glutamax (Invitrogen) in which neurons were cultured for 5 days. Post-hoc genotyping was performed on

DNA extracted from tissue taken from each embryo. Transient transfections were carried out with Lipofectamine 2000 (Invitrogen) at DIV4 using the protocol suggested for neuronal transfection. The original media used to culture cells prior to transfection were replaced in each well/dish following 2 hr of incubation at 37°C with transfection reagents.

Administration of Kdm inhibitors into primary neuronal cultures. KDM inhibitors were dissolved at 10 mM in 100% DMSO. The compounds were diluted to 0.1 nM - to 30 μM in fresh neurobasal media supplemented with B27 and Glutamax. Primary neuronal cultures were placed in neurobasal media with compounds at DIV3 and cultured for 48 hours. Transient transfections of EGFP-construct were carried out with Lipofectamine 2000 at DIV4. Media with compounds used to culture cells prior to transfection were replaced in each well/dish following 2 hr of incubation at 37°C with transfection reagents.

Image acquisition. Images were acquired blind to the genotype.

Primary neuronal cultures: Confocal images of neurons were obtained blind to genotype with the LSM510 using a Zeiss 20× objective with sequential acquisition setting at 2048 x 2048 pixels resolution. Images were taken with setting as z-series stacks (at 1.0 µm depth intervals) with constant parameters for brightness, contrast, and pinhole size. To evaluate the axon length and branching number, all EGFP-expressed cell, which were polarized with single axon

(approximately 50 - 80 neurons/coverslip), were analyzed from at least 5 coverslips derived from 3 embryos for each genotype.

Tissue sections: To analyze the NeuN and PV positive cell number at 6-weeks, images were taken with setting as a single scan using 20× Zeiss objectives with constant parameters for brightness and contrast. To analyze the contralateral axon branch number at P8.5, images were taken with setting as z-series stacks (at 0.5 µm depth intervals) using 40× Zeiss objectives with constant parameters for brightness and contrast. To analyze the dendritic spine number at 6-weeks, images were taken with setting as z-series stacks (at 0.5 µm depth intervals) using 60× Zeiss objectives with constant parameters for brightness and contrast. To analyze the dendritic complexity at 6-weeks, images were taken with setting as z-series stacks (at 0.5 µm depth intervals) using 40× Zeiss objectives with constant parameters for brightness and contrast.

Image quantification. All image analysis was conducted blind to the genotype.

Quantification of axonal morphology in vivo: To analyze the axonal morphology quantitatively, we performed the following procedures for each dataset obtained from the 60-µm coronal sections. For the contralateral axon, terminal branching at P8.5, EGFP expressed axons at contralateral layer 1-4 were analyzed by projecting the *z* series and manually tracing the axon using the overlay tools of the LSM Image Examiner 5 program (Zeiss). Protrusions more than 6 µm long were categorized as axon branches. The axon terminal architecture of each neuron

was also characterized by designating the order of a branch (primary, secondary and so on). Once the elaborations of the axon branch tree were outlined, branch numbers were counted manually and analyzed. We compared at least 20 axons from four brains of each genotype in three independent experiments using the Student's *t*-test.

Quantification of cortical axonal morphology in vivo in adult mice. To analyze and visualize the axon branching in adult brain, we performed the following procedures for each dataset obtained from 40-μm coronal sections. Images acquired for the EGFP fluorescence analysis were taken with sequential acquisition setting at 2,048 × 2,048 pixel resolution as z-stack (3 images taken at intervals of 2.5 μm) from the each sections of frontal cortex. The region of interest, including the EGFP expressing contralateral cortical axon column in the somatosensory cortex (S1), was defined by a 400-μm line parallel to the each cortical layer [layer 1, layer 2/3, layer 4, layer 5, layer 6, and white matter (WM)]. The cortical layers were identified by NeuN immunostaining. The immunofluorescence in each layer was measured by ImageJ software. Three images of strongest EGFP intensity from three sections were averaged for each cortical layer, and normalized by EGFP expressed cell number in ipsilateral layer 2/3 as described below. All settings of threshold were maintained constant across genotypes.

Following a modified method previously described (Mukai et al., 2015), immunofluorescence intensity in the contralateral axon branches was normalized to AAV2-

Syn/EGFP virus infection efficiency in the ipsilateral somatosensory cortex. To quantify the infection efficiency in the injection site of ipsilateral S1 cortex, images acquired for the EGFP and NeuN fluorescence analysis were taken with sequential acquisition setting at 2,048 × 2,048 pixel resolution as a single image. Using ImageJ cell count tool, EGFP and NeuN labeled cells were counted in the region of interest. Three images of highest number of cells from three sections were averaged for virus infection efficiency.

Quantification of axonal morphology in vitro: For *in vitro* analysis, a plasmid expressing EGFP driven by the β-actin promoter was transfected into cultures to fill the cells and facilitate visualization of all processes. Neuronal processes stained for the axonal marker Tau-1 were traced using the overlay tools of the LSM Image Examiner 5 program (Zeiss). Protrusions more than 8 μm long were categorized as axon branches. The axon architecture of each neuron was also characterized by designating the order of a branch (primary, secondary and so on). Once the elaborations of the axon branch tree were outlined, branch points were counted manually, and the length of primary axon (emanating directly from the soma) and its branches were analyzed and compared across at least 3 independent experiments using the Student's *t*-test.

Image analysis of dendritic complexity: The complexity of the basal dendrites of L5 neurons in the PrL area of Thy1-GFP/M mice was analyzed by projecting the z-series to maximum intensity and manually tracing the dendritic tree. Sholl analysis was conducted using concentric radii of 20 μm

extending from the center of the soma. Intersections, branches and terminals were counted by distance from the soma and differences evaluated by repeated measures ANOVA.

Image analysis of dendritic spines: Z-series stacks were subjected to three-dimensional reconstruction using Zeiss LSM 510 Meta. Dendritic protrusions were classified as spines (mushroom, long, stubby or thin) or filopodia, based on the parameters previously outlined (Mukai et al., 2008). Size determination of mushroom spines was carried out using Zeiss LSM image browser overlay tools. The longest straight line in the spine head was counted as the head diameter. The length of the entire spine (including head and stalk) was measured using a bent-line tool. Differences in the number, size and shape of spines were evaluated using the Student's *t*-test.

Electroporation *in utero*. Pregnant mothers with E15.5 embryos were anaesthetized, and a 1 inch incision was made and one uterine horn removed. A total of 0.1–0.5 ml DNA (1–2 mg/ml) was injected into one ventricle. Four to five 40–45 V pulses were applied, each pulse lasting 50 ms, with a 1 s pause between each pulse using an ECM-830 square wave electroporator (BTX) as previously described (Mukai et al., 2015; Saito, 2006).

Virus injection. rAAV2/hSynapsin/EGFP was obtained from Addgene. Viral titer was 3 × 10^{12} particles per ml. Mice (12-14 weeks old) were deeply anesthetized under isoflurane and placed in a stereotaxic frame. Bregma and lambda were used for leveling and zeroing, and

bilateral cranial windows were excised over S1 cortex, centered at 0.97 mm posterior, 2.9 mm lateral. Virus was administered to layer 2/3 by an injection pump at a rate of 100 nL/min via an elongated 10-40 μm-diameter glass micropipette. 200 nl were infused at single injection site. Acrylic dental cement was used to cover the craniotomy after infusion.

Electrophysiology in medial PFC slice preparations. Field recordings were performed on 8- to 10-week-old male *Setd1a*$^{+/-}$ mutant mice and their WT littermates. Isoflurane was used to anesthetize mice that were then decapitated. After a skull incision, the brain was removed and placed in ice-cold dissecting solution, which contained 195 mM sucrose,10 mM NaCl, 2.5 mM KCl, 1 mM NaH_2PO_4, 25 mM $NaHCO_3$, 10 mM glucose, 4 mM $MgSO_4$, and 0.5 mM $CaCl_2$ and was bubbled with 95% O_2/5% CO_2 mixture. Coronal brain slices (300µm) containing the prelimbic and infralimbic prefrontal cortex were cut using a vibratome (VT1200S; Leica, Nussloch, Germany). The freshly cut mPFC slices were immediately transferred to an interface chamber and allowed to recover for at least 2 h at 34–36°C. During all recordings, the slices were continuously perfused with artificial cerebrospinal fluid (aCSF) (bubbled with 5% CO_2/95% O_2) that had the following composition: 124 mM NaCl, 2.5 mM KCl, 1 mM NaH_2PO_4, 25 mM $NaHCO_3$, 10 mM glucose, 1 mM $MgSO_4$, and 2 mM $CaCl_2$. The aCSF was maintained at 34–36°C and fed by gravity at a rate of 2–3 ml/min. Field EPSPs (fEPSPs) were recorded via a glass microelectrode (3–5 MΩ) filled with aCSF and placed in L5 of the mPFC (600–700 µm from midline). The

stimulation site was always aligned ~200 μm away from the recording site along the axis perpendicular to the pial surface. Basic synaptic transmission was characterized at 0.033 Hz, with stimulation intensities of 1–24 V (pulse duration, 0.1 ms). Subsequent experiments were performed at the stimulus intensity that generated a fEPSP one-third of the maximum fEPSP obtained at 24 V. Short-term synaptic facilitation was induced using a paired-pulse protocol with ISIs of 20, 50, 100, 200, 400, and 800 ms. To assess STD, fEPSPs were evoked by using a 40-pulse train at 5, 20 and 50 Hz (pulse duration, 0.1 ms). LTP was induced by one 800-ms (40 pulses), 50-Hz train after a stable 10 min baseline and monitored during 15 min. Then, 15 min after the first tetanus, four additional 50 Hz trains (separated by 10 s) were applied. The fEPSPs were then monitored for 40 min. Fiber volley was quantified by measuring the amplitude of the first peak negativity of the field responses, and the fEPSPs were quantified by measuring the initial slope of the second peak negativity of the responses. Electrophysiological signals were acquired using an extracellular amplifier (Cygnus Technologies) and pClamp 10 (Molecular Devices). Statistical analyses were performed using SigmaPlot 9.0, and GraphPad Prism 4. A two-way repeated-measures ANOVA was used to test differences between genotypes. Data are presented as means ± s.e.m. N indicates number of animals and n indicates number of slices. Asterisks indicate significant differences. All recordings and the majority of data analyses were done blind to the genotype.

Whole cell recordings were performed on 8- to 10-week-old male *Setd1a*$^{+/-}$ mutant mice and their WT littermates. Following brain slicing mPFC slices were immediately transferred to a recovery chamber and incubated at room temperature in recording solution (ACSF) for a minimum of 1h before recording. At the time of recording, slices were transferred to a submerged recording chamber and continuously perfused with ACSF. Whole-cell patch clamp recordings were made using borosilicate glass pipettes (initial resistance 3.0–5.5 MΩ). The internal solution used contained (in mM): $KMeSO_4$ 145, HEPES 10, NaCl 10, $CaCl_2$ 1, $MgCl_2$ 1, EGTA 10, MgATP 5, Na_2GTP 0.5, pH 7.2 with KOH, adjusted to 290 mOsm with sucrose. Layer 2/3 pyramidal neurons in mPFC prelimbic and infralimbic regions were identified by morphology and only neurons exhibiting spike frequency accommodation within action potential trains were included in analysis (Madison and Nicoll, 1984).

For recordings assessing neuronal excitability, action potential firing was assessed in current-clamp mode in response to incremental (20 pA steps), depolarizing current injections of 500 ms duration. Bridge balance of series resistance was employed and recordings with series resistance >20 MΩ were rejected. The resting membrane potential was determined within the first 5 s of achieving whole-cell access. For current-step experiments, the resting membrane potential of all cells was adjusted to approximately -70 mV by injection of a small standing current. Input resistance was calculated in current-clamp mode at -70 mV from the quasi steady-state

(after 450-500 ms) voltage response to a 40 pA hyperpolarizing current-step. The membrane time constant ($\tau_{memb.}$) was calculated by fitting a single exponential to the voltage response to hyperpolarizing current steps (for 20-60 pA steps, using the voltage response during the first 200 ms) employing the built-in Chebyshev routine in Clampfit. For simple comparisons of single metrics, unpaired t-tests were employed. For AP responses across current steps, 2-way repeated measures ANOVA statistics were employed.

Behavior analysis.

Open field: Mouse activity was monitored in an illuminated chamber equipped with infrared sensors to automatically record horizontal and vertical activity (Colbourn Instruments). Each mouse was placed in the center of the chamber and its activity was recorded and collected in 1-min bins for 1 h during day 1 and for a 30-min re-exposure 24 h later (day 2).

T-maze: 6 weeks old animals were trained on a non-match to sample T-maze task. The behavior protocol began with 2 days of habituation during which animals freely explored the maze for 10 minutes, followed by 2 days of 28 shaping when animals were required to alternate between goal arms of the maze to receive food rewards. After shaping was completed, training and testing sessions were conducted. Each trial of the task consisted of a sample and choice phase. In the sample phase, a mouse ran down the center arm of the maze and was directed into one of the goal arms. The mouse returned to the start box where it remained for a delay of 10 seconds. In

the choice phase, the mouse was required to select the arm opposite to that visited in the sample phase to receive a reward. Animals were given daily training of 10 trials until they reached criterion performance, defined as performance of at least 70% correct per day for 3 consecutive days. After criterion was reached, animals completed daily testing sessions composed of 20-25 trials for delays of 10 s, 30 s, and 60 s. In the case of rescue experiments following treatment with tamoxifen, TCP, or ORY-1001, animals completed daily testing sessions composed of 12 trials for 10 s, 20 s, and 40 s delays.

Social memory-Direct interaction with juveniles: All mice were housed two to five in each cage and given *ad libitum* access to food and water. They were kept on a 12-h (6:00 to 18:00) light–dark cycle with the room temperature regulated between 21-23 °C. Behavioral tests were performed during the light cycle in a testing room adjunct to the mouse housing room, which minimizes any transportation stress. The social memory/direct interaction test with juveniles was adapted from previously described (Kogan et al., 2000). Immediately prior to the experimental sessions, 10-12-week-old *Setd1a*$^{+/-}$ and WT littermates were transferred to the testing room and placed into individual cages, identical to the ones used for housing, where they were allowed to habituate to the new environment for 15 min. Male juvenile mice (C57BL/6J, 4-5-week-old) were also placed in the testing room in their home cages and allowed to habituate a similar amount of time. Testing began when a novel juvenile mouse was introduced to a cage with one of the adult

experimental mice. Activity was monitored for 5 min (trial 1) and scored online by a trained observer blind to the genotype of the test mice for social behavior (anogenital and nose-to-nose sniffing, close following and allogrooming) initiated by the. experimental subject, as described in Hitti et. al. (Hitti and Siegelbaum, 2014). After an inter-trial interval of 1 h, the experimental mice were introduced with either the previously encountered mouse or a novel mouse again for 5 min (Trial 2). The time spent in social interaction during trial 1 was subtracted from the social interaction time during trial 2 to obtain the difference score.

Fear Conditioning: Experimentally naive mice were used for fear conditioning in plastic, sound-attenuating cabinets containing a house light. On the 1st day, mice were placed in chambers and received 2 pairings of a tone (30 s, 82 db) and a co-terminating shock (2 s, 0.7 mA). On the 2nd day, initially, mice were placed in the chambers for 6 min, in the absence of tone and shock, and freezing was scored for the whole session, corresponding to hippocampal-dependent contextual fear memory. 2 h later, the context and handling of the mice was changed to assess conditioned fear of the tone alone (cue fear memory). Freezing was assessed before and during the tone presentation to distinguish fear generalization and specific cue fear memory for cue.

Novel Object Recognition: Mice were given a single 10 min trial of habituation to the room, the testing arena and to the procedure over three consecutive days to reduce stress and prevent

a neophobic response and consequently to promote the exploratory activity of mice toward the objects (days 1-3). On the two following days mice were given a 5 min trial of exposure to two identical objects (days 4, 5). A memory test trial was performed 1 h later where one of the familiar objects was replaced with a novel object. 24 h later mice were placed back in the arena and were given another 5 min memory test in which one of the identical objects they explored during the training trial was replaced with a novel one. The location of the novel object was randomized between trials and genotypes for both the 24 h and the 1 h novel object recognition test. A minimal exploration time for both objects during the test phase (~20 s) was used to ensure a similar exploration time of the two objects and between animals, independently of their individual exploratory activity. All trials were video recorded and analyzed blindly by two investigators. Results were expressed as a discrimination index, calculated as the amount of time each mouse spends exploring the novel object divided by the total exploration time for both objects.

Statistical analysis: SPSS 20.0 was used for statistical analysis and Prism 6 (GraphPad) to graph data. Statistical significance was assessed by one-way ANOVA, or two-way repeated-measures ANOVA where appropriate.

Tamoxifen preparation and feeding. A modified version of a tamoxifen feeding protocol was used as previously described in Mei, et al (Mei et al., 2016). Tamoxifen (Sigma, T5648) was dissolved in corn oil (Sigma, C8267) at 20 mg/ml by vortexing. Freshly prepared

tamoxifen was protected from light by aluminium foil. Animal feeding needles (Harvard Apparatus, 52-4025) were used for oral gavage. To avoid toxicity of tamoxifen, the following dosages were used for adult animals (8 weeks): mice at 17–21 g body weight were fed 5 mg/day; mice at 22–25 g body weight were fed 6 mg/day; mice at 26–29 g body weight were fed 7mg/day; mice at 30–35 g body weight were fed 8mg/day. Adult animals were fed for 5 consecutive days followed by 2 days of rest. Animals were then fed for 5 more consecutive days followed by 3 weeks of rest. Corn oil was used as a control. The mice fed with tamoxifen and with corn oil were housed separately to avoid contamination.

TCP preparation and injection. Tranylcypromine (TCP, Sigma, P8511) was dissolved in PBS at 0.75 mg/ml by vortexing. Mice at 17-20 g body weight were injected 2.7 mg/kg/day; mice at 20-28g body weight were injected 3.0 mg/kg/day. Adult animals were injected for 14 consecutive days followed by a non-match to sample T-maze task. During the behavior task, mice were injected with the same dose at 6 pm daily.

ORY-1001 preparation and injection. ORY-1001 (Cayman Chemical, 19136) was dissolved in PBS at 1mg/ml by vortexing. Mice at 17-20 g body weight were injected 9.0 μg/kg/day; mice at 20-28g body weight were injected 10.0 μg/kg/day. Adult animals were injected for 14 consecutive days followed by a non-match to sample T-maze task. During the behavior task, mice were injected with the same dose at 6 pm daily.

RNA sequencing. Total RNA was extracted from prefrontal cortices of 6-week-old mice using RNeasy Mini Kit (Qiagen). We analyzed each four male *Setd1a*$^{+/-}$ mice and their wild-type (WT) male littermates. Quality of RNA samples was assessed by Bioanalyzer (Agilent). Mean (± SD) RNA integrity number (RIN) was 9.3 (± 0.5) in the mutant mice and 9.7 (± 0.2) in WT. An RNA-seq library was prepared from ~400 ng of total RNA by using TruSeq RNA prep kit (Illumina). Messenger RNAs were enriched by poly-A selection. Libraries were sequenced on HiSeq2500/HiSeq4000 (Illumina) with single-end 100bp reads at Columbia Genome Center.

Data processing and analysis of DE genes. We used RTA (Illumina) for base calling and bcl2fastq (Illumina, version 2.17) for converting BCL to fastq format, coupled with adaptor trimming. We mapped the reads to the reference mouse genome (UCSC/mm10) using STAR ((Dobin et al., 2013) (version 2.5.4b). Read counts and normalization of the read count data and analysis of differentially expressed genes (DE genes) were performed by using DESeq2 (Love et al., 2014) (version 1.10.2) with default parameters. By performing principle component analysis (PCA) with our data of eight PFC samples (each five *Setd1a+/-* and WT) and read count data from olfactory bulb (OB) and cerebellum, we confirmed that our data of PFC form a single cluster distant from clusters of OB and cerebellum (data not shown).

ChIP-Seq analysis. Whole prefrontal cortex was dissected from 6-week old wild-type (WT) mice. Dissected tissue was thoroughly minced on ice using razor blades, then fixed with

EGS for 30 min and/or for 5 min at room temperature with methanol-free formaldehyde (Pierce, Thermo Fisher Scientific) diluted to 1% in PBS. Fixation was quenched by adding 1/10th volume of 1.25 M glycine. Fixed tissue was collected by centrifugation at 200 g for 5 min at 4ºC, then washed twice with cold PBS. Washed tissue was flash-frozen with liquid nitrogen then cryofractured using a Covaris CryoPrep Impactor on power setting 6. Nuclei and chromatin immunoprecipitation were performed as in Markenscoff-Papadimitriou *et al*.(Markenscoff-Papadimitriou et al., 2014): briefly, the cryofractured tissue was lysed in ChIP Lysis Buffer (50 mM Tris-HCl pH 7.5, 150 nM NaCl, 0.5% NP-40, 0.25% Sodium Deoxychoalate, 0.1% SDS) in rotation for 40 min at 4ºC. Nuclei were centrifuged (1700x g, 5 min, 4ºC), and the pellet was resuspended in shearing buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8, 0.25% SDS), then sheared on a Covaris S2 sonicator (12 or 16 min, 2% Duty Cycle, Intensity 3, 200 cycles per burst, frequency sweeping). Sheared chromatin was centrifuged (10,000 g for 10 min at 4ºC) to remove insoluble material. Sheared chromatin was diluted 5-fold with ChIP Dilution Buffer (CDB: 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS) and pre-cleared for two hours at 4ºC with protein G dynabeads (Thermo Fisher Scientific). Each ChIP was set up with the appropriate antibody (10 μg of anti-Setd1a, 1 μg of anti-H3K27ac, -H3K4me2 and -H3K4me3, 3 μg of anti-H3K4me1, 5 μg of Mef2) and 10-20 μg of cleared chromatin, incubating then overnight at 4ºC. Protein G dynabeads were blocked overnight with 2 mg/ml yeast tRNA

(Thermo Fisher Scientific), then added to antibody bound chromatin and rotated for 3 hr at 4ºC. Bead-bound chromatin was washed 5 times with LiCl Wash Buffer (100 mM Tris-HCl pH 7.5, 500 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate) and once with TE (pH7.5). DNA was eluted from beads by incubating at 65ºC for 30 min with 25 μL ChIP Elution Buffer (1% SDS, 0.1 M Sodium Bicarbonate). This elution was repeated and the combined elution fractions were incubated overnight at 65ºC with proteinase K (Thermo Fisher Scientific) and RNaseA (Thermo Fisher Scientific). ChIP DNA was purified 1.8X of AMPure XP beads and eluted in 30 μl of nuclease free water. Libraries were prepared with Ovation Ultralow System V2 1-16 (NuGEN technologies) according to manufacturers' recommendation. All samples were multiplexed (in batches of 6) and sequenced using Illumina HiSeq as 50 bp or 75 bp single end. Two biological replicates for each ChIP-Seq for each transcription factor or histone mark were sequenced. As control, 1% chromatin input was also sequenced.

ChIP-Seq data processing, peak calling and motif discovery. Reads were aligned to the *Mus musculus* genome version mm10 using bowtie2 v2.2.6 (Langmead and Salzberg, 2012). Poor quality and duplicated reads were removed using respectively samtools v1.3.1(Li et al., 2009) and Picard (http://broadinstitute.github.io/picard). To measure the reproducibility between biological replicates and to avoid arbitrary thresholding biases for both Setd1a and Mef2, we perform IDR following the ENCODE framework (Landt et al., 2012) using MACS v2.1.1 as peak

caller (DOI: 10.1186/gb-2008-9-9-r137) and IDR 2.0.2 (Li et al., 2011). We defined good biological replicates when the ratio N1/N2 and Np/Nt was below 2. For all subsequent analysis, biological replicates were merged into single alignment files for each antibody using samtools v1.3.1 (Li et al., 2009) and difference in sequencing depth between the libraries was corrected by normalization for library size using HOMER v4.8.2 (Heinz et al., 2010). Setd1a bound promoter elements were defined as such when they were located within 1 Kb of annotated TSS or they overlapped with H3K4me3; Setd1a bound intergenic peaks were defined as such when they were located in unannotated regions of the genome at least 1 kb away from any annotated TSS; Setd1a bound intronic peaks were defined as such when they overlapped with annotated intron of the *Mus musculus* (version mm10) of the genome. *De novo* motif discovery analysis was performed on feature-separated *Mus musculus*, version mm10, of Setd1a peaks using HOMER v4.8.2 (Heinz et al., 2010) using automatically generated background (sequences randomly selected from the genome, matched for GC% content).

Setd1a binding in promoter, intergenic and intronic regulatory regions. To identify the properties of Setd1a bound regions in the PFC, we compared the coverage of PFC ChIP-Seq for Mef2 and histone marks that we sequenced (see above). Heatmaps were generated calculating the normalized coverage per each ChIP-Seq using deeptools v2.5.0.1 (Ramirez et al., 2016), 2 kilobases around the center of Setd1a peaks into 50 bp bins. Density plots were

generated calculating the coverage of the ChIP-Seq reads 2 kilobases around the center Setd1a peaks into 50 bps bins with deeptools v2.5.0.1 (Ramirez et al., 2016) and plotted using the R-package “ggplot2” and R v3.5.1 (http://www.gbif.org/resource/81287).

Tissue specific Setd1a binding. To investigate the Setd1a chromatin binding pattern between different tissues, we compared our PFC epigenetic H3K27ac landscape with different brain tissues using the ENCODE publicly available H3K27ac ChIP-Seq data from the olfactory bulb and the cerebellum (Shen et al., 2012). To make the data analysis comparable between datasets, we reprocessed the ENCODE data as described above. Heatmaps and density plots were generated as described above.

Conservation of Setd1a peaks in the Euarchontoglire clade. Phastcons were obtained using the 59 vertebrate genomes (Phastcons 60ways) (Siepel et al., 2005). We selected data from the Euarchontoglire clade (including 21 vertebrate genomes) and the sources were downloaded from UCSC Genome browser via http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phastCons60way/. Setd1a peaks were first normalized by size and heatmaps were generated as described above. Clusters were defined by hierarchical clustering of the data. For cluster 2 and 3 (Figure S3E), we split peaks in half: conserved and non-conserved half. Mef2 motif distribution was calculated using HOMER v4.8.2 (Heinz et al., 2010) with the function “annotatedPeaks” and plotted using the R package “ggplot2”.

De novo motif analysis was performed with HOMER v4.8.2 (Heinz et al., 2010) using different backgrounds according to clusters: cluster 1 and 4 (Figure S3E) automatic background; for the conserved half of Setd1a peaks in cluster 2 and 3 (Figure S3E) we used as background the non-conserved half of the peaks, and vice versa for the non-conserved half of Setd1a peaks, we used a s background the conserved half of the peaks.

Conservation of murine Setd1a peaks in the human genome. To evaluate the conservation of Setd1a peaks between mouse and human, we submit the FASTA sequence of the 1178 Setd1a peaks to the online blast platform https://blast.ncbi.nlm.nih.gov/Blast.cgi. We used two algorithms (McGinnis and Madden, 2004) that have different sensitivity: (1) megablast to look for very similar sequences (>95% identity) between the selected species; (2) blastn because it works better for cross-species searches allowing mismatches and small deletions. Using both algorithms we found respectively for megablast and blastn 304 and 691 conserved-Setd1a-bound regions between mouse and human. In both cases, we performed *de novo* motif analysis on the hg38 version of the human genome using HOMER v4.8.2 (Heinz et al., 2010) and automatic generated background.

Dissociation of PFC to single cells and scRNA-Seq analysis. Cells were dissociated using the papain dissociation kit (Worthington) following the manufacturers' recommendation. In brief, the PFC of 2 WT and 2 *Setd1a*$^{+/-}$ mice was dissected and transferred to ice-cold PBS. The

PFC was cut in small pieces using a razor blade and dissociated with a papain dissociation system (Worthington Biochemical). The tissue was incubated for 1h at 37°C on a rocking platform with papain-EBSS (1 PFC/mL). The tissue was triturated 20 times and cells were pelleted (400 g, 5 min, RT). Remaining papain was inhibited by resuspending the cell pellet with Ovomucoid protease inhibitor solution diluted 1:10 in EBSS. An extra 5 ml of Ovomucoid protease inhibitor solution was slowly layered at the bottom of the cell suspension to create a discontinuous gradient and cells were pelleted (100 g, 6 min, RT). To enrich for living cells, we removed dead cells from single cell suspension using the MACS® Dead Cell Removal Kit (Miltenyi Biotec) with the MS Columns following manufacturer's recommendation. In brief, 200 µl per PFC of Dead Cell Removal MicroBeads were added the pelleted cells. Cell were resuspended and incubated for 15 min at RT. The MS columns were washed once with 0.5 ml of Binding buffer. After incubation is complete, the cell suspension was diluted with 500 µl 1X Binding Buffer and the suspension was applied to the MS columns. The flowthough was collected. To increase the number of live cells recovered, the column was washed twice with 1 ml of Binding buffer at RT. Flowthough were combined and live cell were pelleted (400 g, 5 min, RT). Cells were then washed twice with PBS containing 0.04% BSA and the final pellet was resuspended in 300 µl of PBS containing 0.04% BSA and passed through a 40 uM cell strainer. Cell were counted and viability was evaluated by trypan blue staining.

The scRNA-Seq was performed using the 10x Genomic platform and the “Single Cell 3' v2” by the Columbia Genome Center. The scRNA-Seq data were aligned using the 10x Genomics package Cell Ranger 2.1.1 on the mm10 mouse genome version. In particular, UMI reads from the different genotypes and biological replicates were counted using “cellranger count” and combined together using “cellranger aggr”. From the downstream analysis, we removed cells in which we detected less then 1000 genes expressed, using “Seurat” R-package (v. 2.3.4). This package was used for all downstream analysis. In details, counts were first normalized (using the normalization method: "LogNormalize"). Linear dimensionality reduction was then performed on the most variable genes after that the data were regress out for cell-cell variability (i.e. batch effect, cell alignment rate, number of detected molecules etc.). All significant principal components were calculated, selected (using the JackStraw function, 20 PCs) and used to defined cell clusters (with the function “FindClusters”). Clusters were then visualized through tSNE plots using the same number of PCs that were used to define clusters. The different cell types were finally defined by the known expression of genes that have been previously describe to mark specific subpopulation of neuronal and non-neuronal cells (Figures S5B, S5C, and S5D). Genes with average expression < 0.1 normalized UMI counts in the whole dataset were removed from all analyses performed.

Sorting of PFC pyramidal neurons and parvalbumin expressing interneurons to identify hypersensitive genomic regions. Emx-Cre mice (Jackson lab) and Parv-Cre mice

(Jackson lab) were crossed to CRE inducible tdTomato mice (Jackson lab) so that pyramidal neurons or parvalbumin expression interneurons expressed the fluorescent protein tdTomato. The PFC of these mice was dissociated as described previously. After dissociation cells were resuspended in sort media supplemented with 100 U/mL DNase I (Worthington Biochemical), 4 mM $MgCl_2$, and 500 ng/mL DAPI (Invitrogen). These cells were passed through a 40 uM cell strainer, and then FAC sorted. Live cells were selected by gating out DAPI positive cells and positive cells were gated by tdTomato fluorescence. Fifty thousand cell were used to prepare ATAC-seq libraries as described in references (Buenrostro et al., 2015; Monahan et al., 2017). In brief, cells were pelleted (500 g, 5 min, 4°C) and resuspended in ice cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630). Nuclei were pelleted (1000 g, 10 min, 4°C) and resuspended in transposition reaction mix prepared from Illumina Nextera reagents (for 50 μL: 22.5 μL water, 25 μL 2xTD buffer, 2.5 μL Tn5 Transposase). Transposed DNA was column purified using a Qiagen MinElute PCR cleanup kit (Qiagen) and amplified using barcoded primers and NEBNext High Fidelity 2x PCR Master Mix (NEB). Amplified libraries were purified using Ampure XP beads (Beckman Coulter). ATAC-Seq data were aligned and processed as previously described for ChIP-Seq data.

Human ATAC-Seq data to evaluate the accessibility of Setd1a conserved regions. ATAC-Seq data from postmortem human brain were downloaded from (Fullard et al., 2018). The fastq file were realigned as described previosly on the hg38 genome.

Interval-based enrichment analysis of Set1a target genes. We converted the genomic positions of Set1a peaks to their human homolog using the Blastn and Megablast tools from (mmu10 to hg19) and used the permutation test implemented RegioneR (Gel et al., 2016) package in R to determine if there is any significant overlap between these peaks and SCZ genomic signals identified in various previous GWAS. We also annotated all ChIP-Seq peaks according to their positions in regard to the gene content in the mouse genome using the PAVIS annotation tool (https://manticore.niehs.nih.gov/pavis2/, mouse Ensembl GRCm38-r82). An intronic interval is defined as a ChIP-Seq peak that hits an intron of a gene. An intergenic interval is defined as a ChIP-Seq peak that is 5 Kbp away from the TSS of a gene and 1kbp away from the end transcription.

Gene-based enrichment analysis of Set1a target genes among GWAS signals. We annotated all target genes of intronic ChIP-Seq peaks using the PAVIS annotation tool. We then map identified mouse target genes of Setd1a and Mef2a/c to their human homologs by their gene symbols based on (http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt) to generate the human target gene set of Setd1a (including 277 high-confidence Setd1a target

genes). We then utilized the INRICH analysis tool (Lee et al., 2012), a permutation-based approach that controls for potential biases caused by SNP density, gene density and gene size to analyze the overlap between the human orthologue targets and previously reported GWAS signals to determine if there is any significant overlap. For the INRICH analysis, the reference gene list was downloaded from the INRICH website (http://atgu.mgh.harvard.edu/inrich/resources/entrez_gene.hg19.map.gz). The reference SNP list of GWAS studies (SNPs on a DNA microarray) was compiled from the table browser (hg19.snpArrayAffy6 and hg19.snpArrayIlluminaHumanOmni1_Quad) of the UCSC genome browser, which covers at least 95% of intervals analyzed. The default setting of 5000 permutation replicates was used for each trait tested. We prepared the various GWAS signals using the summary association data from the NHGRI (National Human Genome Research Institute) GWAS Catalog (Welter et al., 2014). Specifically, we first downloaded the list of SNPs associated with various human traits from the GWAS Catalog (http://www.ebi.ac.uk/gwas, the gwas_catalog_v1.0.1 file, accessed June 2016). For each phenotype, genome-wide significantly associated intervals were defined as follows using the threshold for linkage disequilibrium (LD) used in the SCZ GWAS by PGC ($r^2 > 0.6$) (Schizophrenia Working Group of the Psychiatric Genomics, 2014). We extracted autosomal SNPs with genome-wide significant association ($P < 5 \times 10^{-8}$) and with MAF > 0.05 in the 1000 Genomes Project March 2012 EUR dataset [downloaded as a part of the LocusZoom package

(Pruim et al., 2010)] from the full dataset of GWAS Catalog (we denote these as “index SNPs”). An associated genomic locus for each index SNP was defined as the genomic region containing any SNPs that are: 1) in LD with the index SNP at $r^2 > 0.6$ and 2) within +/-10 Mb of the index SNP. SNPs in LD with the index SNP were identified by using PLINK (Purcell et al., 2007) with the 1000 Genomes Project March 2012 EUR dataset. After performing these procedures, we obtained 124 intervals from the SCZ GWAS signals in the Catalog. As a control, we also performed the same analyses on GWAS signals of 15 non-CNS diseases/traits that have similar associated interval size (from 74 intervals to 174 intervals) that are comparable to the size of SCZ GWAS intervals. This procedure of generating independent GWAS signals is a pre-requisite for the INRICH analysis. To ensure this procedure can produce results comparable to the in-house method used in the original publication of PGC SCZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics, 2014), we also performed an analysis based on the pre-defined data of 108 SCZ-associated loci downloaded from the PGC portal (https://www.med.unc.edu/pgc/files/resultfiles/scz2.regions.zip; results are shown as “SCZ_PGC” in the corresponding Figures/Tables).

Gene-based enrichment analysis of Set1a target genes among LoF DNM-hit genes. Gene-based enrichment analysis of the human target gene set of Setd1a (described above) among LoF DNMs was performed by using DNENRICH (Fromer et al., 2014), a statistical software

package for calculating permutation-based significance of gene set enrichment among genes hit by de novo mutations accounting for potential confounding factors such as gene sizes and local trinucleotide contexts DNENRICH analyses were performed with one million permutations using the following input files: 1) gene name alias file: the default dataset included in the DNENRICH package, 2) gene size matrix file: the default dataset included in the DNENRICH package (for RefSeq genes), 3) gene set file: the set of 277 high-confidence Setd1a target genes (HGNC symbols available, not including *SETD1A*); all genes were equally weighted, 4) mutation list files: lists of genes hit by LoF DNMs in NDD (SCZ, ID/DD and ASD), congenital heart diseases (CHD) and control subjects in denovo-db (Turner et al., 2017) or Refs.(RK et al., 2017; Xu et al., 2012; Xu et al., 2011); all mutations were equally weighted, and 5) background gene file: the default gene list in the gene size matrix. The numbers of LoF DNMs in NDD, SCZ, ID/DD, ASD, CHD and controls were 1,874, 95, 986. 793, 173 and 202, respectively. For analyses restricting the input LoF DNMs to those in genes intolerant to LoF mutations, we used the two metrics, pLI scores (Lek et al., 2016) (threshold = 0.9) and LoF FDR (Petrovski et al., 2013) (threshold = 0.01). Genes for which pLI or LoF FDR was not available were excluded from the background gene list in the corresponding analyses considering pLI or LoF FDR.

GO enrichment and CSEA analysis of Setd1a target genes: Analysis of gene ontology (GO) terms enriched among Setd1a target genes was performed by using ToppGene with default

parameters for four categories (Molecular Function, Biological Process and Cellular Component of GO terms, and Disease terms defined by DisGeNET). To study the expression pattern of Setd1a targets with intronic ChIP-Seq peaks, in the context human brain development, Setd1a target genes were imported into the CSEA tool (Dougherty et al., 2010) (http://genetics.wustl.edu/jdlab/csea-tool-2/). The output FDR *P* values of CSEA analysis were imported into the MultiExperiment Viewer (ver.4.9.0) as heatmap color code to generate the heatmap view. Genes that overlap with developmental stage and cortical signature genes were summarized from SEA analysis (pSI < 0.05).

References

Arguello, P.A., and Gogos, J.A. (2012). Genetic and cognitive windows into circuit mechanisms of psychiatric disease. Trends Neurosci *35*, 3-13.

Barak, O., and Tsodyks, M. (2014). Working models of working memory. Curr Opin Neurobiol *25*, 20-24.

Barch, D.M., and Ceaser, A. (2012). Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn Sci *16*, 27-34.

Barnes, S.A., Wijetunge, L.S., Jackson, A.D., Katsanevaki, D., Osterweil, E.K., Komiyama, N.H., Grant, S.G., Bear, M.F., Nagerl, U.V., Kind, P.C.*, et al.* (2015). Convergence of Hippocampal Pathophysiology in Syngap+/- and Fmr1-/y Mice. J Neurosci *35*, 15073-15081.

Benamer, N., Marti, F., Lujan, R., Hepp, R., Aubier, T.G., Dupin, A.A.M., Frebourg, G., Pons, S., Maskos, U., Faure, P.*, et al.* (2018). GluD1, linked to schizophrenia, controls the burst firing of dopamine neurons. Mol Psychiatry *23*, 691-700.

Bledau, A.S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D.C., Fu, J., Kranz, A., Stewart, A.F.*, et al.* (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. Development *141*, 1022-1035.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-29.

Choi, J., Jang, H., Kim, H., Lee, J.H., Kim, S.T., Cho, E.J., and Youn, H.D. (2014). Modulation of lysine methylation in myocyte enhancer factor 2 during skeletal muscle cell differentiation. Nucleic Acids Res *42*, 224-234.

Crabtree, G.W., Sun, Z., Kvajo, M., Broek, J.A., Fenelon, K., McKellar, H., Xiao, L., Xu, B., Bahn, S., O'Donnell, J.M.*, et al.* (2017). Alteration of Neuronal Excitability and Short-Term Synaptic Plasticity in the Prefrontal Cortex of a Mouse Model of Mental Illness. J Neurosci *37*, 4158-4180.

Deciphering Developmental Disorders, S. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433-438.

Del Pino, I., Rico, B., and Marin, O. (2018). Neural circuit dysfunction in mouse models of neurodevelopmental disorders. Curr Opin Neurobiol *48*, 174-182.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res *38*, 4218-4230.
Fahrner, J.A., and Bjornsson, H.T. (2014). Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. Annu Rev Genomics Hum Genet *15*, 269-293.
Fenelon, K., Xu, B., Lai, C.S., Mukai, J., Markx, S., Stark, K.L., Hsu, P.K., Gan, W.B., Fischbach, G.D., MacDermott, A.B.*, et al.* (2013). The pattern of cortical dysfunction in a mouse model of a schizophrenia-related microdeletion. J Neurosci *33*, 14825-14839.
Flint, J. (2016). Rare genetic variants and schizophrenia. Nat Neurosci *19*, 525-527.
Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M.*, et al.* (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179-184.
Fullard, J.F., Hauberg, M.E., Bendl, J., Egervari, G., Cirnaru, M.D., Reach, S.M., Motl, J., Ehrlich, M.E., Hurd, Y.L., and Roussos, P. (2018). An atlas of chromatin accessibility in the adult human brain. Genome Res *28*, 1243-1252.
Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics *32*, 289-291.
Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landen, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F.*, et al.* (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci *19*, 1433-1441.
Goldman-Rakic, P.S. (1995). Architecture of the prefrontal cortex and the central executive. Ann N Y Acad Sci *769*, 71-83.
Gruter, T., Wiescholleck, V., Dubovyk, V., Aliane, V., and Manahan-Vaughan, D. (2015). Altered neuronal excitability underlies impaired hippocampal function in an animal model of psychosis. Front Behav Neurosci *9*, 117.
Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of, S., Group, P.S.*, et al.* (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518-529.
Hamm, J.P., Peterka, D.S., Gogos, J.A., and Yuste, R. (2017). Altered Cortical Ensembles in Mouse Models of Schizophrenia. Neuron *94*, 153-167 e158.
Harrington, A.J., Raissi, A., Rajkovich, K., Berto, S., Kumar, J., Molinaro, G., Raduazzo, J., Guo, Y., Loerwald, K., Konopka, G.*, et al.* (2016). MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. Elife *5*.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
Hitti, F.L., and Siegelbaum, S.A. (2014). The hippocampal CA2 region is essential for social memory. Nature *508*, 88-92.
Hoshii, T., Cifani, P., Feng, Z., Huang, C.H., Koche, R., Chen, C.W., Delaney, C.D., Lowe, S.W., Kentsis, A., and Armstrong, S.A. (2018). A Non-catalytic Function of SETD1A Regulates Cyclin K and the DNA Damage Response. Cell *172*, 1007-1021 e1017.
Hosseini, A., and Minucci, S. (2017). A comprehensive review of lysine-specific demethylase 1 and its roles in cancer. Epigenomics *9*, 1123-1142.
Kogan, J.H., Frankland, P.W., and Silva, A.J. (2000). Long-term memory underlying hippocampus-dependent social recognition in mice. Hippocampus *10*, 47-56.
Koike, H., Arguello, P.A., Kvajo, M., Karayiorgou, M., and Gogos, J.A. (2006). Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice. Proc Natl Acad Sci U S A *103*, 3693-3697.
Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P.*, et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813-1831.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.
Lao, Z., Raju, G.P., Bai, C.B., and Joyner, A.L. (2012). MASTR: a technique for mosaic mutant analysis with spatial and temporal control of recombination using conditional floxed alleles in mice. Cell Rep *2*, 386-396.
Lee, P.H., O'Dushlaine, C., Thomas, B., and Purcell, S.M. (2012). INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics *28*, 1797-1799.
Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. 1752-1779.

Li, Y., Schulz, V.P., Deng, C., Li, G., Shen, Y., Tusi, B.K., Ma, G., Stees, J., Qiu, Y., Steiner, L.A.*, et al.* (2016). Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. Nucleic Acids Res *44*, 7173-7188.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
Madison, D.V., and Nicoll, R.A. (1984). Control of the repetitive discharge of rat CA 1 pyramidal neurones in vitro. J Physiol *354*, 319-331.
Maes, T., Carceller, E., Salas, J., Ortega, A., and Buesa, C. (2015). Advances in the development of histone lysine demethylase inhibitors. Curr Opin Pharmacol *23*, 52-60.
Maes, T., Mascaro, C., Tirapu, I., Estiarte, A., Ciceri, F., Lunardi, S., Guibourt, N., Perdones, A., Lufino, M.M.P., Somervaille, T.C.P.*, et al.* (2018). ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment of Acute Leukemia. Cancer Cell *33*, 495-511 e412.
Markenscoff-Papadimitriou, E., Allen, W.E., Colquitt, B.M., Goh, T., Murphy, K.K., Monahan, K., Mosley, C.P., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. Cell *159*, 543-557.
McGinnis, S., and Madden, T.L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res *32*, W20-25.
Mei, Y., Monteiro, P., Zhou, Y., Kim, J.A., Gao, X., Fu, Z., and Feng, G. (2016). Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. Nature *530*, 481-484.
Mi, Y., Katkov, M., and Tsodyks, M. (2017). Synaptic Correlates of Working Memory Capacity. Neuron *93*, 323-330.
Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci *16*, 5154-5167.
Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A *98*, 12902-12907.
Mitchell, A.C., Javidfar, B., Pothula, V., Ibi, D., Shen, E.Y., Peter, C.J., Bicks, L.K., Fehr, T., Jiang, Y., Brennand, K.J.*, et al.* (2017). MEF2C transcription factor is associated with the genetic and epigenetic risk architecture of schizophrenia and improves cognition in mice. Mol Psychiatry.
Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E.S., and Lomvardas, S. (2017). Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. Elife *6*.
Mukai, J., Dhilla, A., Drew, L.J., Stark, K.L., Cao, L., MacDermott, A.B., Karayiorgou, M., and Gogos, J.A. (2008). Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. Nat Neurosci *11*, 1302-1310.

Mukai, J., Tamura, M., Fenelon, K., Rosen, A.M., Spellman, T.J., Kang, R., MacDermott, A.B., Karayiorgou, M., Gordon, J.A., and Gogos, J.A. (2015). Molecular substrates of altered axonal growth and brain connectivity in a mouse model of schizophrenia. Neuron *86*, 680-695.
Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V.*, et al.* (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242-245.
Network, and Pathway Analysis Subgroup of Psychiatric Genomics, C. (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci *18*, 199-209.
O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D.*, et al.* (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246-250.
Ow, J.R., Palanichamy Kala, M., Rao, V.K., Choi, M.H., Bharathy, N., and Taneja, R. (2016). G9a inhibits MEF2C activity to control sarcomere assembly. Sci Rep *6*, 34163.
Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet *9*, e1003709.
Plummer, N.W., Evsyukova, I.Y., Robertson, S.D., de Marchena, J., Tucker, C.J., and Jensen, P. (2015). Expanding the power of recombinase-based labeling to uncover cellular diversity. Development *142*, 4385-4393.
Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336-2337.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.
Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
RK, C.Y., Merico, D., Bookman, M., J, L.H., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z.*, et al.* (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci *20*, 602-611.
Rodriguez-Murillo, L., Gogos, J.A., and Karayiorgou, M. (2012). The genetic architecture of schizophrenia: new mutations and emerging paradigms. Annu Rev Med *63*, 63-80.

Rotaru, D.C., van Woerden, G.M., Wallaard, I., and Elgersma, Y. (2018). Adult Ube3a Gene Reinstatement Restores the Electrophysiological Deficits of Prefrontal Cortex Layer 5 Neurons in a Mouse Model of Angelman Syndrome. J Neurosci *38*, 8011-8030.
Rudenko, A., and Tsai, L.H. (2014). Epigenetic modifications in the nervous system and their impact upon cognitive impairments. Neuropharmacology *80*, 70-82.
Saito, T. (2006). In vivo electroporation in the embryonic mouse central nervous system. Nat Protoc *1*, 1552-1558.
Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.
Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.
Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem *81*, 65-95.
Shohat, S., Ben-David, E., and Shifman, S. (2017). Varying Intolerance of Gene Pathways to Mutational Classes Explain Genetic Convergence across Neuropsychiatric Disorders. Cell Rep *18*, 2217-2227.
Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.
Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G.*, et al.* (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci *19*, 571-577.
Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B.*, et al.* (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744-747.
Takata, A., Ionita-Laza, I., Gogos, J.A., Xu, B., and Karayiorgou, M. (2016). De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. Neuron *89*, 940-947.
Takata, A., Xu, B., Ionita-Laza, I., Roos, J.L., Gogos, J.A., and Karayiorgou, M. (2014). Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. Neuron *82*, 773-780.
Tamura, M., Mukai, J., Gordon, J.A., and Gogos, J.A. (2016). Developmental Inhibition of Gsk3 Rescues Behavioral and Neurophysiological Deficits in a Mouse Model of Schizophrenia Predisposition. Neuron *89*, 1100-1109.

Tillotson, R., Selfridge, J., Koerner, M.V., Gadalla, K.K.E., Guy, J., De Sousa, D., Hector, R.D., Cobb, S.R., and Bird, A. (2017). Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. Nature *550*, 398-401.
Titley, H.K., Brunel, N., and Hansel, C. (2017). Toward a Neurocentric View of Learning. Neuron *95*, 19-32.
Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., HA, F.S., Doebley, A.L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). denovo-db: a compendium of human de novo variants. Nucleic Acids Res *45*, D804-D811.
Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L.*, et al.* (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res *42*, D1001-1006.
Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet *44*, 1365-1369.
Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet *43*, 864-868.
Zucker, R.S., and Regehr, W.G. (2002). Short-term synaptic plasticity. Annu Rev Physiol *64*, 355-405.

Acknowledgements. We thank Naoko Haremaki for the maintenance of the mouse colony and technical assistance. This work was partially supported by grants R01MH080234 (to J.A.G) and R01DA036894 (to S.L). Author Contributions: J.M., E.C., B.X., A.T., S.L., and J.A.G designed and interpreted the experiments. J.M validated the mouse model, performed, analyzed and interpreted behavioral, morphological, co-IP, targeted expression and all rescue assays. E.C. performed, analyzed and interpreted all ChIP-Seq and scRNA-Seq assays. Z.S. performed, analyzed and interpreted electrophysiology experiments. A.D performed, analyzed and interpreted behavior

experiments. P.T contributed to the *in vivo* axonal branching assays. C-Y.C performed and analyzed targeted expression assays. B.X performed the bioinformatics analysis of the overlap between ChIP-Seq peaks and risk genes. A.T analyzed the RNA-Seq data and contributed to the validation of the mouse model and the bioinformatics analysis of the ChIP-Seq data. Y.C contributed the bioinformatics analysis. S.L. and J.A.G supervised the work. J.M., E.C., Z.S., A.D, A.T., B.X., S.L., J.A.G wrote the paper. The authors declare no competing financial interests.

Figures and Figure Legends

Figure 1

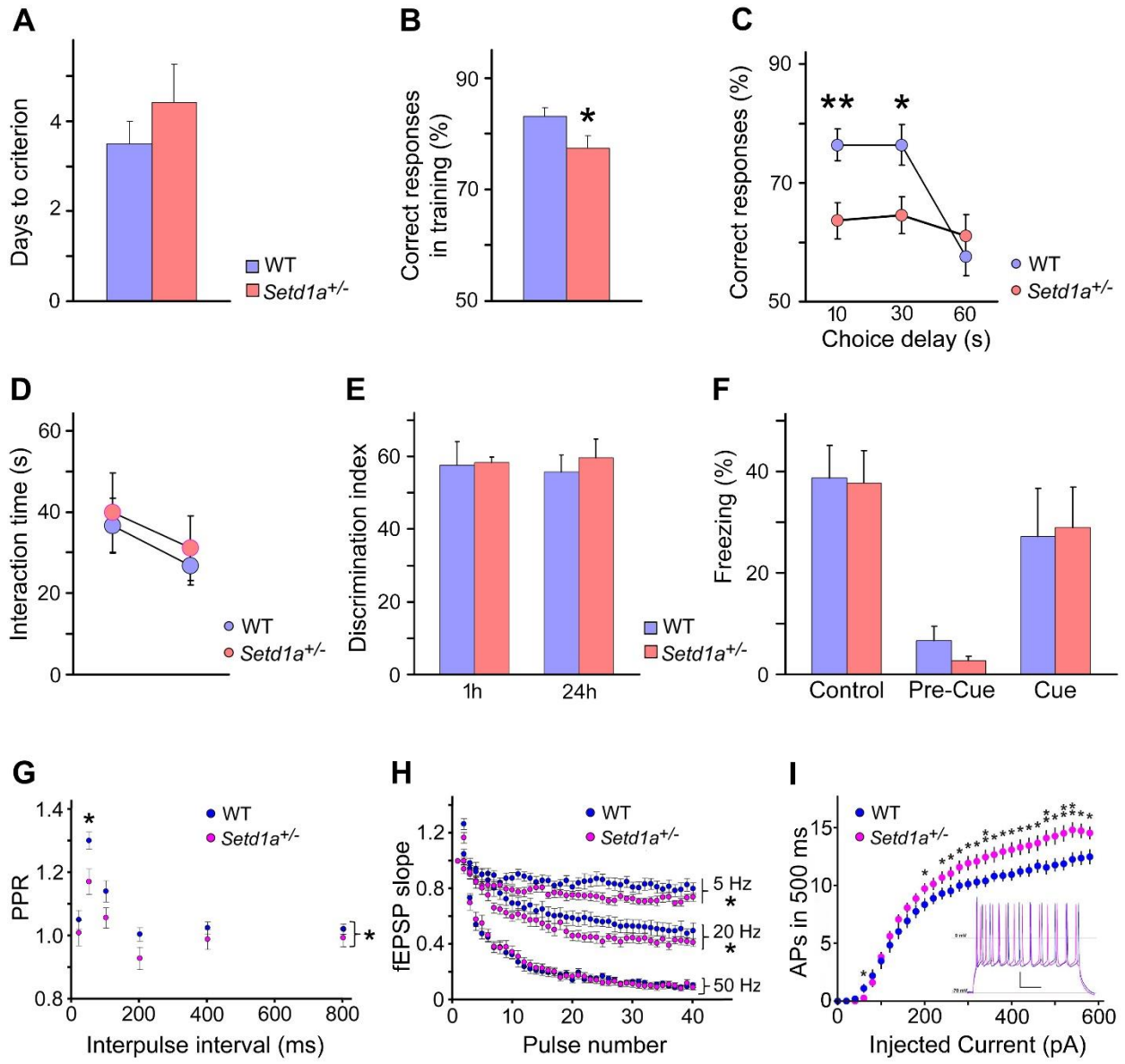

Figure 1. *Setd1a* deficiency impairs specifically WM and short-term plasticity in PFC.

(A) Days taken to reach criterion performance (three consecutive days of > 70% choice accuracy) on the delayed non-match to sample T-maze task. $P = 0.36$. N = 12 each genotype.

(B) Correct responses during training on the delayed non-match to sample T-maze task . $P = 0.041$. N = 12 each genotype.

(C) Working memory performance in a delayed non-match to place task. Mean percentage of correct responses for WT and *Setd1a*$^{+/-}$ mice (N = 12 each genotype). Note that vertical axis starts at 50% correct responses, which represents baseline response accuracy expected by chance. $P = 0.0038$ (10 sec), 0.0172 (30 sec), 0.485 (60 sec).

(D) Intact social memory in *Setd1a* deficient mice. *Setd1a*$^{+/-}$ and WT mice engaged in social interaction with the stimulus juvenile for a similar length of time (N = 8 each genotype, $P > 0.05$).

(E) Intact short and long-term Novel Object Recognition memory in *Setd1a* $^{+/-}$ mice. During an 1 h retention test *Setd1a*$^{+/-}$ and WT mice displayed enhanced preference for the novel object to a similar degree (1 h: N = 8 each genotype, $P > 0.05$). Intact long-term Novel Object Recognition memory in *Setd1a*$^{+/-}$ mice. During a 24 h retention test for assessment of long-term recognition memory, *Setd1a*$^{+/-}$ and WT mice displayed comparable preference for the novel object (24 h: N = 8 each genotype, $P > 0.05$).

(F) *Setd1a* deficiency does not affect fear-related associative memory. *Setd1a*$^{+/-}$ and WT mice spent similar percentage of time freezing upon re-exposure to the context in which fear

conditioning took place 24 h later. Presentation of the conditioned stimulus along 2 h after re-exposure to the conditioning context led to similar freezing times in *Setd1a*$^{+/-}$ and WT mice (N = 8 each genotype). N, number of independent animals. Data are shown as means ± s.e.m. *P < 0.05, **P < 0.01. Student's two-tailed t-test (A, B, E, F) and Two-way ANOVA post hoc pairwise comparisons with Bonferroni correction (C, D).

(G and H) *Setd1a* deficiency impairs short-term plasticity in PFC.

(G) PPR is significantly decreased in *Setd1a*$^{+/-}$ mice (N = 7, n = 23) at all tested ISIs (20, 50, 100, 200, 400 and 800 ms), compared to WT mice (N = 7, n = 23) (2-way repeated-measures ANOVA, P < 0.05; Bonferroni posttests revealed a significant difference at ISI of 50 ms, P < 0.05).

(H) Frequency-dependent differences of STD of the fEPSPs in WT and *Setd1a*$^{+/-}$ mice at 5 Hz, 20 Hz and 50 Hz. At 5 and 20 Hz, the STD is significantly greater in *Setd1a*$^{+/-}$ mice (N = 7, n = 22 at 5 Hz and N = 7, n = 19 at 20 Hz) compared to their WT controls (N = 7, n = 23 at 5 Hz and N = 7, n = 20 at 20 Hz; 2-way repeated-measures ANOVA, P < 0.05 for time × genotype interaction). At 50 Hz, the STD is similar between genotypes across experiments (WT mice: N = 7, n = 18 and *Setd1a*$^{+/-}$ mice: N = 7, n = 16; 2-way repeated-measures ANOVA, P > 0.05). N, number of independent animals. n, number of brain slices. Data are shown as means ± s.e.m. *P < 0.05, **P < 0.01.

(I) Summary data (mean ± s.e.m) of number APs evoked in response to 500 ms currents steps. WT recordings, blue; *Setd1a*$^{+/-}$ recordings, red. (WT, n = 22, *Setd1a*$^{+/-}$, n = 20; 2-way RM ANOVA, P = 0.021). Values represent mean ± s.e.m . (Starred data represent results of pair-wise t-test comparisons at given current steps: * $P < 0.05$, ** $P < 0.01$). *Inset; Setd1a*$^{+/-}$ neurons display enhanced excitability (red trace) compared to WT neurons (blue trace). Representative traces from current-clamp recordings from layer 2/3 pyramidal neurons showing AP responses to near-maximal current step (500 pA, 500 ms). Scale bar: 20 mV, 100 ms. Reference levels, V_{rest} = -70 mV, and 0 mV.

Figure 2

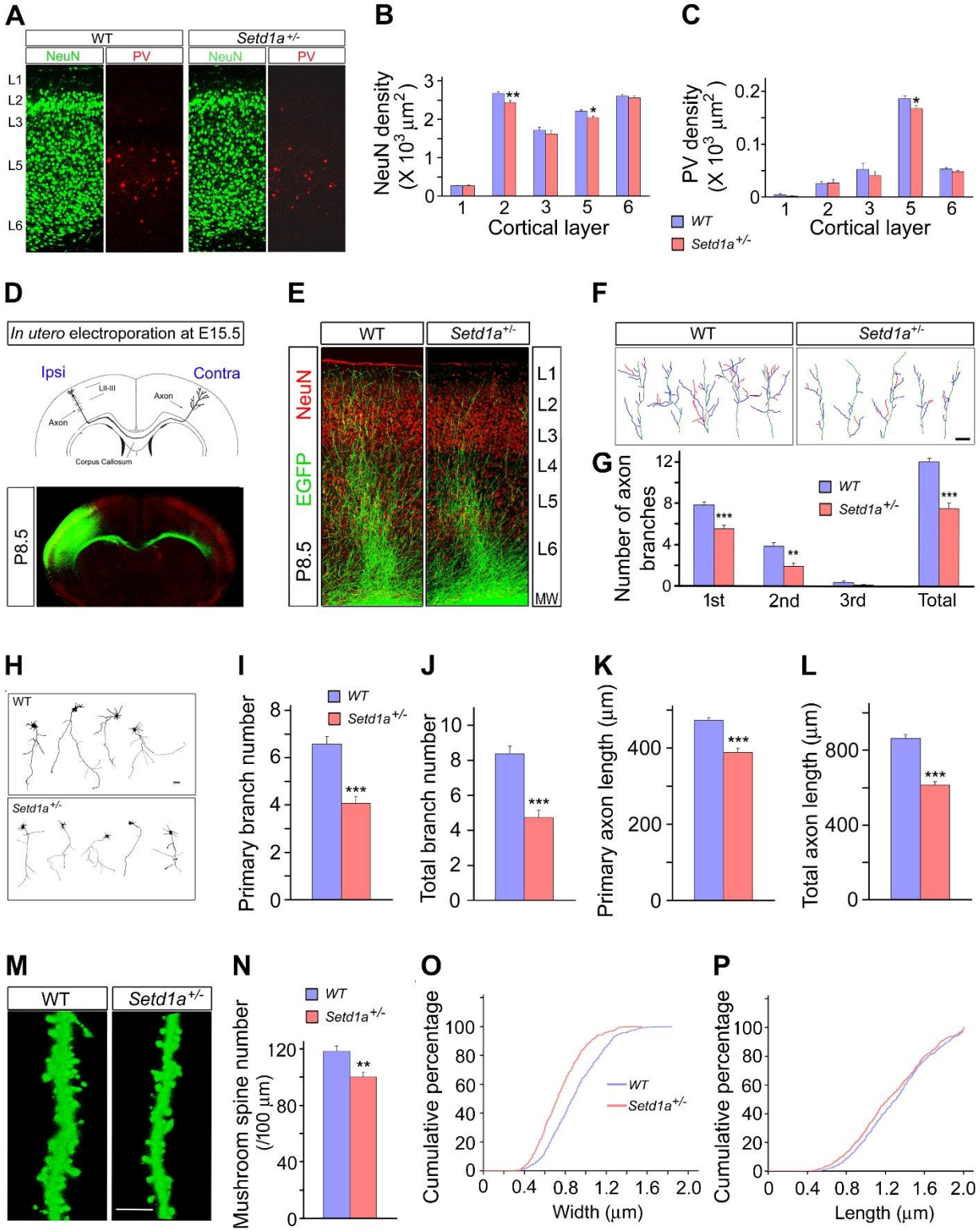
A
WT
Setd1a+/-
NeuN
PV
L1
L2
L3
L5
L6
B
NeuN density (X 10^3 μm^2)
Cortical layer
C
PV density (X 10^3 μm^2)
Cortical layer
WT
Setd1a+/-
D
In utero electroporation at E15.5
Ipsi
Contra
LII-III
Axon
Corpus Callosum
P8.5
E
WT
Setd1a+/-
NeuN
EGFP
P8.5
L1
L2
L3
L4
L5
L6
MW
F
WT
Setd1a+/-
G
Number of axon branches
WT
Setd1a+/-
1st
2nd
3rd
Total
H
WT
Setd1a+/-
I
Primary branch number
WT
Setd1a+/-
J
Total branch number
K
Primary axon length (μm)
L
Total axon length (μm)
M
WT
Setd1a+/-
N
Mushroom spine number (/100 μm)
WT
Setd1a+/-
O
Cumulative percentage
Width (μm)
WT
Setd1a+/-
P
Cumulative percentage
Length (μm)

Figure 2. *Setd1a* deficiency leads to cytoarchitectural alterations and disrupts axonal growth.

(A-C) Cytoarchitectural changes in the cortex of *Setd1a*$^{+/-}$ mice.

(A) Images of coronal sections of the prelimbic area in medial PFC of 8-week-old WT and *Setd1a*$^{+/-}$ mice showing neuronal cells labeled with NeuN (*green*) and interneurons labeled with PV antibody (*red*). Representative images (A) are composites of more than one image acquired from one brain section under identical scanning parameters.

(B) Histograms indicating a modestly diminished NeuN cell density in L2 (9.0% reduction, $p < 0.01$, N = 6, n = 12 per genotype) as well as in L5 (7.5% reduction, $P < 0.02$, N = 6, n = 12 per genotype) of the PFC (A and B).

(C) PV^{+} cell density in *Setd1a*$^{+/-}$ mice was modestly reduced in L5 (11.5% reduction, $P < 0.05$; N = 5, n =15 per genotype] (A and C).

(D-G) *Setd1a* deficiency disrupts axonal terminal branching *in vivo*.

(D) Illustration of the callosal axon pathway of neurons residing at L2/3, electroporated with EGFP at E15.5 and assayed at P8.5 (*upper panel*). Lower panel shows EGFP-labeled neurons (*green*) and NeuN (*red*). Representative images (*lower panel*) are composites of more than one image acquired from one brain section under identical scanning parameters.

(E) Representative images of contralateral axon terminal branching of EGFP-labeled neurons from coronal sections of P8.5 brains of mice *in utero* electroporated at E15.5. Images are composites of more than one image acquired from one brain section under identical scanning parameters.

(F) Tracings of representative neurons in *in utero* electroporated brains expressing EGFP at P8.5, from similar sections as in (E). Depicted are 2D projections of the axon terminal in the contralateral cortical L1–4. Primary (*blue*), secondary (*red*), and tertiary (*pink*) branches from axons (*green*) are highlighted. Scale bar, 20 μm.

(G) Quantitative assessment of contralateral axon branching in L1–4 reveals a reduction of branch number in *Setd1a*$^{+/-}$ mice (N = 10, n = 30 each genotype).

(H-L) *Setd1a* deficiency disrupts axonal growth *in vitro*.

(H) Representative images of axonal deficits in *Setd1a*$^{+/-}$ cortical neurons expressing EGFP at DIV5. *Setd1a*$^{+/-}$ cortical neurons were transfected with plasmids expressing EGFP at DIV4 and fixed at DIV5. Scale bar, 50 μm.

(I) Quantification of primary branchpoint number. *Setd1a* deficiency caused a decrease in the number of primary branch points in *Setd1a*$^{+/-}$ neurons (4.1 ± 0.28, N = 4, n = 315, $P < 0.001$) compared to WT neurons (6.6 ± 0.31, N = 4, n = 304).

(J) Quantification of total branchpoint number. Total branch points were reduced in *Setd1a*$^{+/-}$ (4.7 ± 0.68, N = 4, n = 315, $P < 0.0009$) compared to WT neurons (total branchpoints: 8.4 ± 0.76, N = 4, n = 304).

(K) Quantification of primary axon length. Primary axon length was reduced at DIV5 in *Setd1a*$^{+/-}$ (388.6 ± 11.8 μm, N = 4, n = 315, $P < 0.0008$) compared to WT neurons (472.9 ± 6.9 μm, N = 4, n = 304).

(L) Quantification of total axon length. Total axon length were reduced in *Setd1a*$^{+/-}$ (total axon length: 614.2 ± 22.6 μm, N = 4, n = 315, $P < 0.0003$) compared to WT neurons (total axon length: 869.6 ± 25.9 mm, N = 4, n = 304).

(M) Alterations in spines of basal dendrites of L5 pyramidal neurons at the prelimbic area of medial PFC in *Setd1a*$^{+/-}$ mice. Representative, high-magnification images of eGFP–expressing neurons from WT Thy1-GFP/M (*left panel*) and *Setd1a*$^{+/-}$; Thy1-GFP/M (*right panel*) mice.

(N) Density of mushroom spines estimated over 100 μm of dendritic length is reduced in *Setd1a*$^{+/-}$; Thy1-GFP/M (N = 10, n = 20 each genotype). The density of other subtypes of spine morphology is shown in Figure S2I. Scale bar, 5 μm.

(O and P) Reduction in the width (O; $P < 0.0001$, Kolmogorov-Smirnov test), and length (P; $P < 0.001$, Kolmogorov-Smirnov test) of mushroom spines of *Setd1a*$^{+/-}$; Thy1- GFP$^{+/–}$ mPFC pyramidal

neurons relative to WT Thy1-GFP$^{+/-}$ mPFC neurons. (O, P: WT, N = 10 mice, n = 644, *Setd1a*$^{+/-}$, N = 10, n = 605). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Student's two-tailed *t*-test (B, C, G, I-L, N). N: number of independent animals, n: number of neurons or brain slices. Histograms show means and error bars represent s.e.m.

Figure 3

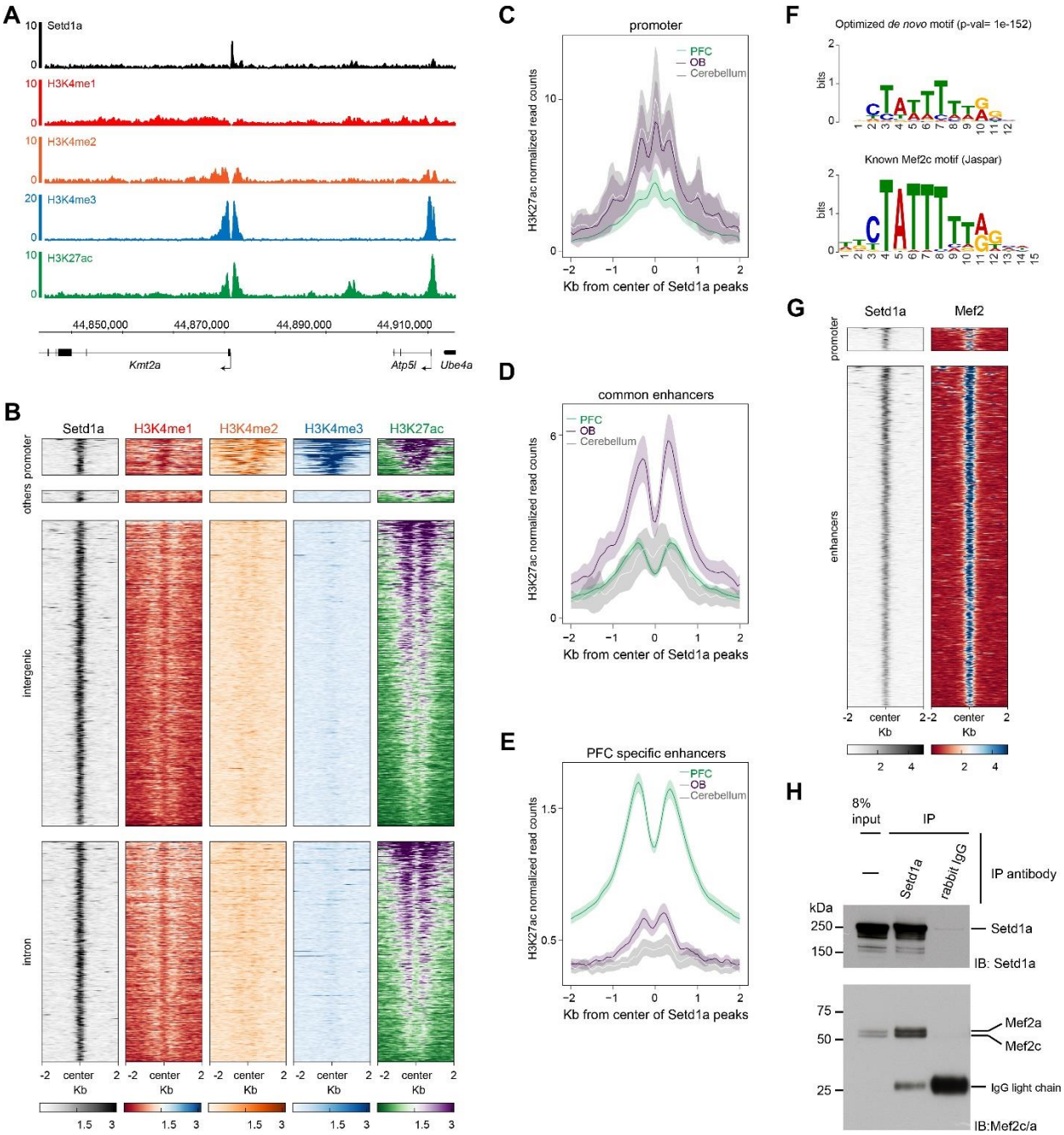

Figure 3. Setd1a binds promoter and enhancers of the PFC in a complex with Mef2.

(A) Representative ChIP-Seq locus showing Setd1a binding at the promoter of *Kmt2a* gene. Normalized reads per genomic coverage from PFC ChIP-Seq for Setd1a (black) and chromatin

marks (H3K4me1, red; H3K4me2, orange; H3K4me3, blue; H3K27ac, green) are shown.

(B) Heatmap of listed ChIP-Seq coverage from PFC, 2 kilobases upstream and downstream from the center of Setd1a peaks. Setd1a peaks were clustered in promoters (top, defined by the overlap with H3K4me3 peaks) or enhancers (defined by absence of H3K4me3 signal and presence of H3K27ac and H3K4me1 signal). Enhancers were sub-clustered by the annotation of Setd1a peaks. top-middle, peaks in annotated 5′UTRs, coding sequence, transcription end site and 3′UTRs (defined as “others”); middle-bottom, peaks in intergenic regions; bottom, peaks in annotated introns. All heatmaps are sorted in the same order, based upon Setd1a and H3K27ac signal.

(C-E) Comparison of the mean normalized coverage H3K27ac per bin (50 bps) 2 Kb around Setd1a peaks in PFC (green), OB (purple) and Cerebellum (gray). Distribution of the mean normalized coverage for promoter (C), common enhancers (D) and PFC specific enhancers (E) of H3K27ac from different tissues is shown. Shading indicates 95% confidence intervals.

(F**)** *de novo* transcription factor motif analysis showed enrichment for Mef2 (top, P = 1e-152) compared to known Mef2c motif (bottom, Jaspar Core Vertebrata database).

(G**)** Heatmap of Setd1a and Mef2 coverage 2 kb around the center of Setd1a peaks in promoter (top) and enhancers (middle). Coverage is shown by color according to the scale (bottom).

(H**)** Co-immunoprecipitation (Co-IP) of Setd1a and Mef2a/c from the cortical plate of 6 weeks old

mice. Western blot of 8% input (left lane), immunoprecipitation (IP) with Setd1a antibody (middle lane) or IgG (right lane) is shown. Detection with Mef2a/c, Setd1a and IgG antibody is shown in methods.

Figure 4

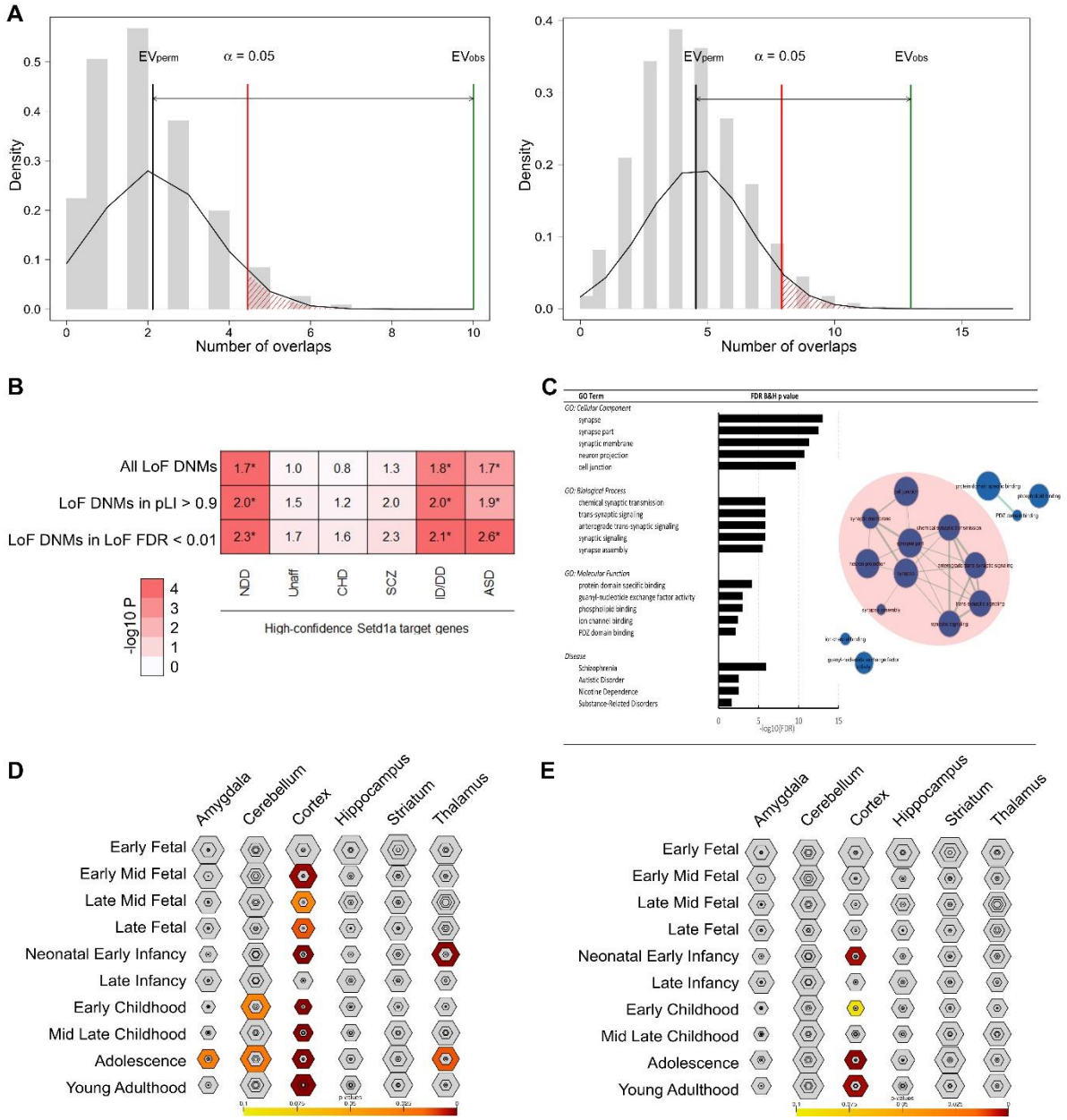

Figure 4. Setd1a targets are enriched for synaptic genes with postnatally-biased expression and overlap with genome-wide significant signals of SCZ and NDD.

(A) Permutation tests (n=20,000, randomization method= circularRandomizeRegions) indicate there is significant overlaps between conserved Setd1a ChIP-Seq peaks (left panel: megablast, *P* = 0.0001, Z-score = 5.547; right panel: blastn , *P* = 0.0004, Z-score = 4.137) and the SCZ-GWAS loci identified by reference (Schizophrenia Working Group of the Psychiatric Genomics, 2014). X-axis is the number of overlaps, Y-axis is the density of expected number of overlaps determined by permutation. EV_{perm}, Expected average number of overlaps by permutation, EV_{obs}, actual observed number of overlaps. Permutation *P* value threshold = 0.05.

(B) DNENRICH analysis of the high-confidence Setd1a target genes. Fold changes (observed/expected number of LoF DNMs in the gene-set of interest) are shown in the cells color-coded by p values. Unaff: Unaffected subjects, CHD: congenital heart diseases, SCZ: schizophrenia, ID/DD: intellectual disability/developmental disorders, ASD: autism spectrum disorders, NDD: neurodevelopmental disorders (SCZ+ID/DD+ASD) [FC; *P* (pLI/LoF-FDR): ID/DD: 2.00/2.14; 2.0×10^{-4} /3.2×10^{-5}, ASD: 1.94/2.61; 0.0053/8.0×10^{-6} and SCZ: 2.05/2.32; ns/ns, total N of LoF DNMs in ID/DD, ASD and SCZ = 986, 793 and 95, respectively].

(C) GO-term analysis of target genes with intronic Setd1a ChIP-Seq peaks and functional clusters formed by these enriched GO terms.

(D) Spatiotemporal expression pattern of target genes with intronic Setd1a Chip-seq in human brain. *P* values are PSI *P* values as defined in the CSEA tool.

(E) Spatiotemporal expression pattern of target genes with intronic Setd1a Chip-seq that overlapped with the mutation intolerant genes hit by LoF DNMs in NDD defined in (C).

Figure 5

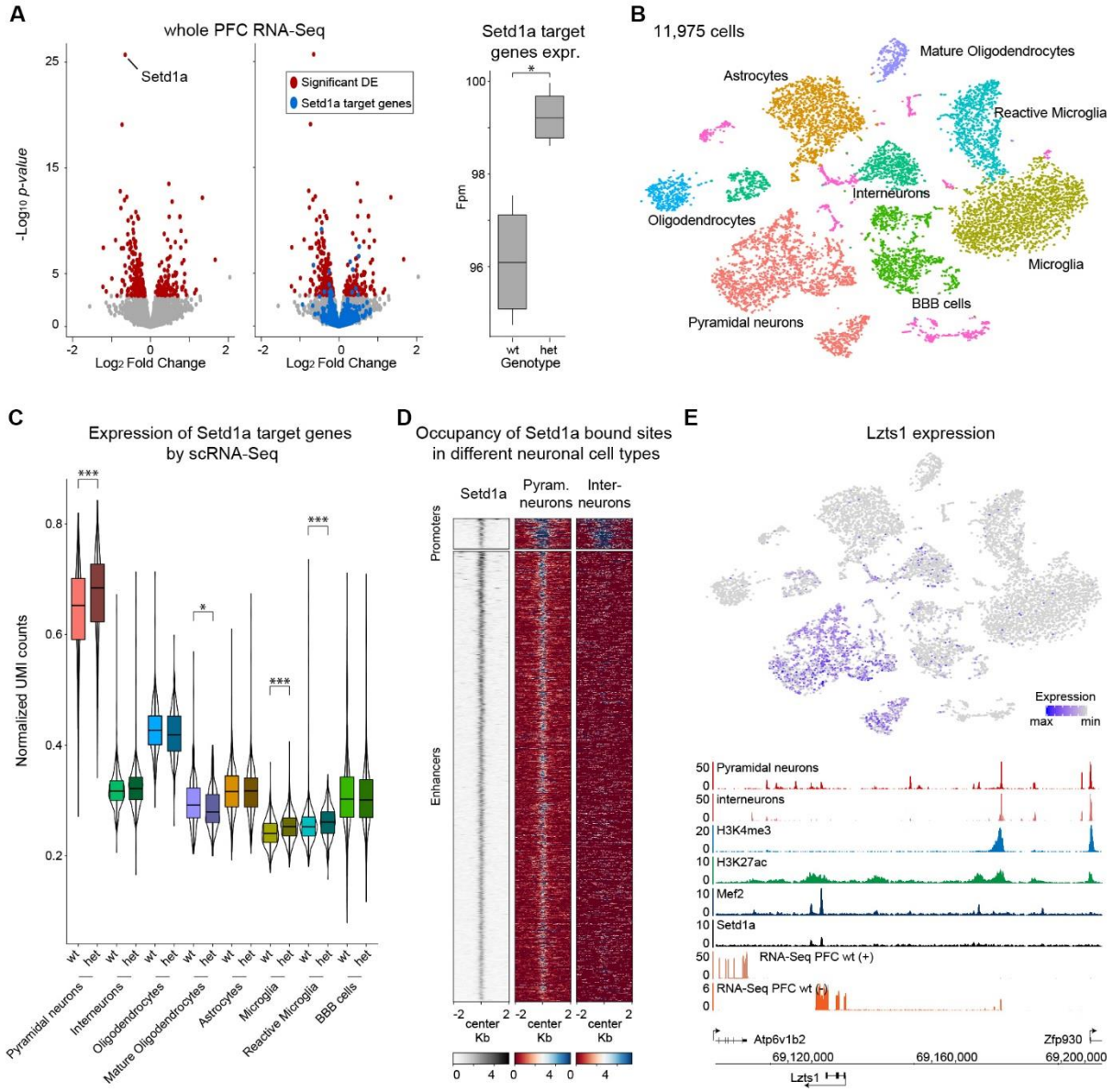

Figure 5. Characterization of genes differentially expressed (DE) in PFC of *Setd1a*$^{+/-}$ mice and cell-type specific expression of Setd1a target genes.

(A) Left and middle panel: Volcano plot of differential expressed genes in whole-PFC-RNA-Seq of *Setd1a*$^{+/-}$ mice. Red; 342 genes that showed differential expression (FDR<5%) (*left and middle panel*). Blue (*middle panel*); Setd1a target genes. Only 28 Setd1a target genes (16 up-regulated

and 12 down-regulated) showed significant changes in expression when we compare WT vs. *Setd1a*$^{+/-}$ mice. Right panel: Mean expression of Setd1a target genes [4 biological replicates per genotype, i.e. 4 WT (wt) and 4 *Setd1a*$^{+/-}$ (het) mice] that by DESeq2 have p-value < 0.05 and FDR <0.43, thus not significantly differentially expressed. Wilcox test, $*P < 0.05$.

(B) tSNE (t-distributed stochastic neighbor embedding) representation of 11,975 single cell transcriptomes from the PFC of WT (5,944) and *Setd1a*$^{+/-}$ (6,031) mice colored by clusters.

(C) Mean expression of Setd1a target genes per cell per cluster from the PFC of WT (wt) and *Setd1a*$^{+/-}$ (het) mice. Only genes with mean expression of 0.1 normalized UMI counts across all clusters were considered. Wilcox test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

(D) Heatmap of Setd1a ChIP-Seq coverage and ATAC-Seq of sorted pyramidal neurons and interneurons, 2 kilobases upstream and downstream from the center of Setd1a bound enhancers and promoters. All heatmaps are sorted in the same order, based upon *Setd1a* signal.

(E) Top. Expression of Lzts1 in PFC neuronal and non-neurons cells. Lzts1 expression is specifically restricted to pyramidal neurons. Bottom. Lzts1 gene locus. Representative RNA-Seq of one biological replicate of WT mice (*orange*), ChIP-Seq of Mef2 (*dark blue*), Setd1a (*black*), H3K27ac (*green*), H3K4me3 (*light blue*) and ATAC-Seq from pyramidal neurons (*dark red*) and interneurons (*light red*) are shown.

Figure 6

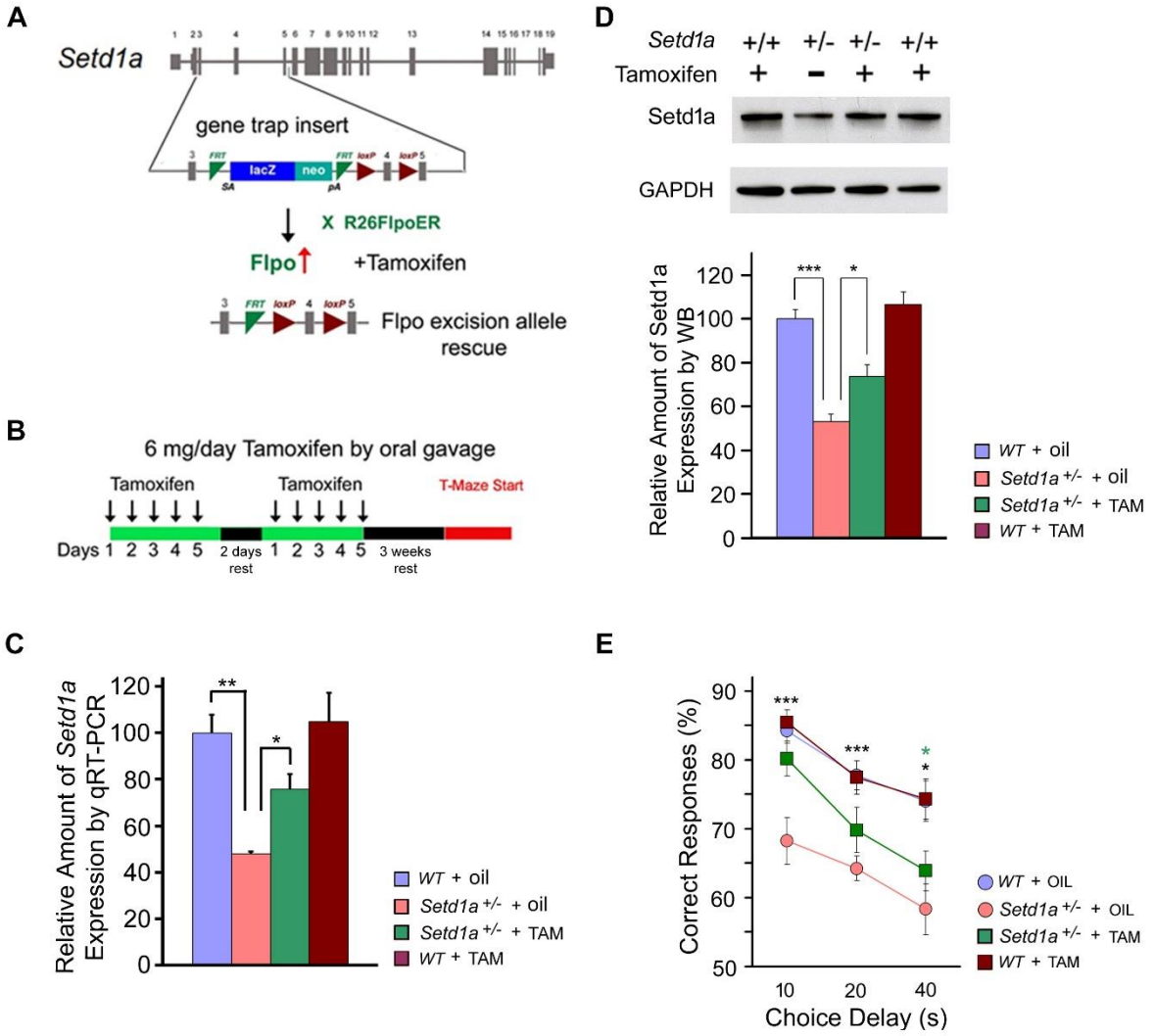

Figure 6. Restoration of Setd1a expression in adulthood rescues WM performance.

(A) Schematic of the restoration of knockout-first allele after tamoxifen treatment. For the rescue experiments, *Setd1a*$^{+/-}$ mice were crossed to *R26FlpoER* mice. FLP (Flpo) induced globally by tamoxifen converts the “knockout-first” allele to a conditional allele, restoring gene activity.

(B) Tamoxifen feeding scheme. Animals were fed tamoxifen for 5 consecutive days followed by 2 days of rest, then fed for 5 more days followed by 3 weeks of rest before the T-maze assay.

(C) qRT-PCR of *Setd1a* expression in PFC of adult mice, demonstrating the restoration of *Setd1a* mRNA levels following tamoxifen treatment. N = 6 mice each genotype, n = 3 each animal. Data are shown as means ± s.e.m. *$P < 0.05$; **$P < 0.01$. Student's two-tailed *t*-test.

(D) Western blot analysis of Setd1a levels in lysates from PFC of adult mice, demonstrating the restoration of Setd1a protein levels by tamoxifen treatment. N = 4 mice each genotype, n = 3 each animal. Data are shown as means ± s.e.m. *$P < 0.05$; **$P < 0.01$, ***$P < 0.001$. Student's two-tailed *t*-test.

(E) WM performance during testing phase. Mean percentage of correct responses for WT and *Setd1a*$^{+/}$ mice with/without tamoxifen treatment. **P*, 0.05, ***P*, 0.01 (black asterisk: WT + oil versus *Setd1a*$^{+/-}$ + oil, green asterisk: *Setd1a*$^{+/-}$ + oil versus *Setd1a*$^{+/-}$ + TAM).

Figure 7

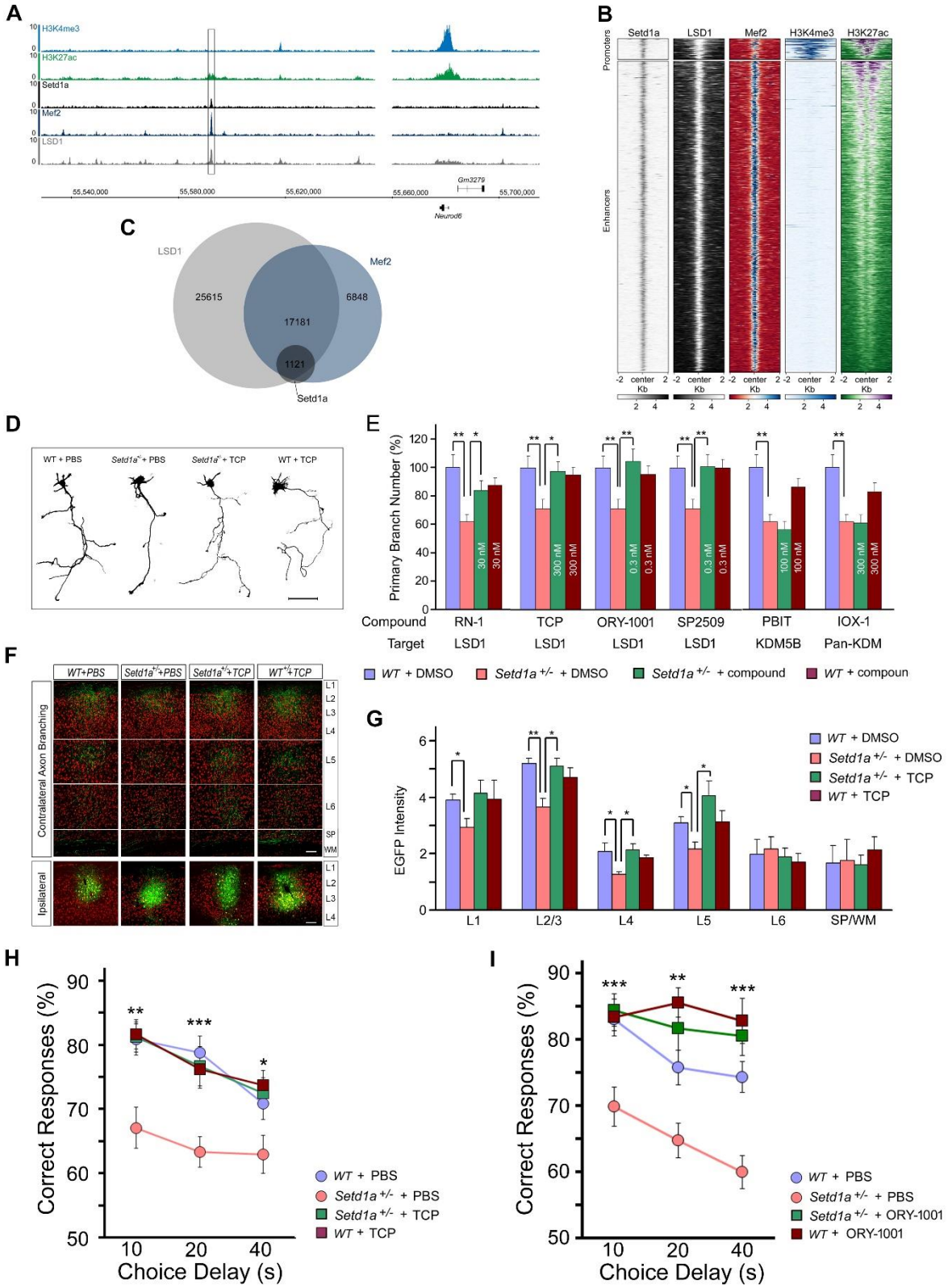
A
H3K4me3
H3K27ac
Setd1a
Mef2
LSD1
55,540,000
55,580,000
55,620,000
55,660,000
55,700,000
Gm3279
Neurod6
B
Setd1a
LSD1
Mef2
H3K4me3
H3K27ac
Promoters
Enhancers
-2 center 2
Kb
C
LSD1
Mef2
25615
6848
17181
1121
Setd1a
D
WT + PBS
Setd1a+/- + PBS
Setd1a+/- + TCP
WT + TCP
E
Primary Branch Number (%)
Compound
Target
RN-1
LSD1
TCP
LSD1
ORY-1001
LSD1
SP2509
LSD1
PBIT
KDM5B
IOX-1
Pan-KDM
30 nM
300 nM
0.3 nM
100 nM
WT + DMSO
Setd1a +/- + DMSO
Setd1a +/- + compound
WT + compoun
F
WT+PBS
Setd1a+/-+PBS
Setd1a+/-+TCP
WT+/-+TCP
Contralateral Axon Branching
Ipsilateral
L1
L2
L3
L4
L5
L6
SP
WM
G
EGFP Intensity
WT + DMSO
Setd1a +/- + DMSO
Setd1a +/- + TCP
WT + TCP
L1
L2/3
L4
L5
L6
SP/WM
H
Correct Responses (%)
Choice Delay (s)
WT + PBS
Setd1a +/- + PBS
Setd1a +/- + TCP
WT + TCP
I
Correct Responses (%)
Choice Delay (s)
WT + PBS
Setd1a +/- + PBS
Setd1a +/- + ORY-1001
WT + ORY-1001

Figure 7. Inhibition of LSD1 activity in adulthood counteracts the effects of *Setd1a* deficiency.

(A) *Neurod6* locus. ChIP-Seq of LSD1 (*grey*), Mef2 (*dark blue*), Setd1a (*black*), H3K27ac (*green*), H3K4me3 (*light blue*) is shown. Overlap between LSD1-Sedt1a-Mef2 peaks is highlighted.

(B) Heatmap of the ChIP-Seq coverage for Setd1a, LSD1, Mef2, H3K4me3 and H3K27ac 2 kilobases upstream and downstream from the center of the 1178 Setd1a peaks (divided in promoter and enhancer element). All heatmaps are sorted in the same order, based upon Setd1a and H3K27ac signals.

(C) Venn diagram of the overlap of the ChIP-seq peaks for LSD1, Setd1a and Mef2.

(D and E) *Setd1a*$^{+/-}$ cortical neurons were treated with LSD1/KDM1a, or JmjC domain KDM inhibitors from DIV3 to DIV5, and transfected with plasmids expressing EGFP at DIV4 and fixed at DIV5. Tracings of representative neurons imaged at DIV5 following TCP or PBS treatment are depicted in (D). The most effective concentrations of each inhibitor for restoring primary axon branch number in *Setd1a*$^{+/-}$ neurons (0.1 nM – 30 μM) are represented in (E). **P*, 0.05, ***P*, 0.01. Histograms show means and error bars represent s.e.m. Scale bar, 50 mm.

(F and G) TCP treatment restores axon branch number *in vivo*.

(F) Representative images of contralateral axon-terminal branching of EGFP-labeled neurons from coronal sections of adult brains of mice injected AAV2Synapsine-EGFP virus at 8 weeks

(*upper panels*) and viral injection sites in ipsilateral L2/3 (*bottom panels*). Immuno-histological staining represents EGFP (*green*) and NeuN (*red*).

(G) Histograms depicting quantitative assessment of contralateral axon branching in L1 to L6 and white matter in adulthood and indicating that TCP treatment restores contralateral axon branching in L2 to L5 in *Setd1a*$^{+/-}$ mice (N = 4 each genotype, n = 3 each brain). **P*, 0.05, ***P*, 0.01. Histograms show means and error bars represent s.e.m.

(H) Complete rescue of WM performance in a delayed non-match to place task in *Setd1a*$^{+/-}$ mice treated with TCP. Mean percentage of correct responses for WT and *Setd1a*$^{+/-}$ mice with/without 3 mg/kg/day TCP or vehicle treatment are depicted. **P*, 0.05. ***P*, 0.01, ****P*, 0.001 for comparison between WT + PBS versus *Setd1a*$^{+/-}$ + PBS groups. No significant difference between WT + TCP and *Setd1a*$^{+/-}$ + TCP groups was observed.

(I) Rescue of WM performance in a delayed non-match to place task in *Setd1a*$^{+/-}$ mice treated with ORY-1001. Mean percentage of correct responses for WT and *Setd1a*$^{+/-}$ mice with/without 10 μg/kg/day ORY-1001 or vehicle treatment are depicted. **P*, 0.05, ***P*, 0.01, ****P*, 0.001 for comparison between WT + PBS versus *Setd1a*$^{+/-}$ + PBS groups.

Figure S1

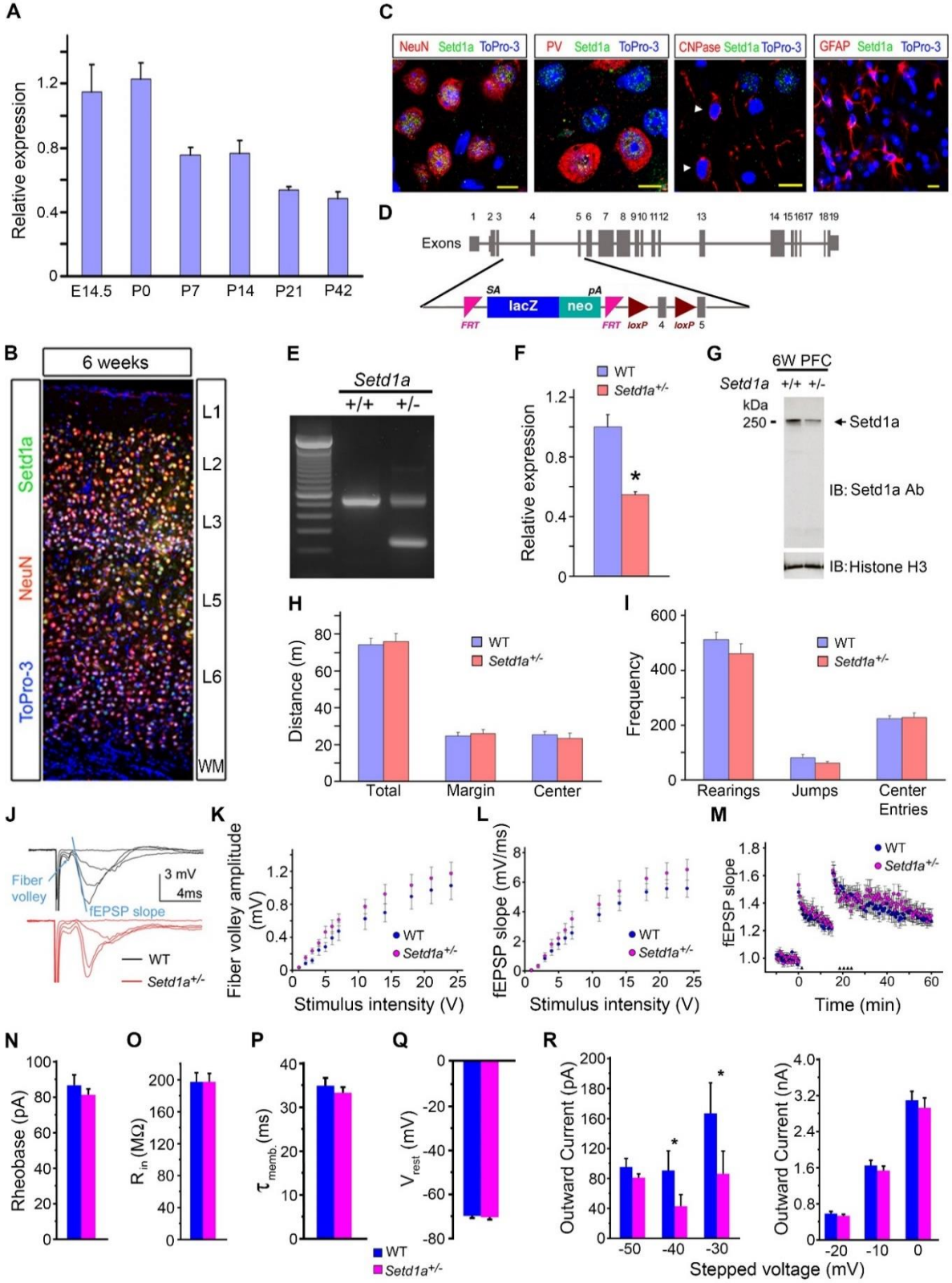
A
Relative expression
E14.5
P0
P7
P14
P21
P42
B
6 weeks
Setd1a
NeuN
ToPro-3
L1
L2
L3
L5
L6
WM
C
NeuN Setd1a ToPro-3
PV Setd1a ToPro-3
CNPase Setd1a ToPro-3
GFAP Setd1a ToPro-3
D
Exons
SA
lacZ
neo
pA
FRT
FRT
loxP
loxP
E
Setd1a
+/+
+/-
F
WT
Setd1a+/-
Relative expression
G
6W PFC
Setd1a +/+ +/-
kDa
250
Setd1a
IB: Setd1a Ab
IB: Histone H3
H
Distance (m)
Total
Margin
Center
I
Frequency
Rearings
Jumps
Center Entries
J
Fiber volley
3 mV
4ms
fEPSP slope
WT
Setd1a+/-
K
Fiber volley amplitude (mV)
Stimulus intensity (V)
L
fEPSP slope (mV/ms)
Stimulus intensity (V)
M
fEPSP slope
Time (min)
N
Rheobase (pA)
O
Rin (MΩ)
P
τmemb. (ms)
Q
Vrest (mV)
R
Outward Current (pA)
Outward Current (nA)
Stepped voltage (mV)

Figure S1. Setd1a expression pattern and molecular, behavioral and electrophysiological characterization of *Setd1a*$^{+/-}$ mice.

(A) Developmental expression of *Setd1a* by qRT-PCR in frontal cortex of WT mice (N = 5 each age). Histograms show means and error bars represent s.e.m.

(B) Representative image of Setd1a and NeuN-labeled cells in the prelimbic area from coronal sections of 6 weeks-old brains. WM: white matter.

(C) Higher magnification images from coronal sections of 6 weeks-old brains. Setd1a (*green*) exhibits mixtures of punctate and diffuse fibrillar staining in NeuN^{+} cells (*red* in *left panel*), and PV^{+} interneurons (*red* in *middle left panel*) in medial PFC superficial layers, but neither in CNPase^{+} oligodendrocytes [*red* (*white arrowheads*) in *middle right panel*) in medial PFC superficial layers nor in GFAP^{+} cells (*red* in *right panel*) in corpus callosum. Setd1a localizes to euchromatin (*left* and *middle left panels*), as neither signal overlaps with ToPro-3 (*blue*) bright heterochromatin regions. Scale bar: 10 μm.

(D) Schematic of the *Setd1a* gene and the targeted locus. The *Setd1a* conditional allele in the *Setd1a*$^{tm1a(EUCOMM)Wtsi}$ mouse line (referred to as *Setd1a*$^{+/-}$) is based on the 'knockout-first' design, a strategy that combines both a reporter-tagged and a conditional mutation. The 'KO-first' allele contains a *LacZ* trapping cassette and a floxed promoter-driven neo cassette inserted into intron 3 of the gene, disrupting gene function.

(E) PCR genotyping showing the expected bands for *Setd1a*$^{+/-}$.

(F) Expression of *Setd1a* by qRT-PCR in medial pre-frontal cortex (mPFC) of *Setd1a*$^{+/-}$ (N = 5) and WT (N = 5) littermate mice at 6-week-old. N, number of individual animals. Data are shown as means ± s.e.m. *P < 0.05; **P < 0.01. Student's two-tailed *t*-test.

(G) Western blot analysis of Setd1a in the frontal cortex of *Setd1a*$^{+/-}$ and WT littermate mice at 6-week-old.

(H and I) *Setd1a*$^{+/-}$ mice show normal locomotor activity and anxiety: No difference in locomotor activity as assessed by total distance covered between *Setd1a*$^{+/-}$ (N = 11) and WT (N = 10) mice during 60 min exposure to Open Field (H). No difference between genotypes was detected in margin distance vs center distance covered, indicators of anxiety (H). Also, no difference was detected in the number of rearings, jumps or center entries between *Setd1a*$^{+/-}$ (N = 11) and WT (N =10) mice (I). N, number of individual animals (H and I). Data are shown as means ± s.e.m. P > 0.05. Student's two-tailed *t*-test.

(J) Sample traces obtained in response to increasing stimulation intensities and showing the fiber volley amplitude (*blue arrow*) as well as the fEPSP initial slope (*blue line*).

(K) Normal afferent volley amplitude in *Setd1a*$^{+/-}$ mice (N = 7, n = 15), compared to WT mice (N = 7, n = 14) (2- way repeated-measures ANOVA, P > 0.05).

(L) Normal stimulus-response curve across experiments in *Setd1a*$^{+/-}$ mice (N = 7, n = 25), compared to WT mice (N = 7, n = 26) (2-way repeated-measures ANOVA, $P > 0.05$). WT mice (*blue circles*); *Setd1a* $^{+/-}$ mice (*pink circles*). N, number of individual animals, n, number of brain slices (J-L). Data are shown as means ± s.e.m.

(M) Comparable STP and LTP in *Setd1a*$^{+/-}$ mice overtime. (WT mice: N = 7, n = 12 and *Setd1a*$^{+/-}$ mice: N = 7, n = 11; 2-way repeated-measures ANOVA, $P > 0.05$). STP was induced at the end of the first 50 Hz train stimulation (1 upwards filled arrowhead), and LTP was induced after the stimulation of 4 consecutive 50 Hz trains (4 upwards filled arrowheads). N, number of independent animals. n, number of brain slices. Data are shown as means ± s.e.m. *$P < 0.05$, **$P < 0.01$.

(N) Minimal current necessary to elicit an action potential (rheobase) does not differ between genotypes. Values represent mean ± s.e.m.. (WT, *Setd1a*$^{+/-}$: n = 22, n = 20; unpaired t-test, $P = 0.87$)

(O) Neuronal input resistances (R_{in}) were not different between genotypes. Values represent mean ± s.e.m. (WT, *Setd1a*$^{+/-}$: n = 22, n = 20; unpaired t-test, $P = 0.72$)

(P) The neuronal membrane time constant did not differ between genotypes. Values represent mean ± s.e.m. (WT, *Setd1a*$^{+/-}$: n = 22, n = 20; unpaired t-test, $P = 0.41$)

(Q) Resting membrane potentials (V_{rest}) were not different between genotypes. Values represent mean ± s.e.m. (WT, *Setd1a*$^{+/-}$: n = 23, n = 21; unpaired t-test, *P* = 0.66)

(R) Outward currents evoked by voltage-steps near AP threshold voltages (from -70mV) were reduced in *Setd1a*$^{+/-}$ mice. Values represent mean ± s.e.m. (WT, *Setd1a*$^{+/-}$: n = 30, n = 33; unpaired t-test, -40mV, *P* = 0.44; -30mV, *P* = 0.38).

Figure S2

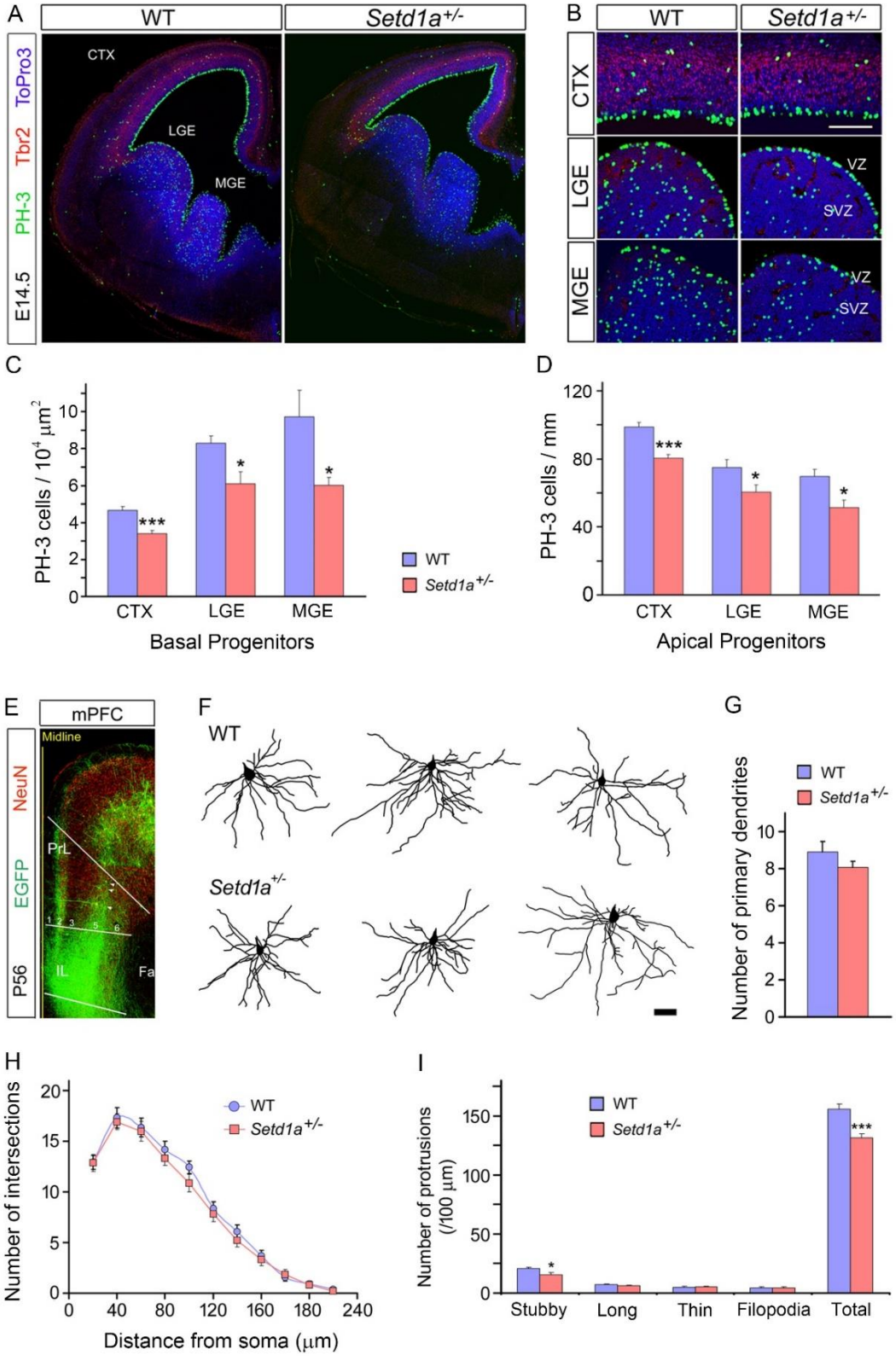
A
WT
Setd1a+/-
ToPro3
Tbr2
PH-3
E14.5
CTX
LGE
MGE
B
WT
Setd1a+/-
CTX
LGE
MGE
VZ
SVZ
C
PH-3 cells / 10^4 μm^2
CTX
LGE
MGE
Basal Progenitors
WT
Setd1a+/-
D
PH-3 cells / mm
Apical Progenitors
E
mPFC
Midline
NeuN
EGFP
P56
PrL
IL
Fa
F
WT
Setd1a+/-
G
Number of primary dendrites
H
Number of intersections
Distance from soma (μm)
I
Number of protrusions (/100 μm)
Stubby
Long
Thin
Filopodia
Total

Figure S2. The effect of *Setd1a*-deficiency on proliferation of NSCs at E14.5, and dendritic morphology of L5 pyramidal neurons.

(A) Decreased PH-3 labeling in the telencephalon in *Setd1a*$^{+/-}$ mice during corticogenesis.

(B) Images of coronal sections through the cortex (*top panel*), lateral ganglionic eminence (LGE) (*middle panel*), and medial ganglionic eminence (MGE) (*bottom panel*) at E14.5 showing mitotically active cells labeled with PH-3 in the ventricular zone (VZ) and subventricular zone (SVZ) of WT and *Setd1a*$^{+/-}$ mice.

(C and D) Histograms indicating reduced frequency both of mitotic basal (C, WT: 4.67 ± 0.18 cells/10^4 μm^2, N = 10, n = 30; *Setd1a*$^{+/-}$: 3.39 ± 0.18 cells/10^4 μm^2, N = 9, n = 27, $P < 0.0001$) and apical (D, WT: 98.77 ± 2.56 cells/mm, N = 10, n = 30 ; *Setd1a*$^{+/-}$: 80.50 ± 2.08 cells/mm , N = 9, n = 27, $P < 0.0001$) NPCs (which give rise to pyramidal neurons) in the cortices of *Setd1a*$^{+/-}$ mice. Similar analysis in the interneuron proliferative zones of the medial ganglionic eminence (MGE) and lateral ganglionic eminence (LGE), showed that basal PH-3$^+$ cells were significantly reduced both in LGE (C, WT: 8.30 ± 0.40 cells/10^4 μm^2, N = 8, n = 16; *Setd1a*$^{+/-}$: 6.11 ± 0.65 cells/10^4 μm^2 , N = 8, n = 16, $P < 0.02$) and MGE (C, WT: 9.74 ± 1.42 cells/10^4 μm^2, N = 8, n = 16; *Setd1a*$^{+/-}$: 6.01 ± 0.44 cells/10^4 μm^2 , N = 8, n = 16, $P < 0.04$) of *Setd1a*$^{+/-}$ mice. Similarly, apical PH-3$^+$ progenitors were reduced both in LGE (D, WT: 75.01 ± 4.01 cells/mm, N = 8, n = 16; *Setd1a* $^{+/-}$:

60.55 ± 4.21 cells/mm , N = 8, n = 16, $P < 0.04$), and MGE (D, WT: 69.60 ± 4.21 cells/mm, N = 8, n = 16; *Setd1a*$^{+/-}$: 51.36 ± 4.27 cells/mm , N = 8, n = 16, $P < 0.01$) of *Setd1a*$^{+/-}$ mice.

N, number of independent animals. n, number of brain slices. Data are shown as means ± s.e.m.

*$P < 0.05$, **$P < 0.01$.

(E) Representative image of EGFP-labeled neurons in layer 5 (*arrow heads*) in PrL of *Thy1-GFP*$^{+/-}$*:Setd1a*$^{+/-}$ mouse. Numbers indicate cortical layers. PrL: prelimbic area, IL: infralimbic area, Fa: corpus callosum anterior forceps. Images are composites of more than one image acquired from one brain section under identical scanning parameters.

(F-H) No alterations in dendritic complexity in the medial PFC of mutant mice.

(F) Representative tracings of the basal dendritic trees of L5 pyramidal neurons at the prelimbic area of medial PFC from *Setd1a*$^{+/-}$; *Thy1-GFP*$^{+/–}$ and WT mice.

(G) No alterations in number of primary dendrites in the basal dendritic tree of L5 pyramidal neurons between genotypes (n = 18 cells for *Setd1a*$^{+/-}$; n = 18 cells for WT, $P = 0.21$).

(H) Sholl analysis of basal dendrite complexity using 20 μm concentric circles around the soma (n = 18 cells for *Setd1a*$^{+/-}$; n = 18 cells for WT, $P > 0.05$).

(I) Distribution of spine morphotypes (other than mushroom spines), as well as total protrusions, in basal dendrites of layer 5 pyramidal neurons in prelimbic area from *Thy1-GFP*$^{+/-}$*;Setd1a*$^{+/-}$ mice. Dendritic protrusions were classified as different spine categories (mushroom, long, thin, or

stubby) or as filopodia, and counted. Data are shown as means ± s.e.m. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Student's two-tailed *t*-test (G and I) and 2-way repeated-measures ANOVA (H).

Figure S3

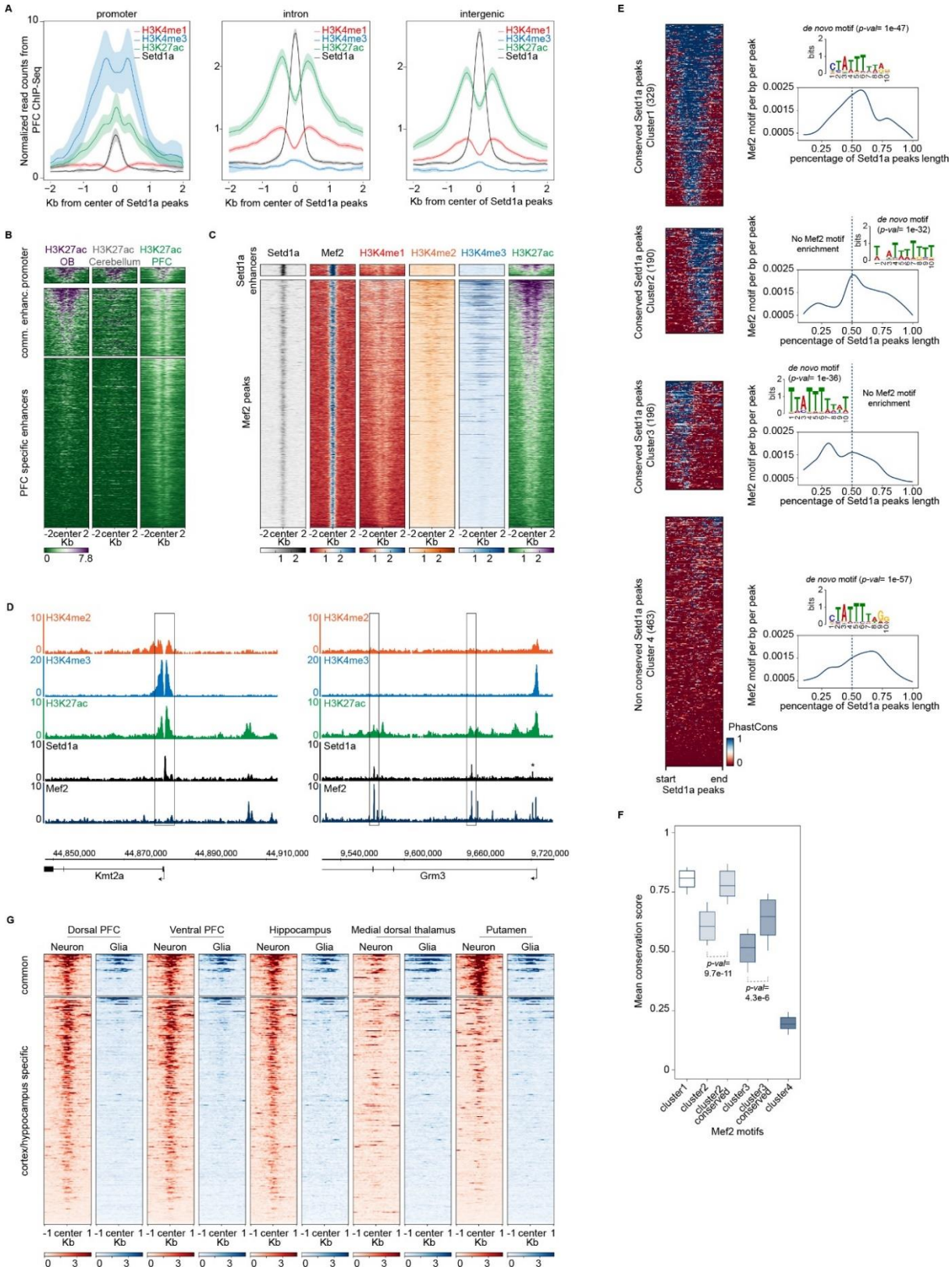
A
promoter
intron
intergenic
H3K4me1
H3K4me3
H3K27ac
Setd1a
Normalized read counts from PFC ChIP-Seq
Kb from center of Setd1a peaks
B
H3K27ac OB
H3K27ac Cerebellum
H3K27ac PFC
comm. enhanc. promoter
PFC specific enhancers
C
Setd1a
Mef2
H3K4me1
H3K4me2
H3K4me3
H3K27ac
Setd1a enhancers
Mef2 peaks
D
H3K4me2
H3K4me3
H3K27ac
Setd1a
Mef2
Kmt2a
Grm3
E
Conserved Setd1a peaks Cluster1 (329)
Conserved Setd1a peaks Cluster2 (190)
Conserved Setd1a peaks Cluster3 (196)
Non conserved Setd1a peaks Cluster 4 (463)
de novo motif (p-val= 1e-47)
de novo motif (p-val= 1e-32)
de novo motif (p-val= 1e-36)
de novo motif (p-val= 1e-57)
No Mef2 motif enrichment
Mef2 motif per bp per peak
pencentage of Setd1a peaks length
PhastCons
start Setd1a peaks end
F
Mean conservation score
p-val= 9.7e-11
p-val= 4.3e-6
cluster1
cluster2
cluster2 conserved
cluster3
cluster3 conserved
cluster4
Mef2 motifs
G
Dorsal PFC
Ventral PFC
Hippocampus
Medial dorsal thalamus
Putamen
Neuron
Glia
common
cortex/hypocampus specific

Figure S3. Setd1a and Mef2 co-bind PFC specific regulatory elements.

(A) Distribution of the mean normalized coverage 2 kb around the center of Setd1a peaks annotated in promoter (*left*), intron (*middle*) and intergenic regions (*right*). Coverage of PFC ChIP-Seq signal for H3K4me1 (*red*), H3K4me3 (*blue*), H3K27ac (*green*) and Setd1a (*black*) are shown. Shading indicates 95% confidence intervals.

(B) Heatmap of the coverage (scale at the very bottom) of H3K27ac in OB (*left*), cerebellum (*center*), PFC (*right*) in promoters (*top*), common enhancers (*middle*), PFC specific enhancers (*bottom*), 2 Kb around the center of Setd1a bound regions.

(C) Heatmap of the mean normalized coverage for Setd1a (*black*), Mef2 (*dark blue*) and chromatin marks (H3K4me1, *red*; H3K4me2, *orange*; H3K4me3, *light blue*; H3K27ac, *green*) ChIP-Seq performed from PFC tissues. Top, Setd1a enhancers that are also bound by Mef2; bottom Mef2 only bound regions.

(D) Examples of two loci bound by Setd1a only (*left*) or by both Setd1a and Mef2 (*right*) in the PFC. Left: Setd1a but not Mef2 binds the promoter region of *Kmt2a* as shown by the normalized reads per genomic coverage for Setd1a (*black*), Mef2 (*dark blue*), H3K4me2 (*orange*), H3K4me3 (*light blue*) H3K27ac (*green*). Right: Two distal enhancers in the *Grm3* locus are bound by both Mef2 and Setd1a. The asterisks indicate a weak non-significant Setd1a peak at the promoter of gene *Grm3*, significantly bound by Mef2.

(E) Heatmap of PhastCons values per base pairs of Setd1a peaks normalized by size (*left*), motif enrichment and distribution of the motif within Setd1a peaks of each cluster (*right*). Four clusters were defined according to their conservation level: one very conserved cluster (*top*); two partially conserved clusters (*top-middle* and *middle-bottom*) and one poorly conserved cluster (*bottom*). Each cluster is enriched for Mef2 motifs (*right panels*) and significant p-value is indicated. In cluster2 and cluster3 (*top-middle* and *middle-bottom*) the enrichment for Mef2 motifs is localized specifically in the most conserved region of peaks.

(F) Mean conservation score of Mef2 motif in each cluster.

(G) Heatmap of ATAC-Seq coverage 2 kb around the center of the most conserved Setd1a peaks between mouse and human (megablast, see Methods and Data Table 1) from different regions (dorsal PFC, ventral PFC, Hippocampus, Medial dorsal Thalamus and Putamen) of postmortem human brain from sorted neuronal (*red*) and glia (*blue*) cells.

Figure S4

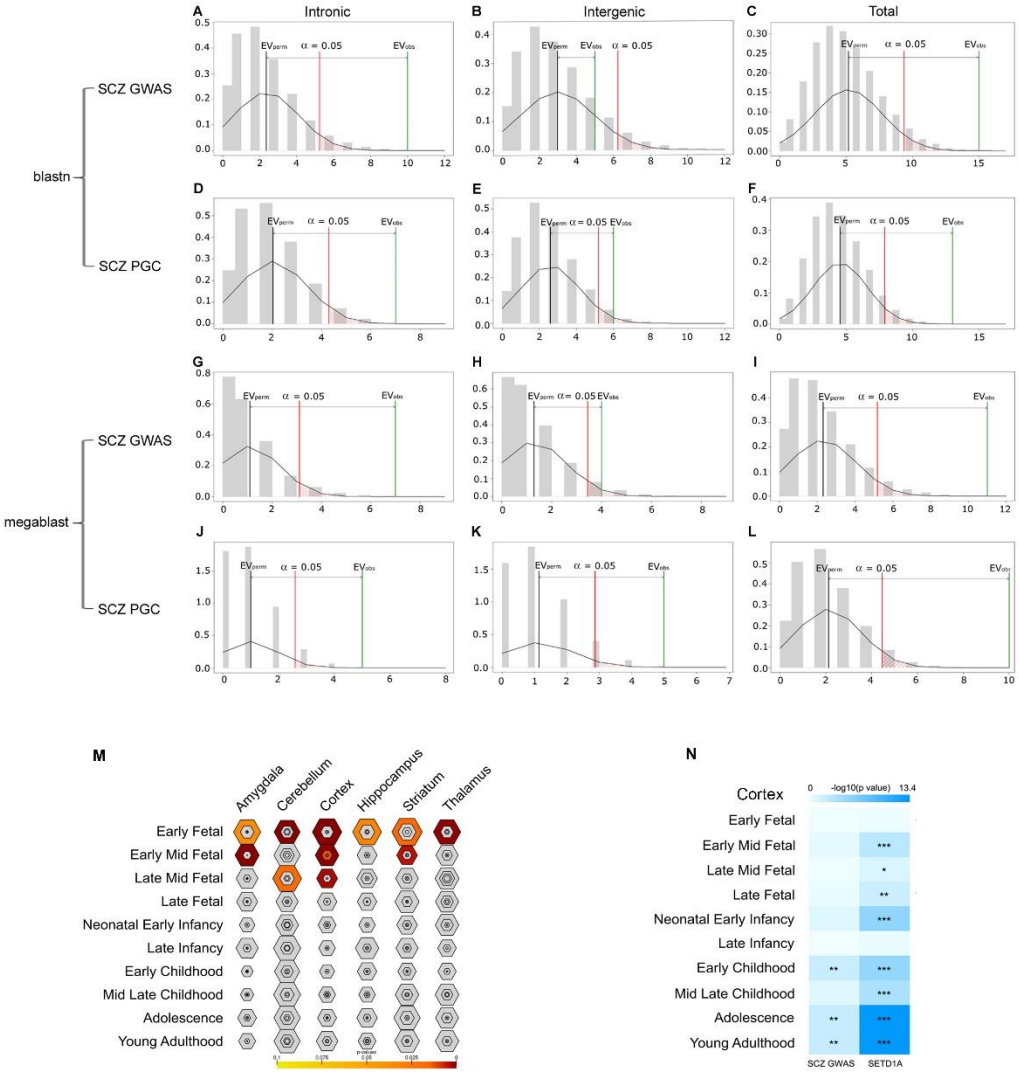

Figure S4. Overlap between genome-wide significant SCZ signals and conserved Setd1a ChIP-Seq peaks by genomic location and spatiotemporal expression analysis of SETD1A intronic target genes.

(A) Conserved Setd1a intronic peaks determine by the blastn method vs SCZ-GWAS intervals. *P* = 0.0015, Z-score = 4.358.

(B) Conserved Setd1a intergenic peaks determine by the blastn method vs SCZ-GWAS intervals. *P* = 0.2016, Z-score = 1.019.

(C) All conserved Setd1a peaks determine by the blastn method vs SCZ-GWAS intervals. *P* = 0.0012, Z-score = 3.863.

(D) Conserved Setd1a intronic peaks determine by the blastn method vs SCZ-PGC intervals. *P* = 0.0028, Z-score = 3.61.

(E) Conserved Setd1a intergenic peaks determine by the blastn method vs SCZ-PGC intervals. *P* = 0.0443, Z-score = 2.164.

(F) All conserved Setd1a peaks determine by the blastn method vs SCZ-PGC intervals. *P* = 0.0004, Z-score = 4.137.

(G) Conserved Setd1a intronic peaks determine by the megablast method vs SCZ GWAS intervals. *P* = 0.0024, Z-score = 4.828.

(H) Conserved Setd1a intergenic peaks determine by the megablast method vs SCZ-GWAS intervals. *P* = 0.0656, Z-score = 2.07.

(I) All conserved Setd1a peaks determine by the megablast method vs SCZ-GWAS intervals. *P* = 0.0004, Z-score = 4.965.

(J) Conserved Setd1a intronic peaks determine by the megablast method vs SCZ-PGC intervals. *P* = 0.0022, Z-score = 4.122.

(K) Conserved Setd1a intergenic peaks determine by the megablast method vs SCZ-PGC intervals. *P* = 0.0054, Z-score = 3.66.

(L) All conserved Setd1a peaks determine by the megablast method vs SCZ PGC intervals. *P* = 0.0001, Z-score = 5.547.

All tests use permutations (n=20,000, randomization method = circularRandomizeRegions) to determine if there are significant overlaps between conserved Setd1a ChIP-Seq peaks and previously identified SCZ-GWAS loci (Schizophrenia Working Group of the Psychiatric Genomics, 2014). X axis represents the number of overlaps, Y axis represents the density of expected number of overlaps determined by permutation. EVperm: Expected average number of overlaps by permutation (*black line*), EVobs: actual observed number of overlaps (*green line*). α, permutation *P* value threshold = 0.05 (*red line*). blastn: conserved Setd1a ChiP-seq peaks that were determined by blastn method. megablast: conserved Setd1a ChiP-seq peaks that were determined by megablast method. SCZ-GWAS: the independent schizophrenia GWAS intervals calculated by the pipeline described in Methods. SCZ-PGC: the independent schizophrenia GWAS intervals calculated in the original publication (Schizophrenia Working Group of the Psychiatric Genomics, 2014).

(M) Mutation intolerant genes with LoF DNMs in various NDD show prenatal-biased expression in various brain regions.

(N) Both SETD1A high confidence target genes and SCZ-GWAS most likely candidate genes (Shohat et al., 2017) manifest postnatal expression bias.

Figure S5

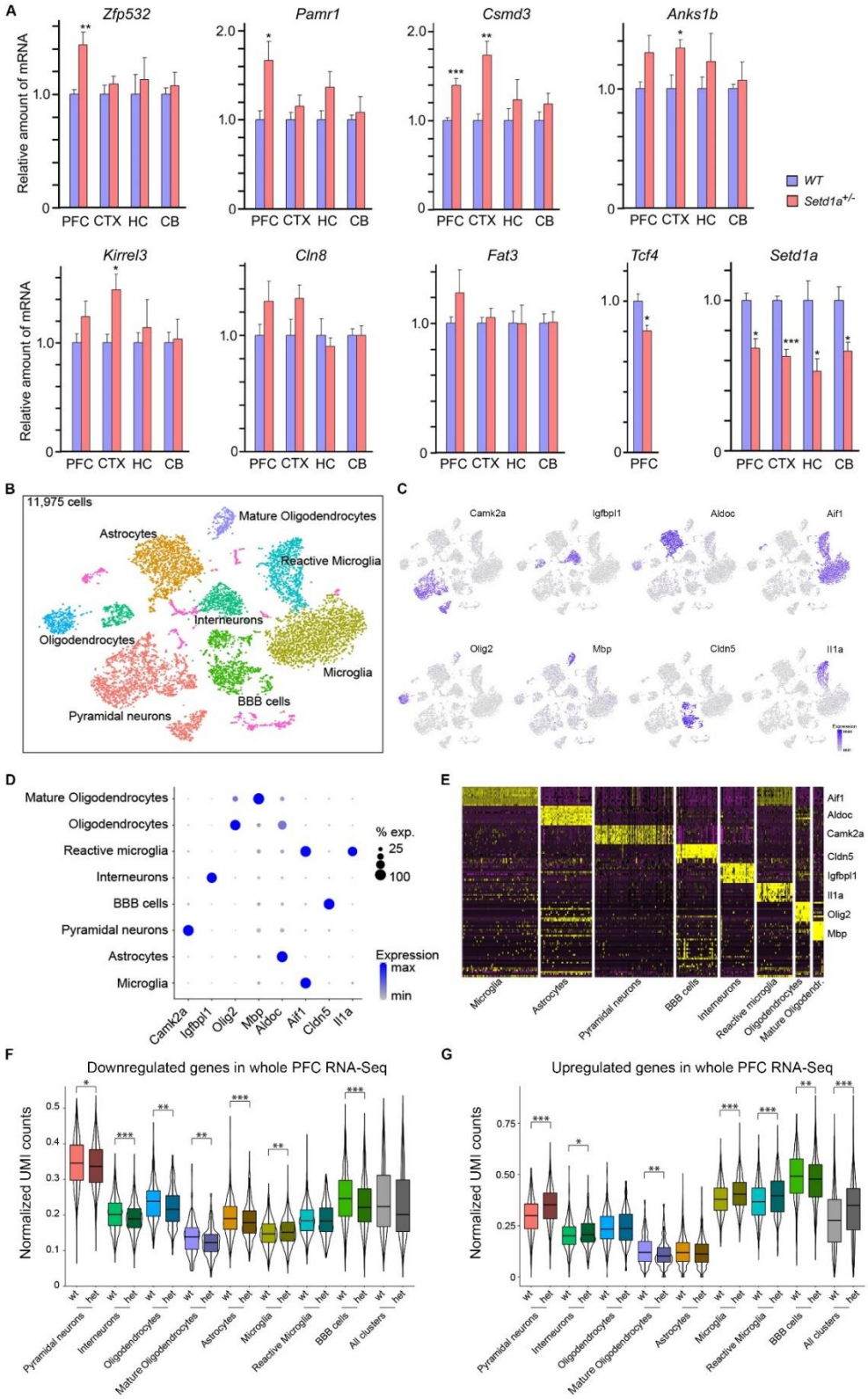
A
Zfp532
Pamr1
Csmd3
Anks1b
Kirrel3
Cln8
Fat3
Tcf4
Setd1a
Relative amount of mRNA
PFC CTX HC CB
WT
Setd1a+/-
B
11,975 cells
Mature Oligodendrocytes
Astrocytes
Reactive Microglia
Interneurons
Oligodendrocytes
Microglia
BBB cells
Pyramidal neurons
C
Camk2a
Igfbpl1
Aldoc
Aif1
Olig2
Mbp
Cldn5
Il1a
D
Mature Oligodendrocytes
Oligodendrocytes
Reactive microglia
Interneurons
BBB cells
Pyramidal neurons
Astrocytes
Microglia
% exp.
25
100
Expression
max
min
E
Aif1
Aldoc
Camk2a
Cldn5
Igfbpl1
Il1a
Olig2
Mbp
F
Downregulated genes in whole PFC RNA-Seq
Normalized UMI counts
wt
het
Pyramidal neurons
Interneurons
Oligodendrocytes
Mature Oligodendrocytes
Astrocytes
Microglia
Reactive Microglia
BBB cells
All clusters
G
Upregulated genes in whole PFC RNA-Seq
Normalized UMI counts

Figure S5. Confirmatory qRT-PCR analysis of a set of high confidence Setd1a targets and scRNA-Seq analysis.

(A) qRT-PCR analysis of a set of high confidence Setd1a targets confirmed the general trend for upregulation observed in our RNA-seq analysis. Only one direct Setd1 target, *Tcf4* (the orthologue of a strong candidate SCZ susceptibility gene), showed decreased mRNA expression in the PFC of 6-week-old *Setd1a*$^{+/-}$ mice (N = 6), compared to WT littermates (N = 6). Data are shown as means ± s.e.m. *$P < 0.05$, **$P < 0.01$, Student's two-tailed *t*-test.

(B) tSNE (t-distributed stochastic neighbor embedding) representation of 11,975 single cell transcriptomes from the PFC of WT and *Setd1a*$^{+/-}$ mice colored by clusters.

(C) Expression of several known cell types maker genes for neuronal and non-neurons cells: *CamK2a* (pyramidal neurons), *Igfbpl1* (interneurons), *Aldoc* (astrocytes), *Aif1* (microglia), *Olig2* (oligodendrocytes), *Mbp* (mature oligodendrocytes), *Cldn5* (endothelial cells), *Il1a* (reactive microglia).

(D) Dot-plot of the expression of the neuronal and glial markers.

(E) Heatmap of the top 10 marker genes that were identified *de novo*. On the right, the genes that were used to identify each cluster of cells.

(F) Average expression of all downregulated genes that were identified through whole-PFC RNA-Seq per cell per cluster per genotype. Genes with average Normalize count below 0.1 were filtered out of the analysis.

(G) Average expression of all upregulated genes that were identified through whole- PFC RNA-Seq per cell per cluster per genotype. Genes with average Normalize count below 0.1 were filtered out of the analysis.

Figure S6

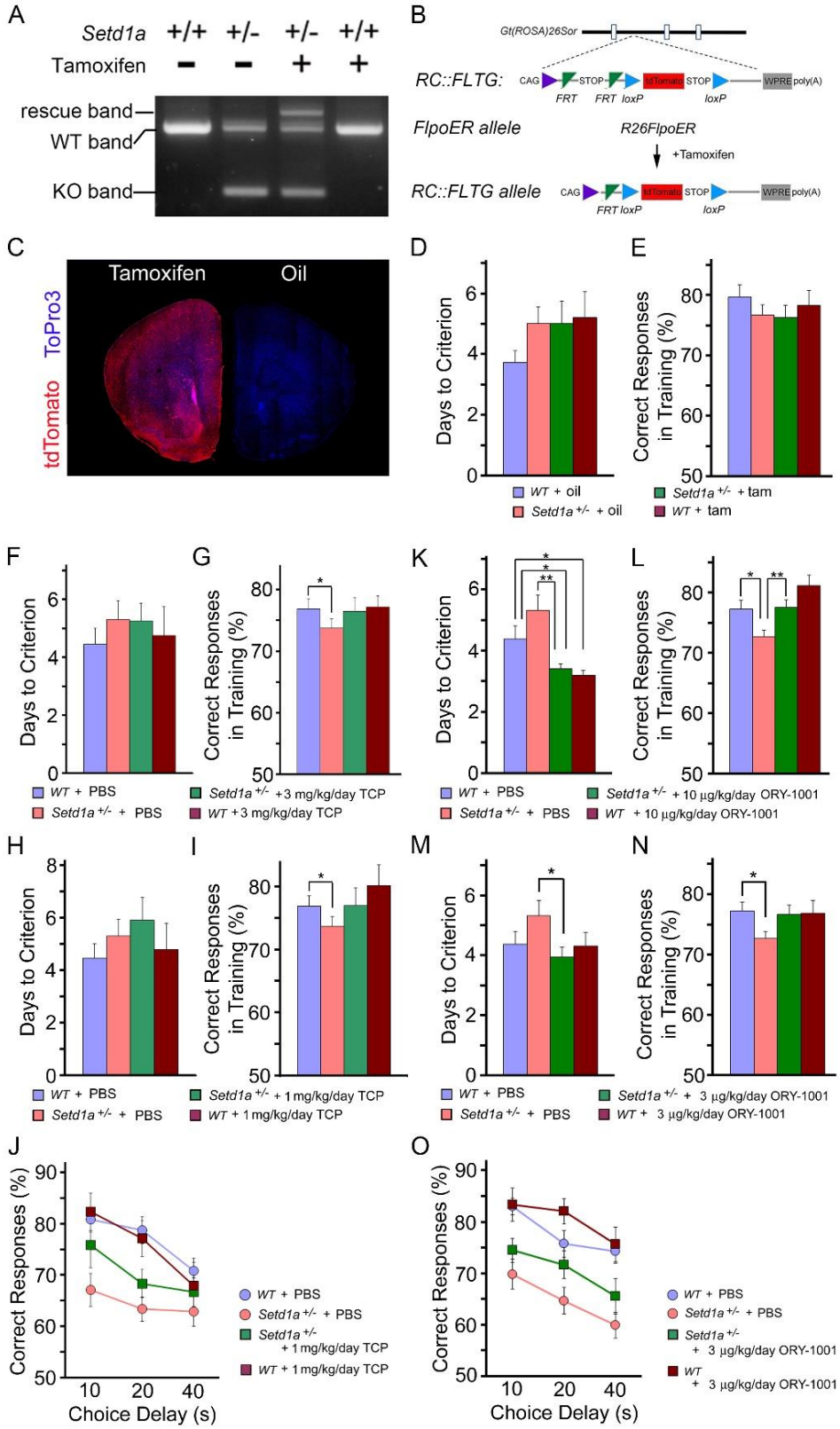
A
Setd1a +/+ +/- +/- +/+
Tamoxifen − − + +
rescue band
WT band
KO band
B
Gt(ROSA)26Sor
RC::FLTG:
CAG STOP tdTomato STOP WPRE poly(A)
FRT FRT loxP loxP
FlpoER allele
R26FlpoER
+Tamoxifen
RC::FLTG allele
FRT loxP loxP
C
Tamoxifen
Oil
ToPro3
tdTomato
D
Days to Criterion
WT + oil
Setd1a +/- + oil
E
Correct Responses in Training (%)
Setd1a +/- + tam
WT + tam
F
Days to Criterion
WT + PBS
Setd1a +/- + PBS
G
Correct Responses in Training (%)
Setd1a +/- +3 mg/kg/day TCP
WT +3 mg/kg/day TCP
K
Days to Criterion
WT + PBS
Setd1a +/- + PBS
L
Correct Responses in Training (%)
Setd1a +/- + 10 µg/kg/day ORY-1001
WT + 10 µg/kg/day ORY-1001
H
Days to Criterion
WT + PBS
Setd1a +/- + PBS
I
Correct Responses in Training (%)
Setd1a +/- + 1 mg/kg/day TCP
WT + 1 mg/kg/day TCP
M
Days to Criterion
WT + PBS
Setd1a +/- + PBS
N
Correct Responses in Training (%)
Setd1a +/- + 3 µg/kg/day ORY-1001
WT + 3 µg/kg/day ORY-1001
J
Correct Responses (%)
Choice Delay (s)
WT + PBS
Setd1a +/- + PBS
Setd1a +/- + 1 mg/kg/day TCP
WT + 1 mg/kg/day TCP
O
Correct Responses (%)
Choice Delay (s)
WT + PBS
Setd1a +/- + PBS
Setd1a +/- + 3 µg/kg/day ORY-1001
WT + 3 µg/kg/day ORY-1001

Figure S6. **WM performance following *Setd1a* re-activation and pharmacological inhibition of LSD1 in adulthood.**

(A-D) Activation of *Setd1a* expression in adulthood and effects on WM performance.

(A) PCR genotyping showing the expected bands (KO, rescue and WT) generated by tamoxifen inducible Flp-mediated recombination at the targeted *Setd1a* genomic locus

(B) Schematic representation of tamoxifen-inducible reporter expression strategy. The Flp responsive fluorescent indicator allele, *RC:FLTG* was previously generated by targeting intron 1 of the *Gt(ROSA)26Sor* locus. FRT flanked transcriptional stop cassettes in *RC::FLTG* prevent expression of tdTomato in the absence of Flp activity. Tamoxifen-inducible Flpo expression excises stop cassettes and leads to tdTomato expression in $R26^{FlpoER+/-}$*:RC: :FLTG*$^{+/-}$ mice.

(C) Tamoxifen-inducible Flp strategy leads to broad reporter expression. Coronal sections from $R26^{FlpoER+/-}$*:RC::FLTG*$^{+/-}$ mice after feeding with tamoxifen (*left*) or corn oil (*right*). Results show widespread induction of tdTomato reporter expression after tamoxifen–induced Flp activation, but not in the absence of Flp activity (corn oil).

(D) Days taken to criterion performance (three consecutive days of > 70% choice accuracy) during training on the delayed non-match to sample T-maze task following activation of Setd1a expression in adulthood.

(E) Correct responses during training on the delayed non-match to sample T-maze task following activation of Setd1a expression in adulthood.

(F and G) TCP administration in adulthood and effects on WM performance. Days taken to criterion performance (three consecutive days of > 70% choice accuracy) (F) and mean percentage of correct responses for WT and *Setd1a*$^{+/-}$ mice (G) during training on the delayed non-match to sample T-maze task following 3 mg/kg/day TCP or vehicle treatment. *$P < 0.05$.

(H-J) Low concentration (1 mg/kg/day) of TCP does not rescue WM performance in the delayed non-match to place task. Days taken to criterion performance (three consecutive days of > 70% choice accuracy) (H), and mean percentage of correct responses during training (I) were not affected. In addition, the impairment in the performance of vehicle-treated mutant mice over vehicle treated WT control littermates during testing with increasing time delays was not completely reversed by 1 mg/kg/day TCP treatment (J) [TCP-treated *Setd1a*$^{+/-}$ mice correct responses: 75.83 ± 4.83%, $P = 0.17$ at 10 sec; 68.33 ± 2.37%, $P = 0.47$ at 20 sec; 66.67 ± 2.78%, $P = 0.79$ at 40 sec, N = 18, compared to vehicle-treated *Setd1a*$^{+/-}$ control correct responses: 67.08 ± 3.23% at 10 sec; 63.33 ± 2.37% at 20 sec; 62.92 ± 2.93% at 40 sec, N = 20].

(K and L) WM performance following administration of 10 μg/kg/day ORY-1001 in adulthood. Days to criterion (K) and percentage of correct choices (L) during training phase were improved in both ORY-1001-treated *Setd1a*$^{+/-}$ and WT mice [ORY-1001-treated *Setd1a*$^{+/-}$ mice: Days to criterion:

3.4 ± 0.16 days, P = 0.0013 versus vehicle-treated *Setd1a*$^{+/-}$; P = 0.040 versus vehicle-treated WT, N = 15. Correct responses: 76.02 ± 1.72%, P = 0.0085 versus *Setd1a*$^{+/-}$ vehicle. ORY-1001-treated WT mice: Days to criterion: 3.2 ± 0.14 days, P = 0.014 versus WT vehicle, N = 15, compared to vehicle-treated (*Setd1a*$^{+/-}$ vehicle: Days to criterion; 5.32 ± 0.50 days, correct responses; 72.69 ± 1.11%, N = 21, WT vehicle; Days to criterion; 4.38 ± 0.42 days, correct responses; 77.25 ± 1.42%, N = 23) . *P < 0.05, **P < 0.01.

(M-O) Low concentration (3 μg/kg/day) of ORY-1001 does not rescue WM performance in the delayed non-match to place task. Shown are days taken to criterion performance (three consecutive days of > 70% choice accuracy) during training (M) and *m*ean percentage of correct responses during training (N). In addition, an impairment in the performance of vehicle-treated mutant mice over vehicle treated WT control littermates during testing with increasing time delays was not effectively reversed by 3 μg/kg/day ORY-1001 treatment (O) [ORY-1001-treated *Setd1a*$^{+/-}$ mice correct responses: 74.44 ± 2.37%, P = 0.22 at 10 sec; 71.67 ± 2.67%, P = 0.07 at 20 sec; 65.56 ± 3.44%, P = 0.31 at 40 sec, compared to vehicle-treated *Setd1a*$^{+/-}$ control correct responses; 69.84 ± 2.96% at 10 sec; 64.68 ± 2.63% at 20 sec; 59.92 ± 2.48% at 40 sec, N = 21]. WT + PBS, N = 23, *Setd1a*$^{+/-}$ + PBS, N = 21, *Setd1a*$^{+/-}$ + 3 μg/kg/day ORY-1001, N = 15, WT + 3 μg/kg/day ORY-1001, N = 15.

Table S1. Overlap between conserved Setd1a Chip-Seq peaks and GWAS signals of various complex disorders

GWAS Category	# of signal peaks	SETD1A target peaks interval (blastn) versus GWAS			SETD1A target peaks interval (megablast) versus GWAS		
		Permutation *P* value (n = 20,000)	# of Overlaps	Z score	Permutation *P* value (n = 20,000)	# of Overlaps	Z score
Schizophrenia	124	0.00125	15	3.86	0.0004	11	4.96
SCZ_PGC*	108	0.00045	13	4.14	0.0001	10	5.55
Blood_metabolite_levels	159	0.34213	2	-0.73	0.44358	2	0.13
Body_mass_index	127	0.22384	6	0.91	0.50297	2	0.1
Bone_mineral_density	77	0.0676	5	1.95	0.25339	2	0.95
Breast_cancer	79	0.26199	0	-1.14	0.59217	0	-0.71
Cholesterol,_total	81	0.67507	2	-0.01	0.58757	1	0.09
Crohn's_disease	126	0.62902	3	-0.06	0.64292	1	-0.22
HDL_cholesterol	97	0.017	0	-1.9	0.08985	0	-1.51
Inflammatory_bowel_disease	112	0.52397	2	-0.36	0.29299	2	0.89
LDL_cholesterol	74	0.21014	0	-1.16	0.50817	0	-0.75
Menarche_(age_at_onset)	128	0.53337	3	-0.28	0.46348	2	0.28
Prostate_cancer	99	0.18694	0	-1.22	0.49188	0	-0.79
QT_interval	77	0.58512	1	-0.38	0.58257	0	-0.64
Rheumatoid_arthritis	109	0.31248	3	0.54	0.19854	2	1.02
Type_2_diabetes	86	0.49508	1	-0.53	0.46823	0	-0.84
Ulcerative_colitis	74	0.25244	0	-1.14	0.57072	0	-0.72

NOTE: SCZ_PGC data was obtained from Ref. (Schizophrenia Working Group of the Psychiatric Genomics, 2014).. For other GWAS signals see Methods

Table S2. Overlap between conserved Setd1a Chip-Seq peaks and GWAS signals of various CNS-related disease

GWAS Category	# of signal peaks	SETD1A target peaks interval (blastn) versus GWAS			SETD1A target peaks interval (megablast) versus GWAS		
		Permutation *P* value (n = 20,000)	# of Overlaps	Z score	Permutation *P* value (n = 20,000)	# of Overlaps	Z score
Autism	20	0.286	2	0.863	0.085	2	2.136
Attention deficit	16	0.255	2	1.002	0.076	0	2.268
ALS	17	0.204	1	1.291	0.927	0	-0.256
Alzheimer's disease	53	0.007	8	3.745	0.518	0	-0.686
Parkinson's disease	42	0.361	3	0.835	0.683	0	-0.572

NOTE: SCZ_PGC data was obtained from Ref. (Schizophrenia Working Group of the Psychiatric Genomics, 2014).. For other GWAS signals see Methods

Table S3. Enrichment of target genes with intronic Setd1a peaks in SCZ-GWAS intervals

GWAS	Permutation *P* value	# of genes in the GWAS intervals
Schizophrenia	0.007	11
SCZ_PGC*	0.005	12
body_mass_index	0.294	6
blood_metabolite_level	0.968	3
Bone_mineral_density	0.198	5
Breast_cancer	1	1
Rheumatoid_arthritis	0.882	2
Cholesterol_total	1	1
Crohn_disease	0.721	4
HDL_cholesterol	0.91	2
Inflammatory_bowel_disease	1	1
LDL_cholesterol	1	0
Menarche	0.755	4
Prostate_cancer	1	1
QT_interval	1	1
Type-2-diabetes	0.229	5
Ulcerative colitis	1	0

NOTE: SCZ_PGC data was obtained from Ref. (Schizophrenia Working Group of the Psychiatric Genomics, 2014). For other GWAS signals see Methods

Table S4. Enrichment of target genes with intronic Setd1a peaks in CNS-related disease GWAS intervals

GWAS	Permutation *P* value	# of genes in the GWAS intervals
Autism	1	1
Attention_deficit	1	1
Alzheimer's_disease	0.298	3
Amyotrophic _lateral_sclerosis	0.117	2
Parkinson's_disease	0.315	2

NOTE: SCZ_PGC data was obtained from Ref. Schizophrenia Working Group of the Psychiatric Genomics, 2014).

For other GWAS signals see Methods

Table S5. Enrichment of Setd1a and Mef2 bound targets in the SCZ GWAS intervals

	# of genes tested	*P* value	# of genes in GWAS intervals
Setd1a bound intronic targets	263	0.007	11
Mef2 bound intronic targets	3302	0.021	60
Setd1a and Mef2 co-bound targets	260	0.005	11
Mef2 only bound intronic targets	3042	0.271	49