Shared molecular consequences of epigenetic machinery disruption in neuronal progenitors

Juan Ouyang[1μ, €], Kaan Okay[1μ], Katrín Möller[1μ], Katrín Wang[1], Stefán Pétursson[1], Tessa Pierce[1,2], Arsalan Amirfallah[1δ], Adam Davidovich[4], Kimberley Anderson[1,3], Kasper D. Hansen[4,5,6*], Hans Tomas Bjornsson[1,3,6*]

[1]Louma G. Laboratory of Epigenetic Research, Faculty of Medicine, University of Iceland, Reykjavik, Iceland; [2] California Institute of Technology, Pasadena, California, USA; [3]Department of Genetics and Molecular Medicine, Landspitali University Hospital, Reykjavik, Iceland; [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA; [5]Department of Bioengineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [6]Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[μ]These authors had equal contribution

[€]Current position: Emory University, Atlanta, GA, USA.

[δ]Current position: The Blood Bank, Landspitali University Hospital, Reykjavik, Iceland.

*Corresponding authors: Hans Tomas Bjornsson, htb@hi.is and Kasper Daniel Hansen, khansen@jhsph.edu

Abstract

The Mendelian disorders of the epigenetic machinery (MDEMs) are an emerging cause of intellectual disability and growth abnormalities. Hippocampal neurons are a consistently affected cell type in these disorders. To investigate the effects of epigenetic machinery (EM) disruption in a disease-relevant context, we established a murine hippocampal neurodifferentiation model enabling systemic knockout (KO) of multiple EM factors. We then interrogated each EM-KO using Oxford Nanopore DNA methylation (DNAm) sequencing and RNA sequencing. Disruption of *Kmt2a* and *Dnmt1* produced strikingly convergent gene expression changes. However, while *Dnmt1*-KO induced extensive DNAm alterations, KO of *Kmt2a* had little effect on methylation. Loss of either EM factor led to premature neuronal differentiation, partially explaining the transcriptional convergence. The transcription factor GSX1 emerged as a potential regulatory node, with altered methylation and expression in *Kmt2a*-KO cells, and ~20% of its known targets were differentially expressed. Extending our methylation analysis to 46 EM genes, we found that loss of DNA methyltransferases caused the strongest DNAm changes, whereas other EM-KOs induced milder-to-no effects. Clustering of EM factors based on promoter DNAm revealed two EM subgroups enriched for interactions with the DNAm machinery. Allele-specific analysis identified a shared differentially methylated region (DMR) at *Zic4* across the entire group of EM-KOs, and *Zic4* overexpression appears to maintain the NPC state. Taken together, our results reveal both gene-specific and convergent effects across diverse EM KOs and provide new insight into the molecular etiology of the MDEMs.

Keywords: Epigenetics, chromatinopathies, DNA methylation machinery, intellectual disability, Oxford Nanopore Technology, NPCs

Introduction

In the past 15 years clinicians have witnessed the rapid discovery of a group of at least 90 distinct disorders, caused mainly by heterozygous variants in the components that directly maintain epigenetic modifications. This group of disorders is described by different names, including the Mendelian disorders of the epigenetic machinery (MDEMs) (Bjornsson, 2015), chromatinopathies (Ciptasari & van Bokhoven, 2020), or rare diseases of epigenetic origin (Fu et al., 2023). Despite a multitude of names, all parties agree that the disruption of these genes is likely to lead to secondary epigenetic dysfunction. Remarkably, although these genes have been associated with broad phenotypic presentations (Bjornsson, 2015), they also have some unifying disease phenotypes, such as intellectual disability and growth abnormalities (Bjornsson, 2015), seen in the vast majority of patients. We have previously hypothesized that shared phenotypic abnormalities may be caused by shared epigenetic disruption. In support of this hypothesis, we have observed shared epigenomic abnormalities in three such MDEMs (Kabuki syndrome 1&2 and Rubinstein-Taybi syndromes): in blood cells for all three (Luperchio et al., 2021) and in neurons for the first two (Boukas et al., 2024).

Another defining feature of these disorders is that the majority of them (42) have been shown to have distinct blood cell DNAm episignatures with diagnostic utility (Aref-Eshghi et al., 2020). Remarkably, although blood may not always be the most disease-relevant tissue, many of these disorders have shared epigenetic abnormalities (Aref-Eshghi et al., 2018). A question emerges whether such DNA methylation (DNAm) changes can be found in disease-relevant cells, such as neurons, and whether they can point to underlying shared pathogenic mechanisms in these disorders. Here, we have focused on a hippocampal neuronal model, given prior evidence of disruption of hippocampal neurogenesis in many of the MDEM (Carosso et al., 2019; Feng et al., 2013; Noguchi et al., 2016; Park et al., 2014; Schoof et al., 2019; Sheu et al., 2019; Syal et al., 2018; You et al., 2015; Zhang et al., 2014; Zhou et al., 2016).

Recently, Oxford Nanopore Technologies (ONT) DNAm sequencing has become available, allowing the discovery of genome-wide DNAm levels and offering the possibility of phasing to look for haplotype-specific effects or effects on genomic imprinting (Gigante et al., 2019). Here, we describe a catalogue of DNAm abnormalities seen in 46 EM KOs in a single disease-relevant cell-type context, murine neuronal progenitor cells (mNPCs). Our work identifies several TFs which appear to mediate aspects of the observed chromatin abnormalities or show dysregulation secondarily to the epigenetic abnormality. This is

interesting since the EM machinery in general has limited sequence preferences (Rothbart & Strahl, 2014) and DNAm can interfere with transcription factor binding (Heberle & Bardet, 2019). Therefore, it may not be surprising that systematic KO of EM factors may reveal key TFs whose role in neuronal development is modulated by DNAm dynamics. Our approach, which allows phasing of the genomes, also highlights the considerable effect that genetic variation has on genome-wide levels of DNAm. Together, our results lay the groundwork for understanding how the epigenetic machinery (EM) helps orchestrate hippocampal neurogenesis and how disruption of this process through EM loss uniformly leads to learning and neurological defects.

Results

Systematic KO of EM factors in a single neuronal system

To comprehensively study the role of DNAm abnormalities in EM-deficient neurons, we utilized CRISPR-Cas9 technology to knock-out (KO) EM genes in mNPCs. To obtain homogenous and efficient CRISPR-based gene KOs, we used F1 offspring from two fully sequenced mouse-strains (C57BL/6J and FVB), both expressing transgenic Cas9-GFP protein from the ROSA locus, a well-known epigenetic safe harbor region (Friedrich & Soriano, 1991; Irion et al., 2007). We isolated hippocampal mNPCs at postnatal day 0 (P0) from this F1-outcross and grew them in culture media that helped maintain the mNPC state. We verified cell identities (**Supplementary Figure 1A**) prior to using these for individual EM-KO experiments using 46 dual guide-RNA (sgRNA) lentiviral plasmids (**Supplementary Table 1**), each containing two sgRNA sequences targeting the same gene. We also created eight sgRNA plasmids with random sequences (scramble gRNAs, **Supplementary Table 1**). These guide-vectors were individually packaged and transduced using lentiviral delivery into the Cas9-expressing mNPCs (**Figure 1A**). Cells were selected for stable integration of a construct, and four days after transduction, harvested for profiling DNAm (ONT sequencing) or expression (Illumina RNA-seq).

To confirm the efficiency of the KO in our system, we performed qPCR, western blot (WB), and immunofluorescence (IF) assays on individual EM-KO mNPCs. As expected, four days after transduction, we routinely saw a decrease in mRNA and protein levels for individual EM KOs. As an example, we observed significant ($p < 0.05$) downregulation of *Kmt2d* and *Ezh2* at the RNA level (**Figure 1B-C**) as well as a significant ($p < 0.05$) decrease

in protein levels of CHD1, KMT2A, and DNMT1 (**Figure 1D-E, Supplementary Figure 1B-C**).

To enable accurate and systematic estimation of CRISPR KO efficiency from our ONT long-read sequencing data, we developed *samCRISPR*, a custom pipeline designed to leverage the unique features of this sequencing technique. This approach utilizes Minimap2 (Li, 2018) for sequence alignment and SAMtools (Danecek et al., 2021) for downstream processing (**Figure 1F**). When we compared *samCRISPR* with *CRISPResso2* (Pinello et al., 2016), the gold standard for estimating KO efficiency from short-read data, it yielded fewer false positive CRISPR events on average in control samples (2.1% vs 4.9% for sgRNA1 ($p < 0.001$) and 1.4% vs 4.5% for sgRNA2 ($p < 2.2e-16$); **Supplementary Figure 1D-E**). Moreover, *samCRISPR* yielded a similar KO efficiency for each guide (39.1% vs 37.5% for sgRNA1 and 36.85% vs 35.3% for sgRNA2; **Figure 1G, Supplementary Figure 1F-G**) across 46 EM KOs. As an example, when we ran two replicates of *Kmt2a*-KO and two relevant controls, we found almost no false positives in the control samples, while we detected on average 40% KO efficiency for either allele of the *Kmt2a*-KOs (**Figure 1H**). An underestimate of the actual KO efficiency is expected for both methods due to the small detection window (+/- 1bp around cut-site), and both WB and IF showed higher KO efficiency, i.e. lower protein expression (40-90%) in *Chd1*-, *Kmt2d*- and *Dnmt1*-KOs (**Figure 1D-E, Supplementary Figure 1B-C**) supporting the notion that a portion of the functional variants are not captured by either samCRISPR or CRISPResso2.

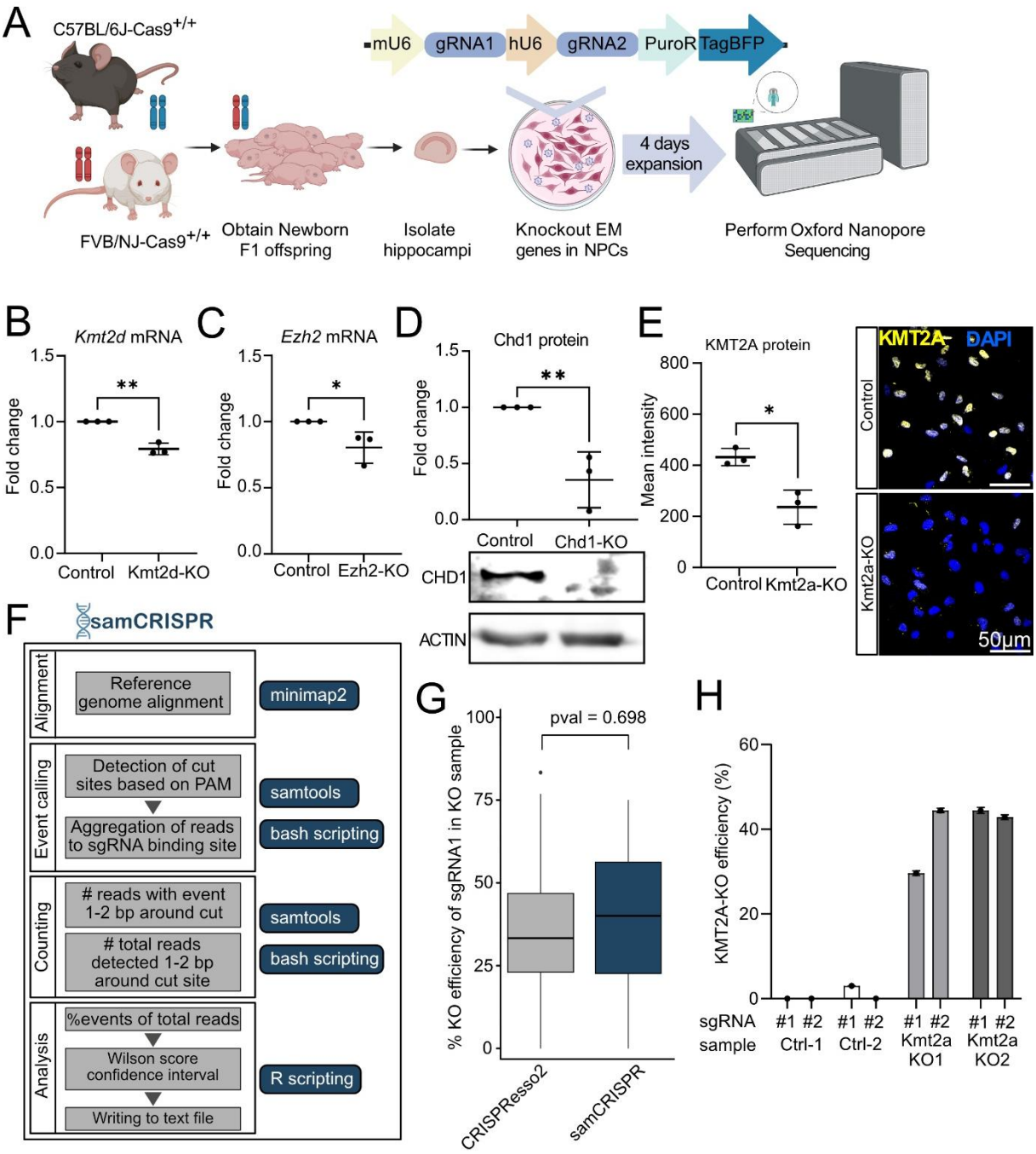


Figure 1: Effective CRISPR-Cas9 targeting in a disease relevant cellular system. (A) Schematic workflow of EM-KO in mNPCs using lentiviral vectors followed by evaluation of DNAm with ONT sequencing. (B) Gene expression changes of *Kmt2d* (p=0.0013) and (C) *Ezh2* (p=0.0229) in mNPCs after knocking out *Kmt2a* or *Ezh2* with CRISPR/Cas9 compared with scrambled control (n=3). An unpaired t-test was used to calculate significance. (D) Protein levels and quantification of CHD1 after knocking out *Chd1* with CRISPR/Cas9, compared with scrambled control (n=3, p=0.0054) An unpaired t-test was used to calculate significance. (E) Quantification *Kmt2a* mean expression per cell in *Kmt2a*-KO and scrambled controls (n=3, p=0.0001). An unpaired t-test was used to calculate significance. (F) Schematic of the *samCRISPR* analysis tool for KO efficiency estimation of long-read sequencing data. (G) The KO efficiency estimation by samCRISPR method vs CRISPResso2, for individual

142 sgRNAs in 46 EM KO from ONT WGS sequencing data (Two tailed t-test). (H) Bar plot
143 showing *Kmt2a*-KOs efficiencies for sgRNA #1 and #2 in two *Kmt2a*-KOs and controls (Ctrl)
144 calculated by samCRISPR. *p<0.05, **p<0.01, ***p<0.001, ns=not significant.

145 **DNAm and transcriptional changes in *Dnmt1*-KO and *Kmt2a*-KO mNPCs**

146 Having verified the KO efficiency of our system, we investigated if disrupting
147 individual EM genes in our mNPCs would result in DNAm abnormalities. As *Dnmt1* encodes
148 the key maintenance DNA methyltransferase, we used it as a positive control for assessing the
149 ONT sequencing approach. We performed ONT sequencing (Simpson et al., 2017) on *Dnmt1*-
150 KO and scrambled controls in biological triplicates. Differential methylation analysis revealed
151 a global demethylation pattern in *Dnmt1*-KO cells, including 141,721 hypomethylated
152 differentially methylated regions (DMRs) (qval < 0.05, **Figure 2A, Supplementary Table 2**)
153 and a sole hypermethylated region in *Zic1* gene body (**Figure 2B**). This is consistent with
154 prior studies showing that loss of DNMT1 in human NPCs causes global demethylation
155 (Jonsson et al., 2019). *Dnmt1*-KO DMRs were enriched in 3´UTRs, introns, and exons but
156 depleted in promoters, intergenic regions, and 5´UTRs (**Figure 2C, Supplementary Table 3**).

157 Given the extensive DNAm changes in *Dnmt1*-KO, we explored whether *Dnmt1*-KO DMRs
158 show enrichment of TF binding motifs using an Analysis of Motif Enrichment (AME;
159 MEME-suite, mouse HOCOMOCO v11 core motifs). This analysis identified 178 known TF
160 binding motifs that were significantly enriched (Fisher's exact test, adjusted pval <0.05,
161 **Figure 2D and Supplementary Table 4**), including a motif for MECP2, which recruits
162 DNMT1 to hemi or fully methylated CpGs for its maintenance function (Kimura & Shiota,
163 2003). However, we also found other noteworthy TFs, some with CpG sites in their binding
164 motifs, such as HIF1α, AHR and ARNT, and others lacking CpGs in their motifs, including
165 SMAD3, ESR2, NKX2-5 and STF1 (**Figure 2D, Supplementary Table 4**). Several hypoxia-
166 related factors (HIF1α, EPAS1 and ARNT) are of particular interest given our recent
167 discovery of dysregulated hypoxia signaling in Kabuki syndrome, another MDEM (Carosso et
168 al., 2019). We next asked if the enrichment of TF motifs in the *Dnmt1*-KO associated DMRs
169 could reflect known DNMT1 interaction partners. We compared these motifs with known
170 DNMT1 interactors from (Oughtred et al., 2021) and found that of the 24 interactors
171 represented in the HOCOMOCO database, 14 of TFs exhibited motif enrichment in *Dnmt1*-
172 KO DMRs (**Supplementary Figure 2A**).

173 To investigate whether methylation changes in *Dnmt1*-KOs coincide with changes in
174 transcription, we performed RNA-seq on *Dnmt1*-KO cells which resulted in 3,454

differentially expressed genes (DEGs) compared with controls (**Supplementary Table 5**). We used HOMER to annotate the *Dnmt1*-KO DMRs and found that 2,713 DEGs are associated with 26,909 DMRs (**Figure 2E, purple**). Among these were, for example, *Ppfibp1* (**Supplementary Figure 2B**), a gene encoding a factor involved in synaptic protein-protein interactions (PPIs) (Wei et al., 2011), the immediate early genes *Jun*, *Fos* and *Egr1* (**Figure 2F-G, Supplementary Figure 2C)** (Kim et al., 2018; Veyrac et al., 2014), as well as *Rora* and *Itga8* (**Supplementary Figure 2C-D**), previously identified DNMT1 target genes (Tao et al., 2022). We also found that mouse imprinted genes (Isles, 2022) were significantly overrepresented in the *Dnmt1*-KO associated DEGs (Fisher's exact test, OR = 1.83, p=0.0118), including for example *Zim1* (**Figure 2H).** This enrichment increased when we focused on the subset of DMR-associated DEGs (Fisher's exact test, OR = 2.13, p=0.003). While DEGs did not overlap with DMRs more frequently than other detected genes (Fisher's exact test, OR = 1.07, p=0.069), DEGs carried on average more DMRs than non-DEGs (7.6 vs 6.4, respectively, Wilcoxon rank sum test, $p = 3.57x10^{-06}$ **Figure 2I**). Together, this suggests that the wide-spread methylation changes we observe after *Dnmt1*-KO have relatively little impact on gene expression at this time point in our system.

We next examined KMT2A, a histone methyltransferase which is expected to have a more indirect influence on DNAm. Pathogenic variants in *KMT2A* are the cause of Wiedemann-Steiner syndrome (OMIM, 605130) which has a distinct DNAm signature in peripheral blood (Foroutan et al., 2022). To investigate whether *Kmt2a* loss also perturbs DNAm in neurons, we performed ONT sequencing on *Kmt2a*-KO NPCs in replicates, which revealed only three genome-wide significant DMRs compared to controls. One of these DMRs was located in the 3'UTR of the *Gsx1* gene (mm10 coordinates, meanDiff = 0.001121656, qval = 0.042, **Figure 2J**). We also performed RNA-Seq after *Kmt2a* KO and found 2105 DEGs compared to controls (**Supplementary table 6**). Among the previously identified DMRs, only one had significant changes in gene expression, *Gsx1* (**Figure 2K**). GSX1 is a TF expressed early in NPCs and is known to promote maturation in specific neuronal subtypes (Coltogirone et al., 2023; Pei et al., 2011). Notably, a prior study reported reduced H3K4me1 near this DMR in *Kmt2a* knockdown neurons derived from a mouse model (Kerimoglu et al., 2017) (**Supplementary Figure 2E**), supporting that *Gsx1* may be a direct target of KMT2A in mNPCs. We then asked if GSX1 targets were differentially expressed in *Kmt2a*-KOs. Using the TFBS database, which predicts TF-target gene interactions from DNA motifs and ENCODE DNase I footprints (Plaisier et al., 2016), we identified 133 GSX1 target

genes (19.2% of those expressed) that overlapped with *Kmt2a*-KO DEGs. However, this overlap was not significantly enriched (Fisher's exact test, OR = 1.18 and p = 0.055). To further explore whether *Gsx1* regulation might also be influenced by other EM genes, we next examined the *Dnmt1*-KO. While we identified two intergenic DMRs near *Gsx1* in *Dnmt1*-KO (**Supplementary Figure 2F-G**), *Gsx1* did not exhibit corresponding transcriptional changes in *Dnmt1*-KO. Interestingly, its downstream targets were significantly enriched amongst *Dnmt1*-KO DEGs (Fisher's exact test, OR=1.16, p=0.044). Finally, unlike what was found for the *Dnmt1*-KOs, we found no enrichment of imprinting genes in *Kmt2a*-KO (Fisher's exact test, OR = 1.23, p=0.295). Thus, although imprinting is regulated through chromatin state, the effect is likely EM-gene specific.

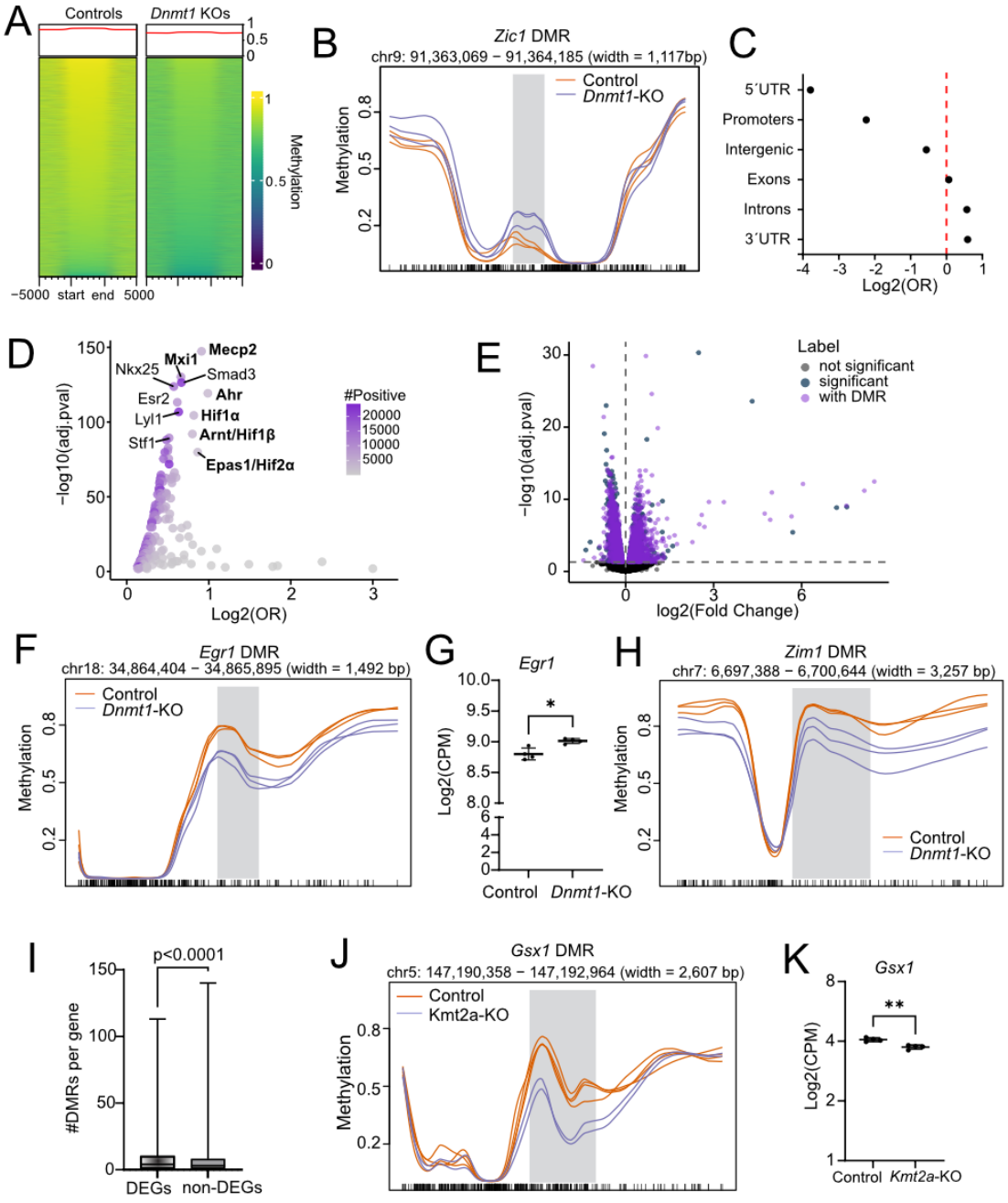


Figure 2: Extent and impact of DNA methylation changes in *Dnmt1*-KO and *Kmt2a*-KO mNPCs. (A) Enrichment heatmaps of smoothed CpG methylation across 141,722 *Dnmt1*-associated DMRs. Each horizontal line represents the average smoothed methylation profile across CpG sites within a DMR and ±5kb flanking regions. Yellow shows higher smoothed methylation but blue shows lower smoothed methylation. (B) The *Zic1* DMR (qval < 0.05, gray vertical bar) in *Dnmt1*-KOs (purple) compared to control NPCs (orange). Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome. (C) The enrichment analysis of the DMR distribution in *Dnmt1*-KOs, using Fisher's exact test. All categories are significantly enriched or depleted ($p < 0.05$). The exact value is listed in **Supplementary Table 3**. (D) Scatter plot showing enriched TF motifs in *Dnmt1*-KO DMRs (padj < 0.05). (E) A volcano plot showing the distribution of *Dnmt1*-KO DEGs, with significant genes (padj<0.05) shown in dark blue, and DEGs with DMRs associated shown in purple. (F) *Egr1* DMR (qval < 0.05, gray vertical bar) in *Dnmt1*-KOs (purple) compared to control NPCs (orange). Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome. (G) Significant upregulation of *Egr1* in *Dnmt1*-KO compared to controls, $p < 0.05$. (H) Three *Zim1* DMRs (qval < 0.05, gray vertical bar) in *Dnmt1*-KOs (purple) compared to control NPCs (orange). Flanking regions of middle DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome. (I) Plot showing mean DMR numbers per gene with error bars representing standard error. Wilcoxon-rank test shows statistical significance, $p<0.0001$. (J) *Gsx1* DMR (qval < 0.05, gray vertical bar) found in *Kmt2a*-KO NPCs (orange) compared to control NPCs (purple). Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome. (K) RNA sequencing results showing normalized *Gsx1* counts in *Kmt2a*-KOs and controls. *p<0.05, **p<0.01, ns=not significant.

Substantial convergence of gene expression changes after *Dnmt1* and *Kmt2a* disruption

Given the phenotypic overlap between multiple MDEMs, we next investigated the overlap in molecular consequences between *Kmt2a*-KOs and *Dnmt1*-KOs. When we compared the DEGs from both *Dnmt1*-KO and *Kmt2a*-KO cells, we found a substantial overlap of 1543 DEGs, including 679 upregulated and 861 downregulated genes (Fisher's exact test, OR=11.6, $p < 0.0001$, **Figure 3A, Supplementary Table 7**). Moreover, gene expression changes were highly correlated between the two datasets (Spearman's $\rho = 0.94$, $p<0.0001$, **Figure 3B**), and almost all DEGs were concordant in the direction of change. In both datasets, the upregulated genes were enriched in neuronal differentiation and maturation GO terms (**Figure 3C, Supplementary Figure 3A-B**), while the downregulated genes were enriched in GO terms linked to cell cycle and chromosome organization (**Figure 3D, Supplementary Figure C-D**). These patterns suggest that disruption of *Dnmt1* and *Kmt2a* may drive cells towards premature differentiation.

257 To determine whether *Kmt2a*-KO and *Dnmt1*-KO mNPCs undergo premature
258 differentiation, we explored markers for neuronal and glial differentiation in both KOs and
259 scrambled control mNPCs, using WB and IF staining. WB analysis revealed a significant
260 increase in the neuronal marker TUBB3 in *Kmt2a*-KO mNPCs compared to controls (**Figure**
261 **3E**), whereas *Dnmt1*-KO mNPCs did not exhibit a similar change at the protein level (**Figure**
262 **3E**). No notable differences in the proliferation marker PCNA were detected in either KO
263 model (**Supplementary Figure 3E**). Interestingly, IF staining demonstrated a greater number
264 of TUBB3-positive cells in both *Kmt2a*-KO and *Dnmt1*-KO cultures, compared to scrambled
265 controls, along with distinct morphological changes indicative of differentiation, such as
266 smaller and elongated cell-bodies (**Figure 3F-G**). Increased expression of the astrocyte
267 marker EAAT2 was also observed in *Dnmt1*-KO cultures, whereas no significant change was
268 detected in *Kmt2a*-KO cells (**Supplementary Figure 3F-G**). Prior work has shown
269 prematurely poised neuronal enhancers in primary mNPCs from a *Kmt2a*-haploinsufficient
270 model, a phenotype which normalized after genetic rescue of *Kmt2a* (Reynisdottir et al.,
271 2024). Additionally, NPCs from this model exhibit a similar premature neuronal
272 differentiation phenotype as in our full KO (**Supplementary Figure 3H-I**), supporting that
273 this is a feature of *Kmt2a* disruption. These findings support that loss of both factors
274 contributes to premature differentiation of mNPCs but they differ in impact. Specifically, loss
275 of *Kmt2a* leads to a robust and uniform neuronal differentiation phenotype, whereas *Dnmt1*
276 deficiency promotes both neuronal and glial differentiation, but potentially in a more variable
277 manner.

278 We next asked whether the shared DEGs were driven by common regulatory
279 programs, using motif enrichment analysis of their promoter regions. Analysis of Motif
280 Enrichment (AME; MEME-suite) of promoter regions of the shared DEGs identified common
281 TFs in the promoter regions, including members of the specificity protein (SP) and E2F
282 families, notably SP2 and E2F4, both of which are known regulators of cell cycle progression
283 and differentiation in neuronal cells (**Figure 3H, Supplementary Table 8**)(Liang et al., 2013;
284 Persengiev et al., 1999). We wondered if any of these TFs can interact with both KMT2A and
285 DNMT1, thereby possibly explaining the overlapping expression changes. By comparing
286 known interactors for either protein from BioGRID, that were expressed in our system, we
287 found a significant overlap of 53 interactors (Fisher's exact, OR = 18.5, $p<0.0001$, **Figure 3I**).
288 Five of these had enriched motifs in the shared DEGs, most notably SP1 and E2F6 (**Figure**
289 **3H, boxes**).

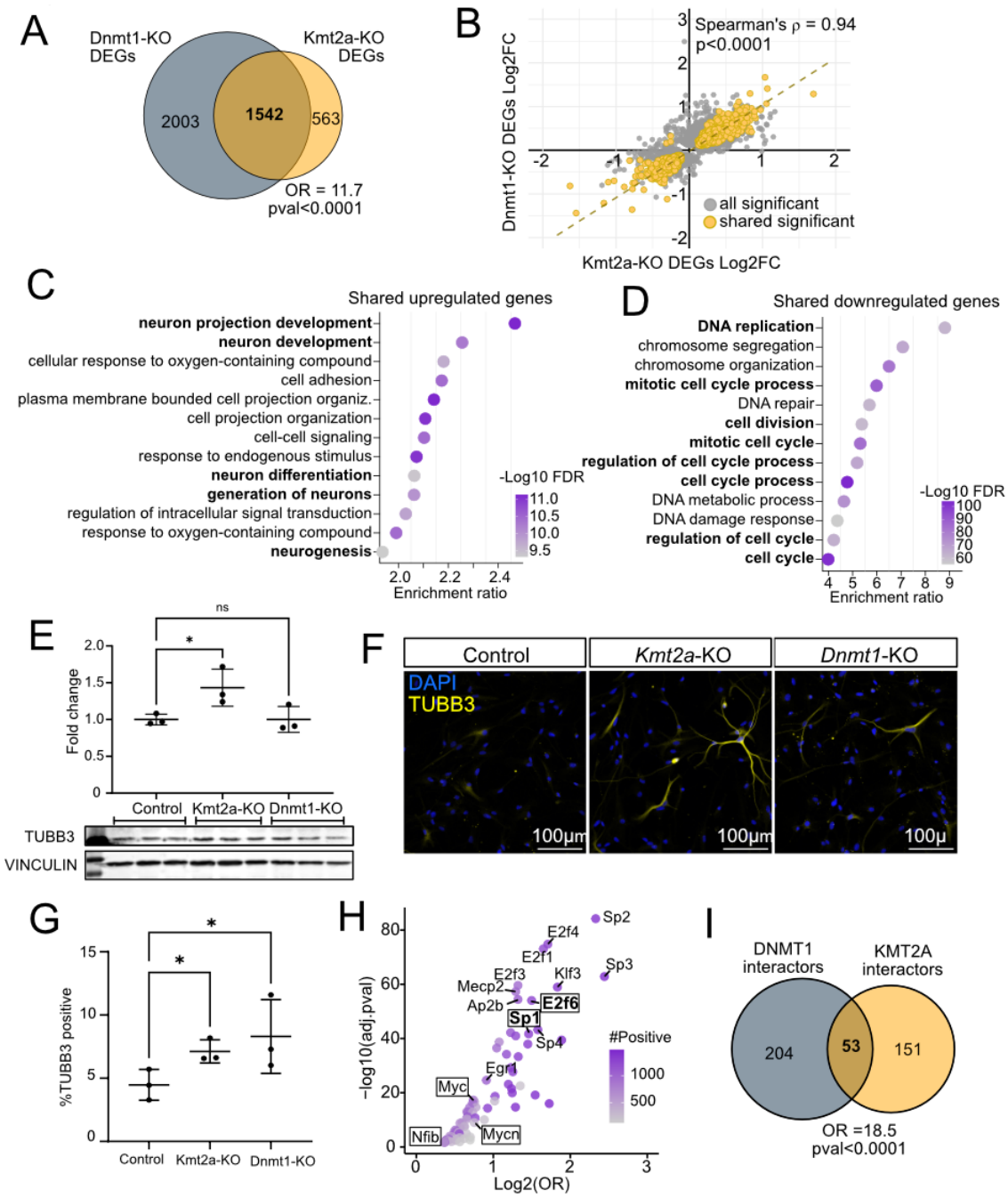


Figure 3: Premature differentiation in *Kmt2a*-KO and *Dnmt1*-KO. (A) Venn diagram showing extensive overlap of DEGs in *Dnmt1*-KO and *Kmt2a*-KO cells compared to control NPCs, with results from Fisher's exact test (OR=11.7, p<0.001). (B) Scatter plot showing Log2FC of significant DEGs from both *Dnmt1*-KOs and *Kmt2a*-KOs in gray, and the shared significant DEGs in yellow. Spearman's correlation was calculated for the shared DEGs (R = 0.94 p-val < 0.0001). (C) GO term analysis (Biological Process) of overlapping upregulated genes, showing enrichment ratio of each GO-term on the x-axis and color depicting the FDR. (D) GO term analysis (Biological Process) of overlapping downregulated genes showing enrichment ratio for each GO-term on the x-axis and color depicting the FDR. (E) WB analysis of TUBB3 protein after knocking out *Kmt2a* (padj=0.0474) and *Dnmt1* (padj>0.05, ns) with CRISPR/Cas9. A one-way ANOVA with Dunnett´s multiple comparisons test was used to calculate significance. (F) Immunofluorescence staining of mNPCs after *Kmt2a*-KO

303 and *Dnmt1*-KO compared with controls (blue = DAPI, yellow = TUBB3, magenta = EAAT2).
304 (G) Quantification of the percentage of TUBB3 positive cells in *Kmt2a*-KO, *Dnmt1*-KO and
305 controls. A one-way ANOVA with an FDR correction was used to calculate significance. (H)
306 Scatter plot of TFs ranked by the number of promoters containing their binding motifs. The
307 odds ratio indicates motif overrepresentation in promoters of shared DEGs. Boxes label
308 shared BioGrid interactors. (I) Venn diagram showing the overlap of expressed BioGrid
309 interactors between KMT2A and DNMT1, Fisher's exact test OR=18.5, p <0.0001.*p<0.05,
310 ns=not significant.

311

312 **Evaluation of changes that occur during normal differentiation**

313 Given the premature differentiation observed in EM KOs, some of the expression
314 changes may reflect differences in cell state composition across genotypes. To explore the
315 extent of this in our EM KOs, we differentiated wild-type (WT) mNPCs into mixed neuronal
316 cultures (see methods) and demonstrated that these cultures express typical markers for
317 neurons and astrocytes, as well as synaptic markers at Day 6 (**Supplementary Figure 4A**).
318 We then performed ONT long-read RNA-seq on mNPCs, Day 2, and Day 6 differentiating
319 cells (schematic in **Figure 4A**), which yielded 3706 and 3441 DEGs at Day 2 and Day 6,
320 respectively, compared with mNPCs (**Supplementary Table 9-10, Supplementary Figure**
321 **4B**). Both developmental timepoints were enriched in neurogenic and synaptic terms,
322 supporting neural differentiation and maturity of the culture (**Supplementary Figure 4C-D**).
323 Additionally, a gene-set enrichment analysis of cell-types showed significant enrichment in
324 cell-type markers for the three main brain-cell lineages (neurons, astrocytes and
325 oligodendrocytes, **Supplementary Figure 4E**) as would be expected.

326 To understand the extent of transcriptional changes that may be secondary to
327 premature differentiation in *Dnmt1*-KOs and *Kmt2a*-KOs, we compared DEGs from each KO
328 to those identified during normal neuronal differentiation (mNPCs vs. Day 2 and Day 6).
329 Interestingly, DEGs from both KOs showed strong and consistent overlap with normal
330 differentiation associated gene expression changes (ORs > 2 in all comparisons, $p < 0.0001$,
331 Fisher's exact test; **Figure 4B**) and displayed a linear correlation (Spearman's $\rho > 0.45$ for all
332 comparisons, $p < 0.0001$; **Figure 4C-F**). These results suggest substantial convergence
333 between KO-induced and differentiation-related transcriptional programs, consistent with a
334 premature differentiation phenotype in EM-KOs.

335 To map the temporal landscape of DNAm during neuronal differentiation, we profiled
336 genome-wide DNAm using ONT in mNPCs at Day 0, Day 2, and Day 6 of differentiation

337 (schematic in **Figure 4A**). A principal component analysis (PCA) of genome-wide DNAm
338 showed that the developmental stage was the main driver of methylation variance, as captured
339 by PC1 (**Supplementary Figure 4F**). While differential methylation analysis identified no
340 significant changes at Day 2 compared to Day 0 mNPCs, it revealed 8,627 genome-wide
341 significant DMRs (qval < 0.05) at Day 6 compared to Day 0 mNPCs, indicating that most
342 DNAm remodeling occurs relatively late in differentiation (**Figure 4G, Supplementary**
343 **Table 11**). Genome-context analysis showed that, similar to *Dnmt1*-KO DMRs, on Day 6
344 DMRs were enriched in introns, 3' UTRs, and exons, but they were depleted in promoters, 5'
345 UTRs, and intergenic regions (**Supplementary Figure 4G**). To assess whether the
346 methylation dynamics observed during normal differentiation are recapitulated in the KOs, we
347 examined the methylation state of the Day 6 DMR regions in *Dnmt1*-KO and *Kmt2a*-KO
348 cells. *Kmt2a*-KO cells maintained a methylation profile similar to undifferentiated mNPCs at
349 these loci, while *Dnmt1*-KO cells exhibited a methylation pattern resembling that of Day 2
350 neurons (**Figure 4H**). However, methylation levels adjacent to these Day 6 DMRs were also
351 lower in *Dnmt1*-KO cells, suggesting that the effects are not specific to those regions but
352 rather a global effect. Additionally, 1,154 of the Day 6 DMRs were also identified in *Dnmt1*-
353 KO cells (OR = 1.48, p < 2.2e-16, **Supplementary Table 12**). We conclude that the
354 methylation changes over these Day 6 DMRs are driven by global hypomethylation in *Dnmt1*-
355 KO cells but that they are also consistent with aspects of premature differentiation. Taken
356 together, these results suggest that transcriptional changes precede widespread DNAm
357 remodeling, supporting a model in which DNAm acts as a secondary layer that consolidates,
358 rather than initiates, neurodevelopmental programs.

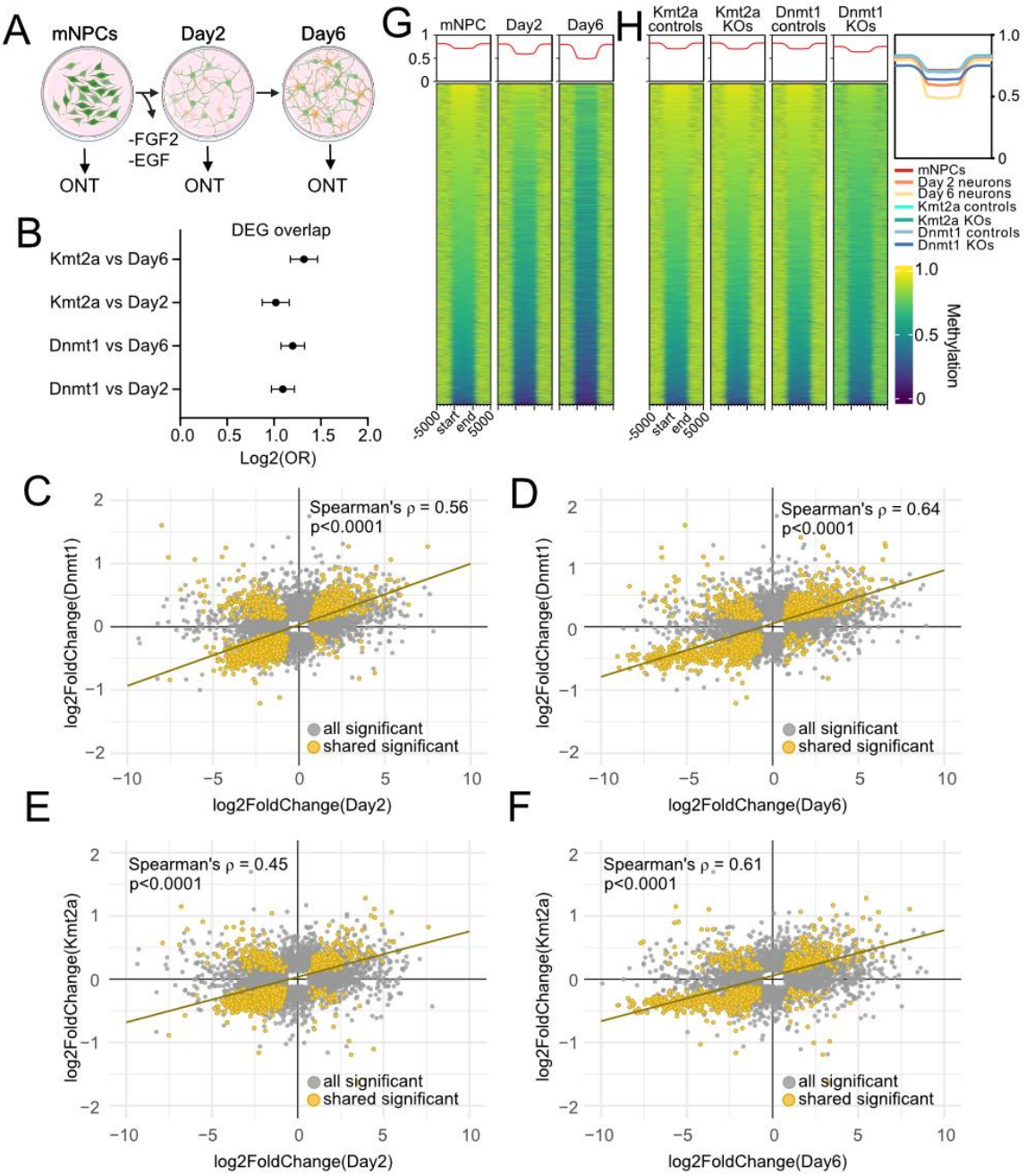


Figure 4: Changes observed in DNAm and gene expression during differentiation of NPCs. (A) A schematic of the neurodevelopmental model. (B) Results of Fisher's exact test shown for the overlap between *Kmt2a*-KO and *Dnmt1*-KO DEGs with NPC-Day 2 and NPC-Day 6 DEGs, with 95% confidence intervals shown. (C-F) Significant Spearman correlations between DEGs of (C) *Dnmt1*-KO vs Day 2, (D) *Dnmt1*-KO vs Day 6, (E) *Kmt2a*-KO vs Day 2, (F) *Kmt2a*-KO vs Day 6. Axis represents log2(Fold change) for each dataset. (G) Day 6 DMRs are shown in mNPCs, Day 2 and Day 6 neurons. (H) Day 6 DMRs in KO and control samples. In the enriched heatmaps in (G) and (H), each line demonstrates the average smoothed methylation values across all CpGs in corresponding the DMR. DMRs were aligned at the DMR start site (start) and end site (end), ±5kb on either side. Yellow shows higher smoothed methylation but blue shows lower smoothed methylation values.

372 **An expanded cohort of EM KOs shows subtle DNAm changes**

373 We next broadened our analysis to assess DNAm abnormalities across the wider EM.
374 Using ONT long-read whole genome sequencing, we profiled 44 additional EM-KO lines
375 alongside matched scrambled controls. These data were pooled with previously generated
376 data from *Dnmt1*-KO and *Kmt2a*-KO, resulting in a total of 46 EM-KOs and 16 control
377 samples. To systematically evaluate genome-wide DNAm changes, we partitioned the mouse
378 genome into ~2.7 million 1 kb tiles (each containing >2 CpGs) and compared methylation
379 levels across groups. A PCA plot of these tiles revealed that two known DNAm machinery
380 components, *Dnmt1* and *Dnmt3b*, showed the most dramatic DNAm abnormalities
381 (**Supplementary Figure 5A**). When these two were excluded from the analysis, *Kat6a*, *Chd5*,
382 and *Tet1* emerged as outliers, clustering separately from the remaining EM-KOs
383 (**Supplementary Figure 5B**). Despite this variation, no shared DMRs were identified across
384 the 46 EM-KOs.

385 **Promoter DNAm alterations depend on interactions with the DNAm machinery**

386 Given the diverse roles of EM genes in modulating DNAm, we hypothesized that
387 shared DMRs may exist within specific subgroups of EM-KOs with related molecular
388 functions. Therefore, classifying EM genes based on regional DNAm patterns could help
389 uncover convergence. Promoter regions are well-established regulatory elements that are
390 easily assigned to the corresponding genes and involved in transcriptional control (Deaton &
391 Bird, 2011). To investigate whether a subset of EM-KOs exhibit similar DNAm changes at
392 gene promoters, we applied non-negative matrix factorization (NMF) to cluster the 46 EM-
393 KOs based on DNAm patterns at the promoters of 20,482 mouse protein-coding genes
394 (**Supplementary Figure 5C)**. Hierarchical clustering of EM-KOs using the NMF-derived
395 coefficients of latent factors further characterized three latent factors (**Figure 5A**).
396 Specifically, cluster 1 included *Chd3, Setd4, Setd1a, Setd1b, Kdm5b*, and *Kdm1a*. In contrast,
397 cluster 2 included the controls and showed low coefficients in all three latent factors. Finally,
398 cluster 3 included the genes *Kmt2a*, *Dnmt1*, *Chd1*, *Kdm6a*, *Hdac6*, *Ezh2*, and *Crebbp* and
399 showed the highest coefficient in latent factor 2 (**Figure 5A**). *Dnmt3b*-KO showed the highest
400 coefficient in latent factor 1 and did not cluster with other EM genes (**Figure 5A**).
401 Interestingly, genes with the highest weight in latent factor 2 were significantly enriched in
402 GO terms related to generation of neurons and neurodifferentiation (**Figure 5B**) as well as the
403 Wnt signaling pathway (**Supplementary Figure 5D**). In contrast, no significant enrichment in
404 GO terms was found for the other latent factor groups. Interestingly, in the two clusters that

did not group with controls (clusters 1 and 3), nearly all the gene members (excluding SETD4) had STRING defined interactions with the maintenance methyltransferase DNMT1 (**Figure 5A**). Indeed, components of these clusters had a significant enrichment in interactions with all the enzymatic DNAm machinery components (**Figure 5A,C**). The enrichment was strongest for the maintenance methyltransferase (DNMT1) but less so with *de novo* methyltransferases (**Figure 5C**). Furthermore, the same groups (Clusters 1&2 and DNMT3B), had a significantly higher number of STRING interactions with other members of the DNAm machinery (n=25) (Boukas et al., 2019) compared to what was seen in cluster 2 (**Figure 5D**, one-sided Wilcoxon rank-sum test, W = 349.5, p-value = 0.0014). These results suggest that EM factors outside the core DNAm machinery may influence promoter DNAm through interactions with components of the DNAm machinery.

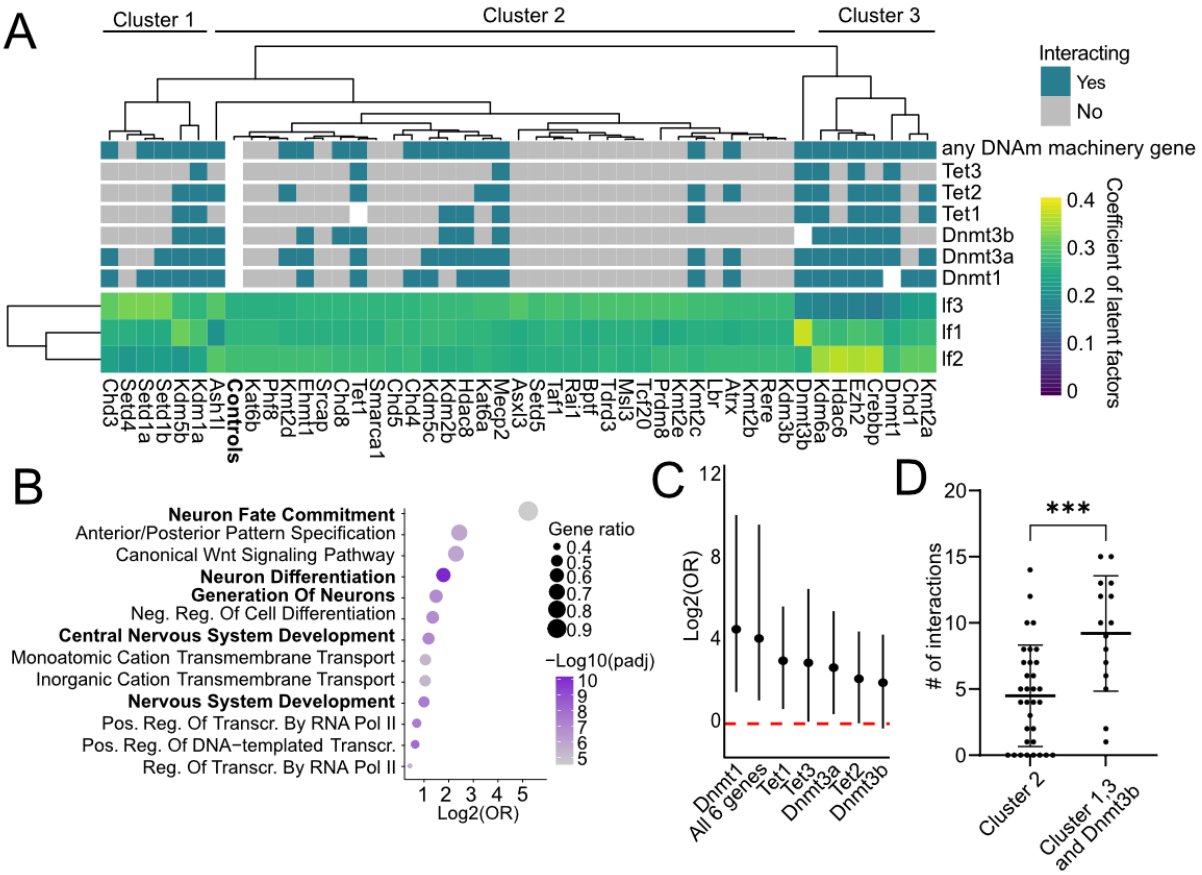


Figure 5: DNAm analysis of 46 EM KO genes in NPCs reveals shared promoter effects. (A) A composite graph showing STRING interactions of EM components with enzymatic DNAm components on top, the heatmap showing coefficient of latent factors for each EM-KO across three latent factors. Color scale represents coefficient of each latent factor (yellow shows higher latent factor coefficient but blue shows lower latent factor coefficient). (B) Gene ontology (GO) term enrichment (Biological Process) of the genes associated with latent factor 2. Size represents the ratio of genes in the relevant GO term, color indicates the -log10(adjusted pvalue). (C) Forest plot showing enrichment of enzymatic DNAm machinery-

426 interacting EMs in two relevant clusters compared to rest. Significance testing was done with
427 a Fisher's exact test. Dots represent Fisher's exact test OR (log_2 scale), with lines indicating
428 95% confidence intervals. The horizontal dotted red line represents statistical significance.
429 (D) Comparison of number of STRING interactions with 25 DNAm machinery components
430 between the clusters. A Wilcoxon rank sum test was used for significance. lf1 = latent factor
431 1, lf2 = latent factor 2, lf3 = latent factor 3. ***<0.001, ns=not significant.

432

433 **Haplotype-specific DNAm within the context of EM KOs**

434 We used mNPCs derived from F1 offspring of the two strains (FVB maternal × B6J
435 paternal, **Figure 1A**). Because the C57BL/6J (B6J) and FVB mouse strains differ genetically
436 across roughly half the genome, it is possible to phase this part of the genome and perform an
437 allele-specific methylation analysis. PCA of the phased data shows the same pattern between
438 KOs and controls as for the unphased data (**Supplementary Figure 6A**) but supports the
439 presence of haplotype-specific methylation between the two strains.

440 Using the control samples only, we found 1074 haplotype-specific DMRs (FWER <
441 0.05) (**Figure 6A, Supplementary Table 13**) (Do et al., 2016; Tycko, 2010), including one
442 located in the exon 4 of *Tet3* (**Figure 6B**). Because of the consistent parental origin of the two
443 genomes, these DMRs should include imprinted loci. Indeed, our DMRs include 21 out of 57
444 known mouse imprinted loci (**Supplementary Table 14**) (Gigante et al., 2019; Xie et al.,
445 2012); note that imprinting is partly cell type specific. We next examined the genetic variation
446 at these loci. As expected, we found that 95% of the haplotype-specific DMRs (n = 1018)
447 contained haplotype-specific genetic variants, including insertions/deletions (indels), SNPs
448 and CpG changes (**Figure 6C**). This variation appeared enriched in exons, 5' UTRs, 3' UTRs,
449 and promoters but depleted in intergenic regions (**Supplementary Figure 6B**). Within these
450 DMRs we found 11285 single nucleotide polymorphisms (SNPs), 7836 indels, 70 insertions
451 and 64 breakpoints (i.e. 17.92 variants per DMR). In contrast, only about half of the ICRs
452 defined as DMRs (52,35%, n= 11) harbored genetic variants and these regions carried fewer
453 variants on average (7.57 variants per DMR). This pattern is consistent with imprinted regions
454 being primarily regulated epigenetically, whereas non-imprinted haplotype-specific DMRs are
455 more often mediated by cis-acting regulatory polymorphisms (Tycko, 2010).

456 To determine if the EM-clusters, identified via the NMF analysis (**Figure 5A**) yielded
457 any shared DMRs we performed a grouped analysis on the phased data for each of these
458 clusters, compared to the controls. Cluster 1 (*Chd3*, *Setd4*, *Setd1a*, *Setd1b*, *Kdm5b*, and
459 *Kdm1a*) showed a shared, FVB-specific, hypomethylated DMR, 4.6 kB upstream of the gene

460 *Ebf1* (FWER=0.03, **Figure 6D**), which encodes a transcription factor involved in early
461 neurogenesis (Garel et al., 1999; La Manno et al., 2021). In contrast, cluster 2 and 3 did not
462 yield any shared DMRs, compared to controls. Finally, we searched for DMRs associated
463 with our entire group of 46 EM KOs, separately on each haplotype. We found a shared
464 hypomethylated DMR on the C57BL/6J haplotype (FWER<0.05) and none on FVB. The
465 DMR is located in the 3' UTR of *Zic4* (**Figure 6E, Supplementary Figure 6C**), a TF that
466 plays a role in cerebellar development and malformations (Grinberg et al., 2004). Although
467 *Zic4* was not differentially expressed in *Dnmt1*-KO nor *Kmt2a*-KO mNPCs compared to
468 scrambled controls (**Supplementary Figure 6D-E**), it was significantly downregulated
469 (log2FC = -1.978, adjusted p-val = 0.00197) at Day 6 compared to Day 0 mNPCs in our
470 neurodevelopmental model (**Supplementary Figure 6F**). Similarly, prior studies on Sotos
471 syndrome (Choufani et al., 2015) have reported DNAm changes in the 3' UTR of *ZIC4* (CpG
472 cg26386385), suggesting this region is sensitive to epigenetic regulation across multiple EM
473 factors. To evaluate the functional consequences of *Zic4*, we overexpressed Myc-tagged *Zic4*
474 in NPCs and compared with Myc-overexpression alone. The results showed reduced
475 expression of neuronal maturation markers (**Supplementary Figure 6G-H**), implicating ZIC4
476 may maintain the NPC state and raising the possibility that it may act as a brake on neuronal
477 maturation.

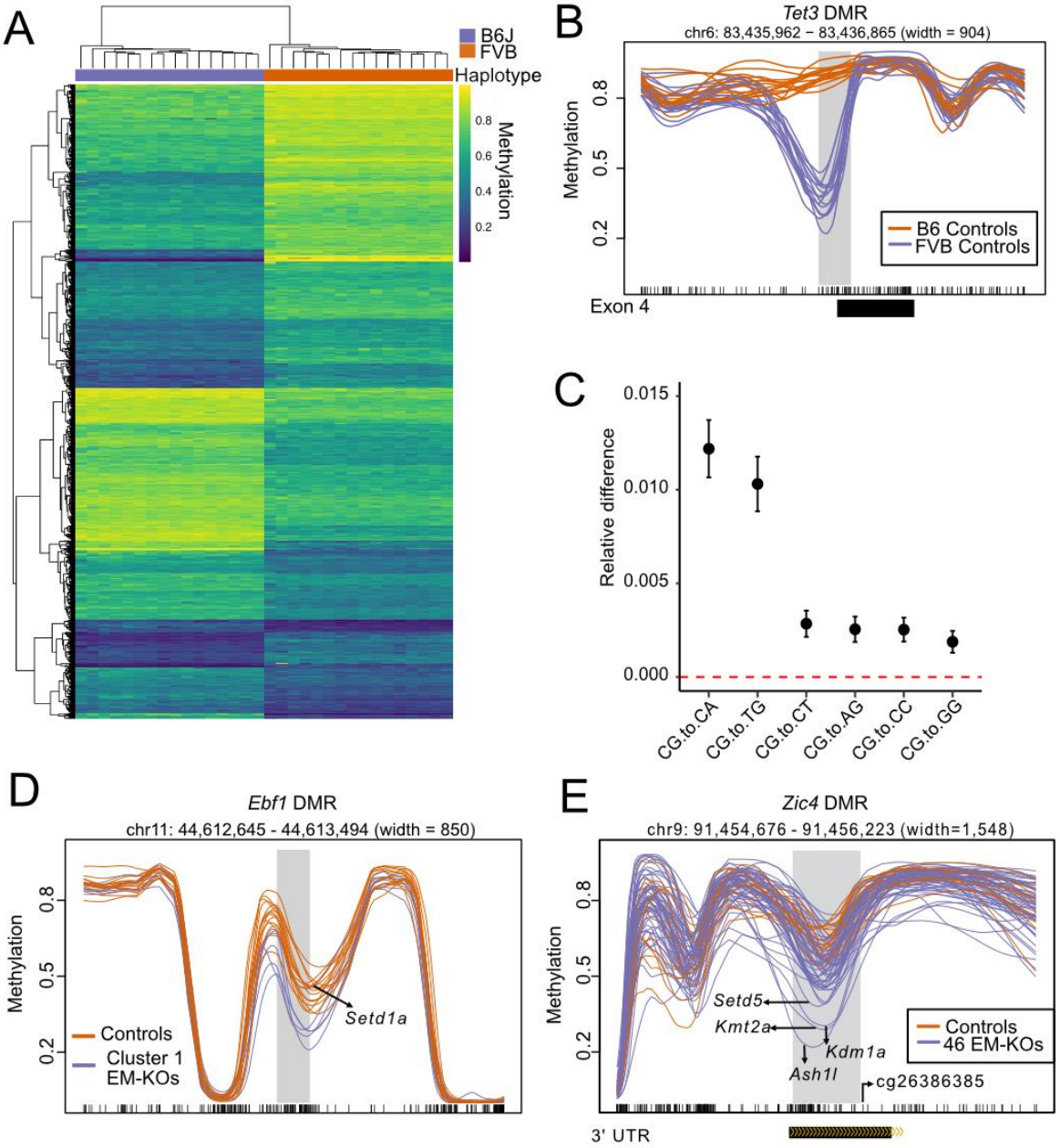


Figure 6: Haplotype-specific changes in DNAm. (A) A heatmap plot showing smoothed methylation of 1074 haplotype-specific DMRs across B6J and FVB haplotypes. Yellow and blue represent higher and lower DNAm, respectively. Each line in y-axis represents a single DMR across haplotyped controls but x-axis represents haplotyped controls. B) A haplotype-specific DMR in the *Tet3* gene. C) An enrichment of CpG dinucleotide mutations within the haplotype-specific DMRs compared to rest of genome. D) A line plot showing a DMR in 5' UTR of *Ebf1* between FVB haplotyped control mNPC (orange) and FVB haplotyped Cluster 1 EM-KOs (purple). The DMR is highlighted in gray. Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome. T (E) A line plot showing a DMR in 3' UTR of *Zic4* between B6J haplotyped control mNPC (orange) and B6J haplotyped 46 EM KOs (purple). The DMR is highlighted in gray. Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome.

491 **Discussion**

492 Our study presents a comprehensive investigation into how EM disruption affects
493 DNAm and gene expression in a disease-relevant neuronal system, offering insights into the
494 presence and functional impact of DNAm changes beyond peripheral blood. By combining a
495 mNPC model with ONT sequencing, haplotype phasing, and transcriptomic profiling, we
496 characterized the functional consequences of 46 EM gene KOs. Disruption of *Dnmt1* leads, as
497 expected, to extensive changes in the DNAm. A PCA of genome-wide DNAm suggests that
498 the second largest effect on DNAm is from *Dnmt3b* and, interestingly, *Kat6a*, although more
499 work will be needed to map these changes genome wide. Across the 46 EM genes, we
500 observed little-to-no shared DNAm changes. This could be because of the relatively short
501 time lapse between KO and methylation profiling in our work. Our data suggests that TFs
502 such as GSX1, HIF, SP and the E2F family transcription factors may mediate or interpret the
503 EM-related epigenetic disruption. For instance, *Gsx1* was differentially expressed in *Kmt2a*-
504 deficient cells, and its target genes were enriched among DEGs in both *Kmt2a*-KOs and
505 *Dnmt1*-KOs. Similarly, SP and E2F family TF binding motifs were enriched in promoters of
506 shared DEGs while SP1 and E2F6 specifically are known binding partners of both KMT2A
507 and DNMT1. These findings point to a model in which the EM indirectly modulates
508 transcriptional networks through TFs, either by altering DNAm at TF binding sites or via
509 interactions with the TFs themselves. Importantly, some TFs such as EGR1 are known to
510 recruit TET1, further linking transcriptional networks to active demethylation pathways (Sun
511 et al., 2019). Similarly, our clustering of promoter DNAm reveals that at least in some EM
512 contexts, the physical interaction with the DNAm machinery plays a role.

513 Interestingly, we uncovered a hypomethylated DMR in the 3' UTR of *Zic4* which was
514 shared across all 46 EM-KOs in one of the two haplotypes. This gene is implicated both
515 murine and human neurodevelopment and DNAm changes in the gene has been identified in
516 peripheral blood from individuals with Sotos syndrome, a neurodevelopmental disorder
517 caused by mutations in the epigenetic regulator *NSD1* (Choufani et al., 2015). However, the
518 methylation changes are modest, and this gene is not differentially expressed in our system.
519 Nevertheless, *Zic4* is downregulated during differentiation and overexpression of *Zic4*
520 reduced expression of markers for neuronal and glial differentiation (**Supplementary Figure**
521 **5G-H**) suggesting that it may play a role in maintaining the NPC state. It is not clear whether
522 this gene is important in explaining the phenotypic overlap in the MDEM group, but it could
523 have potential utility as a biomarker.

We have established a premature differentiation phenotype arising as a result of KO of either *Dnmt1* or *Kmt2a*. This is something we and others have observed as a consequence of disrupting several other EM factors. For example, a recent study of KMT2A-deficient mice revealed a H3K4me1 enrichment at neuronal enhancers in undifferentiated NPCs (Reynisdottir et al., 2024), suggesting that these enhancers show premature poising. Additionally, in prior work on Kabuki syndrome, another MDEM caused by mutations in *KMT2D*, where the same phenotype has been reported in two disease relevant contexts (Carosso et al., 2019; Fahrner et al., 2019). Moreover, studies of Kabuki syndrome neurons from a mouse model have shown an overrepresentation of aging-related loci (Boukas et al., 2024), further suggesting that premature differentiation may represent a unifying phenotype across a subset of MDEMs. Interestingly, in our NMF analysis, KMT2A and DNMT1 clustered with EZH2, KDM6A and CREBBP, all of which have been shown to regulate neuronal stem cell maintenance and differentiation (Lei & Jiao, 2018; Pereira et al., 2010; Wang et al., 2010), and genes associated with the latent factor of this group were enriched in genes relating to neuronal differentiation (**Figure 5B**). This could suggest that even though the DNAm changes are very small, the disruption of these EM factors affects a core NPC gene network.

Despite the modest DNAm changes, we observed an impressive convergence of transcriptional dysregulation between DNMT1 and KMT2A, along with the aforementioned shared cellular phenotype of premature differentiation. Comparing the transcriptional changes in these two EM-KOs to transcriptional changes during normal neuronal differentiation, we conclude that part of the overlap is a consequence of the premature differentiation. Nonetheless, we note that the overlap between *Dnmt1*-KO and *Kmt2a*-KO cells appears stronger than either of the overlaps with normal differentiation. Combined with the lack of methylation changes in *Kmt2a-KOs*, we have shown that the expression changes are not a consequence of DNA methylation changes and are more likely to drive this particular phenotype. It will be interesting and important to understand how the shared expression changes arise, and especially to identify which epigenetic changes, if any, are causal. In any case, differentiation state should be carefully considered when analyzing gene expression, since premature or altered differentiation can be a major source of the observed changes.

Beyond providing mechanistic insight, this work establishes a framework for using long-read sequencing, haplotype phasing, and transcriptomic profiling in disease-relevant cells to dissect how epigenetic regulators act in health and disease.

557 **Materials and methods**

558 **Animals**

559 B6J.129(Cg)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J(Strain#026179) (B6J) and
560 FVB.129(B6)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J(Strain#026558) (FVB) mice,
561 each carrying a heterozygous Cas9 transgene, were obtained from Jackson laboratory. Mice
562 were bred within their respective background strains (B6J and FVB) to generate homozygous
563 Cas9-expressing offspring. DNA was isolated from ear clips, and Polymerase Chain Reaction
564 (PCR) was used to confirm the Cas9 cassette using primers design provided by the Jackson
565 laboratory (**Supplementary Table 15**). All mice were kept on a 12 h dark-light cycle at room
566 temperature (21-23°C) with relative humidity around 40% and maintained with ad libitum
567 diet and water. Mouse importation and experimental protocols were in accordance with
568 Icelandic Food and Veterinary Authority (license no. 2208602) and accepted by National
569 Expert Advisory Board on Animal Welfare.

570 **Dissections and primary NPCs isolation**

571 The NPCs were isolated from an F1 hybrids of homozygous Cas9 mice (B6J and FVB). The
572 isolation procedure was based on published protocols (Bernas et al., 2017; Rybachuk et al.,
573 2019) with minor modifications, see Supplementary Materials for details. NPCs were grown
574 in neuronal growth medium (NGM), which included Neurobasal medium supplemented with
575 1% B27 supplement (ThermoFisher, 17504044), 1% Glutamax (ThermoFisher, 35050061),
576 0.25% Penicillin-streptomycin (ThermoFisher, 15140122) 20 ng/ml EGF (PeproTech, AF-
577 100-15), 20 ng/ml FGF2 (PeproTech, 100-18B) and 2 µg/ml Heparin (MP Biomedicals,
578 0210193125) on PDL/Laminin coated 6-well plates, and frozen down in NGM + 10% DMSO
579 (Santa Cruz, sc-358801).

580 **Neuronal progenitor differentiation**

581 NPCs (passage 3-4) were differentiated into mixed neuronal cultures by collecting them as
582 previously described, counting and seeding 150.000 cells/well on a PDL/Laminin coated 24-
583 well plate in 0.5ml N2-Neuronal Induction Medium (N2NIM) consisting of Neurobasal Plus
584 Medium (ThermoFisher, A3582901) with 1x B-27 Plus Supplement (ThermoFisher,
585 A3582801), 1x CultureOne Supplement (ThermoFisher, A3320201), 1x N2 Supplement
586 (ThermoFisher, 17502048), 0.25x Penicillin-streptomycin and 0.25x GlutaMax. This was
587 supplemented with 20ng/ml FGF2 and 10ng/ml BDNF (Peprotech, 450-02). On day 2 and 4,

588 medium was changed to N2NIM supplemented with 10ng/ml FGF2, 10ng/ml BDNF and
589 10ng/ml GDNF (ORF Genetics, IK0900). At Day2 and Day6 of differentiation, neuronal
590 cultures were collected and DNA and RNA isolated as described below.

591 **Guide RNA design and isolation of plasmids**

592 CRISPR/Cas9 guide RNAs (sgRNAs) were designed with an online tool
593 (https://chopchop.cbu.uib.no/) (Labun et al., 2019). Two sgRNAs, with the highest predicted
594 efficiency and minimal mismatches per gene, were selected to ensure out-of-frame upon
595 Cas9-mediated cleavage. All sgRNAs were verified with BLAT to ensure that none target any
596 other sequences in either of the two genomes. Eight scrambled control sgRNA pairs (random
597 sequence not binding in the murine genome) were designed for controls. Lentiviral CRISPR
598 vectors (Replogle et al., 2020), each including one pair of sgRNAs (**Supplementary Table 1**)
599 were synthesized by Vectorbuilder. Bacterial stabs were streaked onto LB plates and plasmids
600 isolated using ZymoPURE II plasmid Midiprep kit (Zymo, D4201) and stored at -20°C until
601 needed. All 54 dual guide-RNA lentiviral plasmids were Sanger sequenced to confirm correct
602 sgRNA pairs after amplification.

603 **Cell culture**

604 For virus production, HEK293T cells were cultured in culturing media, DMEM/F12
605 supplemented with GlutaMAX™ (ThermoFisher, 31331093) and 10% Fetal Bovine Serum
606 (FBS) (ThermoFisher, 10270106). All cells were incubated at 37°C with 5% CO_2. HEK293T
607 cells were frozen down in culture media supplemented with 10% DMSO.

608 **Virus production and transduction on NPCs**

609 HEK293T cells in a T75 flask were transfected to produce lentivirus. The transfection mixture
610 contained 2.29 µg lentiviral CRISPR plasmid, 2.08 µg psPAX2 (Addgene, 12260), 0.63 µg
611 pMD2.G (Addgene, 12259), combined with 15 µl Lipofectamine 3000 and 10 µl P3000
612 (ThermoFisher, L3000015), and brought to a final volume of 250 µl with Opti-MEM™
613 (ThermoFisher, 31985047). The transfected HEK293T cells were cultured for 72 hours, and
614 viral particles collected from the medium after 48 and 72hours, passing the supernatant
615 through a 0.45 µm syringe filter (Santa Cruz, sc-358814) to remove cell debris. After the
616 second collection, 4× Lentivirus concentrator solution was prepared following the protocol
617 from the MD Anderson Cancer Center Functional Genomics Core, and was used to precipitate
618 viral particles (Kutner et al., 2009; Lo & Yee, 2007). Fifty µl viral aliquots were stored in -

619 80°C until use. Transduction efficiency for each virus was tested in NPCs at several
620 concentrations, using serial dilutions of the virus together with 3.2ug/ml polybrene (Santa
621 Cruz, sc-134220). Transduction efficiency was determined using fluorescence-activated cell
622 sorting (FACS) of BFP fluorescent cells in NPC cells, 48 hours after transduction, and an
623 efficiency of ~30% was used to conduct all experiments.

624 **EM KO in NPCs and validation**

625 For EM KO experiments, the NPCs were seeded on a 6-well plate 24 h prior to transduction.
626 Cells were transduced with lentivirus in the presence of 3.2 µg/ml polybrene. Each lentivirus
627 targeted a single EM gene and non-targeting controls were used as negative controls for viral
628 infection and culturing effects. After 24 h, cells were split and selected by 0.5 µg/ml
629 puromycin (ThermoFisher, A1113802) for 2 days. Cells surviving puromycin treatment were
630 cultured for another day without puromycin to allow them to recover, then harvested for
631 experiments using Accutase as described above. Timing was selected based on the fact that
632 average protein half-life in mouse primary neurons is ~86 hours (Mathieson et al., 2018) and
633 for EM proteins it is considered even less (~40 hours), thus most of the EM proteins would be
634 expected to be degraded at the time of our evaluation. The KO efficiency was determined by
635 PCR, WB and immunofluorescence staining.

636 **RNA isolation, cDNA synthesis and Real Time Quantification PCR (RT-qPCR)**

637 Cultured cells were harvested and lysed in TRIzol™ Reagent (ThermoFisher, 15596018).
638 Total RNA was isolated using Direct-zol™ RNA Microprep Kit (Zymo Research, R2062)
639 based on manufacturer's protocol. The concentration and quality of RNA was evaluated with
640 Nanodrop (ThermoFisher). One µg total RNA was used for cDNA synthesis with High-
641 Capacity cDNA Reverse Transcription kit (ThermoFisher, 4368814). RT-qPCR reactions
642 were performed using Luna® Universal qPCR Master Mix (NEB, M3003) and a CFX384
643 Touch Real-Time PCR Detection System (BIO-RAD), according to manufacturer's protocol.
644 Any primers used are provided in **Supplementary Table 15.**

645 **Protein isolation and WB**

646 NPCs were collected using Accutase and counted with a Countess II FL Automated Cell
647 Counter (ThermoFisher). Cells were lysed at 15000-20000 cells/µl in 1x sample buffer
648 (60mM Tris HCl pH 6.8, 2% SDS, 10% Glycerol, 0.01% Bromophenol blue, 1.25% Beta-me)
649 supplemented with 1:100 Halt™ Protease Inhibitor Cocktail (ThermoFisher, 78437), 1:100

650 phosphatase inhibitor Cocktail (Cell signaling, 5870) and 1:100 PMSF Protease Inhibitor
651 (ThermoFisher, 36978). The lysate was heated at 95°C for 5 min and 1:100 Benzonase
652 nuclease (Sigma-Aldrich, sc-202391) added to degrade genomic DNA. The lysate was
653 aliquoted and stored in -80°C. Western blotting was performed as has previously been
654 described (Mahmood & Yang, 2012), where 15-25 µg of protein was loaded per sample,
655 blotted onto PVDF membrane (ThermoFisher, 88520) and stained overnight using the
656 following primary antibodies: rabbit anti-CHD1, 1:1000 (Cell signaling, 4351), mouse anti-
657 ACTIN, 1:1000 (Abcam, ab8224), rabbit anti-EAAT2, 1:1200 (Abcam, ab41621), mouse
658 anti-beta III Tubulin, 1:800 (Abcam, ab78078), rabbit anti-DNMT1, 1:1000 (Proteintech,
659 24206-1-AP), mouse anti-VINCULIN, 1:1000 (Sigma, V9131), rabbit anti-Myc-Tag, 1:1000
660 (Cell signaling, 2272). Mouse anti-PCNA, 1:7000 (Proteintech, 60097-1-Ig) was incubated at
661 RT for 1.5 h. Blots were stained with secondary antibodies: Donkey anti-Mouse IgG, 1:15000
662 (LICOR, 926-68072), Donkey anti-Rabbit IgG, 1:15000 (LICOR, 926-32213) for 1h at RT.
663 After washing, blots were visualized with Odyssey® CLx Imager and protein quantification
664 was evaluated by Image J (PMC3855844).

665 **Immunofluorescence**

666 NPCs were cultured on PDL/Laminin-coated 8-well culture slide (Falcon, 354118) for 24-48
667 hours, fixed with 4% paraformaldehyde (PFA), blocked with 5% NGS in 0.1% PBS-T and
668 stained using primary antibodies against rabbit anti-KMT2A, 1:200 (Cell signaling, 14197),
669 rabbit anti-EAAT2 1:400 (Abcam, ab41621) and mouse beta III Tubulin 1:200 (Abcam,
670 ab78078), overnight. The following day, cells were stained using secondary antibodies: Alexa
671 goat anti-rabbit 647, 1:1000 (ThermoFisher, A-21244) and Alexa goat anti-mouse 555,
672 1:1000 (ThermoFisher, A-32727). Cells were mounted with Fluoromount-G™ Mounting
673 Medium, with DAPI (ThermoFisher, 00-4959-52). Images were captured with Olympus
674 Confocal microscopy (Olympus Fluoroview, FV4000), processed with Fiji and quantified
675 with CellProfiler. Statistical analysis was done using GraphPad Prism v10.4.

676 **Overexpression**

677 NPCs were seeded at a density of 250,000-300,000 cells per well in PDL and Laminin coated
678 6-well plates, 24 hours prior to transfection. The following day, cells were co-transfected with
679 2.5 µg of either the Myc-tagged *Zic4* overexpression plasmid (Origene, MR216949) or the
680 corresponding Myc-alone plasmid backbone (control), along with 7.5 µl of Lipofectamine

681 3000, following the manufacturer's protocol. Cells were harvested 24 hours post-transfection
682 for downstream analysis.

683 **DNA sample isolation and preparation**

684 Genomic DNA isolated from cells using Monarch® HMW DNA Extraction Kit for Cells &
685 Blood (NEB, T3050) followed the manufacturers protocol for DNA extraction from cells. The
686 concentration and quality of DNA was measured by Nanodrop (ThermoFisher) and Qubit to
687 quantify the precise amount of double stranded DNA via Qubit dsDNA HS assay kit
688 (ThermoFisher, Q33231). High molecular weight (HMW) DNA was sheared to 20-30 kb to
689 achieve consistent size and to optimize ONT sequencing yield. HMW DNA was loaded to a
690 hydro-tube assembled with hydropore-syringe (Diagenode, E07010003), inserted into the
691 cassette, installed on a base unit and then sheared by Megaruptor® 3 (Diagenode) using the
692 following settings: a volume of 70 µl, a speed of 30, and a concentration of 50 ng/µl. The size
693 of sheared DNA was evaluated with HS Large Fragment 50 kb kit (Agilent, DNF-464-0500)
694 on a Fragment Analyzer, proceeded by High Sensitivity Large Fragment 50 kb Analysis
695 software.

696 **ONT-DNA sequencing library construction**

697 Library preparation for ONT-DNA sequencing used the ONT Ligation Sequencing kit V14
698 (ONT, SQK-LSK114), using 1 µg sheared genomic HMW DNA in 47 µl nuclease-free water
699 (ThermoFisher, Am9937) as input, followed by library preparing and loading R10.4.1 flow
700 cell (FLO-PRO114M, ONT) on PromethION 24 sequencer following the manufacturer's
701 instructions, see supplementary methods for details. Nanopore sequencing was performed
702 using the following settings: a Run limit of 96 h, a Minimum read length of 200 bp, and
703 Super-Accurate Basecalling Model with 5mC and 5hmC modified base context. To increase
704 the output, flow cells were washed once after 22-24 h of sequencing using Flow Cell Wash kit
705 XL (ONT, EXP-WAS004-XL), followed by 1-1.5 h waiting time. Then priming and loading
706 library procedures were the same as described previously. Around 70 Gigabites estimated
707 bases with an average read length of 20 kb were achieved per sample.

708 **Detection of KO efficiency from ONT sequencing data.**

709 We developed a computational tool, *samCRISPR*, to estimate CRISPR/Cas9 KO efficiency
710 using ONT long-read whole-genome sequencing (WGS) data. The name "samCRISPR"
711 references both SAMtools, from which the mpileup format is used in KO efficiency

712 calculations, and CRISPR-Cas9 genome editing technology. SAMtools is used to sort, index,
713 and filter aligned reads, and to generate mpileup output at the target site, which provides base-
714 level resolution of read alignments. This output is parsed by our tool, *samCRISPR*, to identify
715 substitutions, insertions, and deletions based on mpileup letter codes, allowing precise
716 quantification of editing events and KO efficiency. The tool identifies sequencing reads that
717 carry CRISPR-induced events within designated target regions using the mpileup output from
718 SAMtools (Danecek et al., 2021) (**Figure 1F**). To eliminate the chance of counting
719 sequencing errors in the vicinity of target regions as actual CRISPR events, the counting of
720 CRISPR events was restricted to 1bp upstream and 1bp downstream of Cas9 cut sites.
721 CRISPR events were defined as mutations occurring within ±1 base pairs of the predicted cut
722 site, located 3 bp upstream of the PAM sequence. This narrow window was chosen to
723 minimize false positives caused by sequencing or alignment errors near Cas9 cleavage sites.
724 KO efficiency was calculated as the ratio of reads with CRISPR events in the target region to
725 the total number of reads mapped to that region. To validate samCRISPR results, KO
726 efficiencies for selected EM KOs were manually inspected using the Integrative Genomics
727 Viewer (IGV) (Robinson et al., 2011). The *samCRISPR* accepts a BAM file, a reference
728 FASTA, and BED-formatted sgRNA coordinates as input, and supports batch analysis of
729 multiple BAM files and sgRNAs. It requires a Linux environment with SAMtools for mpileup
730 generation and R computing environment for calculating Wilson score confidence intervals
731 (Agresti & Coull, 1998).

732 **Phasing of ONT data and obtaining CpG-level methylation data**

733 We implemented a custom graph genome-based approach integrating both FVB and B6
734 genomic sequences into a unified non-linear reference. A genome graph, a structure
735 composed of nodes and edges, captures alternative alleles, structural variants, and indels
736 across multiple genomes (Liao et al., 2023; Zhang et al., 2025). Unlike linear references,
737 genome graphs support the simultaneous representation of multiple haplotypes, making them
738 particularly suited for admixed or hybrid genomes. Initially, reads were aligned to the linear
739 B6 and FVB references to generate haplotype-specific BAM files. The genome graph was
740 then used to resolve homologous CpG coordinates between strains, for instance, reconciling
741 differences where a CpG site appears at position 6 in B6 but at position 20 in FVB. Whole-
742 genome alignment of ONT WGS data from F1 B6 × FVB crosses was performed using
743 MUMmer4 v4.0 (Marcais et al., 2018) to identify heterozygous variants. Next, Jellyfish
744 (Marcais & Kingsford, 2011) was used to enumerate all 45-mers in each genome, which were

classified into conserved nodes and variant edges. Edges shorter than 3 kb were aligned between genomes to derive CIGAR strings, enabling base-level coordinate conversion. From these, a pseudo-hybrid intermediate genome was constructed by selecting the longer edge between B6 and FVB, thus forming a unified coordinate system encompassing diversity of both strains. Reads were aligned to both references using Minimap2 v2.26 (Li, 2018), retaining only high-confidence alignments (MAPQ > 20). Strain-specific SNPs and alignment scores were used to assign reads to B6 or FVB haplotypes. WhatsHap v2.8 (Patterson et al., 2015) was applied for read phasing based on heterozygous variants. Guppy v6.5.7 in MinKNOW software v23.04.5 was used to call canonical and modified bases (5mC, 5hmC) from raw current signal files fast5/pod5. Methylation of each CpG was calculated using modkit v0.1.2 (https://github.com/nanoporetech/modkit). Importantly, all DNAm data were projected onto the pseudo-hybrid coordinate system, minimizing reference bias during haplotype-specific methylation analysis.

Identification of differentially methylated regions

Methylation levels of each CpG for each sample were extracted from bedMethyl files generated by modkit, which contain per-CpG entries detailing genomic coordinates and read counts. Each file included approximately 21 million CpG sites. These datasets were imported into R to create BSseq objects (Hansen et al., 2012) for downstream DMR analysis. To ensure data reliability, unphased datasets retained only CpGs with ≥3× coverage in at least 80% of samples. For phased data, where coverage was inherently lower, CpGs with ≥1× coverage across all samples were included. DMRs were identified in phased data using the bsseq 1.39.4 (Hansen et al., 2012) and unphased data using dmrseq 1.22.1 (Korthauer et al., 2019) R packages. The dmrseq controls the false discovery rate (FDR) whereas bsseq controls the family-wise error rate (FWER). For the phased data smoothing was performed separately for each allele. Strain-specific CpGs were included during smoothing but were removed prior to DMR identification, following previous work (Wulfridge et al., 2019). This strategy enhanced statistical power while reducing spurious DMR detection due to strain-specific sequence differences. DMRs were filtered using the following criteria phased analysis: ≥3 CpGs per region, >5% mean methylation difference, and inverse CpG density ≤300 (defined as the number of CpGs divided by the region's genomic width), unphased analysis ≥5 CpGs, and inverse CpG density ≤120.

Annotation of differentially methylated regions

We used annotatePeaks.pl program to annotate DMRs by using UCSC mm10 mouse reference fasta and UCSC mm10 refGene gtf files. That program calls a separate program, assignGenomeAnnotation, to efficiently assign DMRs to possible annotations genome wide. By default HOMER, in case some annotation overlap, HOMER assigns a priority based on the following: TSS (by default defined from -1kb to +100bp), TTS (by default defined from -100 bp to +1kb), CDS Exons, 5' UTR Exons, 3' UTR Exons, introns, and intergenic. Intergenic regions were considered either upstream (TSS) or downstream (TTS) of genes.

To annotate DMRs in phased data, a custom pipeline was employed: The UCSC mm10 GTF file was converted to BED format using a Bash script. A Python script mapped coordinates between the mm10 genome and the pseudo-hybrid intermediate genome. This allowed accurate functional annotation of DMRs in a unified coordinate space representing both B6 and FVB haplotypes.

Enrichment analysis of DMRs in genomic features

Canonical transcripts of protein-coding genes in *Mus musculus* were downloaded from Ensembl using the biomaRt package v2.58.2 (Durinck et al., 2005) and annotated using GenomicFeatures package v1.54.4 (Lawrence et al., 2013) in conjunction with the UCSC genome annotation TxDb.Mmusculus.UCSC.mm10.knownGene v3.10.0. A subset of these genomic features was then selected to include only the canonical protein-coding transcripts. For CpG enrichment, we used R computing environment to construct a 2×2 contingency table with the number of CpGs for each genomic feature and DMRs. A Fisher's exact test was used to assess enrichment.

Calculation of CpG mutation proportion in haplotype-specific DMRs

We calculated the proportion of mutated CpG dinucleotides in haplotype-specific DMRs compared to the rest of the genome. The B6 and FVB genome were used as reference and query for alignment in MUMmer4 (Marcais et al., 2018), respectively. A python script was utilized to count each CpG dinucleotide mutation category from output of MUMmer4. To visualize results, for each CpG dinucleotide mutation category, we subtracted the proportion obtained for the rest of the genome from the proportion obtained for haplotype-specific DMRs. These proportion differences are displayed as coefficients in the forest plot. The 95% confidence intervals for the coefficients were derived from the output of prop.test() function in R.

809 **DNA methylation overlap across individual datasets**

810 An R-based pipeline was developed to visualize DMRs between mNPCs and Day 6 neurons,
811 incorporating additional conditions including *Dnmt1*-KO and *Kmt2a*-KOs and Day 2 neurons.
812 Unphased DMR coordinates (from mNPC vs. Day 6 comparisons) were used to extract
813 methylation data from the two experiments. CpGs with zero coverage in any sample were
814 excluded. Remaining CpGs were smoothed with the *BSmooth()* function from the bsseq
815 v1.39.4 package. The *normalizeToMatrix()* function from the EnrichedHeatmap v1.32
816 package was then used to map smoothed average methylation values from each group to a
817 fixed-width matrix centered on the extended DMR intervals (±5kb) with a 100 bp bin size,
818 and heatmaps were generated using the same package.

819 **PCA plot of genome-wide DNAm**

820 To assess global DNAm variation, the mouse genome was segmented into 1 kb non-
821 overlapping tiles using the *tileGenome()* function from the GenomicRanges R package. Tiles
822 containing more than two CpGs were retained, resulting in approximately 2.7 million
823 genomic tiles. A custom R function was used to generate a PCA plot based on raw genome-
824 wide DNAm levels across samples.

825 **RNA isolation and sequencing**

826 RNA was isolated as before using TRI Reagent and purified using ZYMO Direct-zol™ RNA
827 Microprep Kit. RNA concentration was measured using Qubit with the Qubit RNA HS Assay
828 Kit (ThermoFisher, Q32852), and RNA integrity was assessed with a Bioanalyzer using
829 Agilent RNA 6000 Nano Kit (Agilent, 5067-1511) according to the manufacturer's
830 instructions. The RNA samples wereFor *Kmt2a*-KOs and *Dnmt1*-KOs, the two genes were
831 knocked out individually with a scrambled control in 4 biological replicates each, RNA
832 isolated and sent to Novogene Europe for mRNA Illumina sequencing. For wild-type
833 neurodevelopmental cells, 3 biological replicates were isolated for each timepoint, RNA
834 extracted and purified as before and 200ng total RNA from each sample processed using the
835 ONT PCR-cDNA Barcoding kit (ONT, SQK-PCB111.24), according to the manufacturer's
836 instructions (version v111-revG). Samples were barcoded and pooled at equimolar ratios, then
837 loaded onto 3x R9.4.1 flow-cells (FLO-PRO002) and sequenced on a PromethION 24
838 instrument.

839 **RNA sequencing analysis**

840 Illumina short-read RNA-seq (paired-end reads) with 150bp read length on average was
841 generated from *Dnmt1*-KO and *Kmt2a*-KO mNPCs. The libraries were sequenced at
842 approximately 78 million reads per sample. Reads were aligned to the UCSC mm10 reference
843 genome using HISAT2 (Kim et al., 2019). Alignments with >50 mapping quality were used in
844 featureCounts (Liao et al., 2014) to create a gene-level count matrix. This count matrix was
845 normalized in R using the cpm() function from the edgeR (Robinson et al., 2010), accounting
846 for gene lengths and library size. Protein coding genes with >1 Counts Per Million (CPM) in
847 at least one sample were retained for differential expression analysis with DESeq2 (Love et
848 al., 2014), considering genes with an adjusted p-value <0.05 as differentially expressed. For
849 long-read ONT RNA-seq, libraries were sequenced with an average of ~10 million reads and
850 a median read length of ~1.1kb. Reads were aligned to the UCSC mm10 reference genome
851 with Minimap2 (Li, 2018) using *-uf* and *-ax splice* arguments. The featureCounts was utilized
852 to generate the gene-level count matrix with the *-L* argument for long reads. DE analysis
853 followed the same procedure as for short-read data. To ensure reliability, a histogram plot of
854 unadjusted p-values for all genes with >1 CPM was created in R. To stabilize variance,
855 regularized log transformation (rlog) was applied to raw expression values of retained genes
856 using the DESeq2 package. PCA was performed on the transposed rlog-transformed
857 expression matrix using *prcomp()*, and the first two principal components (PC1 and PC2)
858 were visualized using ggplot2.

859 **Obtaining promoters of mouse protein-coding genes for NMF analysis**

860 The *promoters()* function of the R package TxDb.Mmusculus.UCSC.mm10.knownGene v3.2
861 was utilized to fetch out UCSC mm10 coordinates of promoters. The promoter definition was
862 considered 2kb upstream and 500bp downstream of transcription start site (TSS) of canonical
863 transcript of each gene. The mouse protein coding gene list was downloaded from
864 https://ftp.ncbi.nih.gov/gene/DATA/gene_info.gz.

865 **Clustering of EM KOs by NMF based on promoter DNAm abnormalities**

866 We used NMF on smoothed DNAm levels of promoters of protein-coding genes (Cemgil,
867 2009). CpGs with more than 3x coverage across at least more than 80% of samples were kept
868 in our analysis. The promoters with no retained CpGs were discarded. The number of latent
869 factors (parameter k) was chosen to be three based on diagnostic plots of NMF and PPI-
870 interactions of EM components. The brunet model in the R package nmf v0.28 was employed
871 in decomposition.

872 **Inspection of STRING interactions of EM components**

873 These interactions consisted of three categories: known interactions derived from curated
874 databases and experimental data; predicted interactions based on gene neighborhood, gene
875 fusions, and co-occurrence; and other interactions inferred from text mining, co-expression,
876 and protein homology. We used STRING interactions (confidence score cutoff of >0.400)
877 (Bernas et al., 2017) to determine whether the shared impact of EM-KOs on promoter DNAm
878 is associated with PPI-nets of EM factors with DNAm EM components.

879 **Enrichment analysis and visualization**

880 To test whether the number of overlapping genes between different RNA sequencing datasets
881 was higher than expected, Fisher's exact tests were performed using the function fisher.test()
882 in R. Shared detected genes were considered background. To test if TF targets or imprinted
883 genes were enriched in DEGs lists, Fisher's exact tests were performed using the function
884 fisher.test() in R, using detected genes as background. TF target genes were obtained from
885 https://tfbsdb.systemsbiology.net/ (Plaisier et al., 2016) and the TFlink (https://tflink.net/)
886 (Liska et al., 2022) databases. A list of mouse imprinted genes was obtained from the
887 Imprinted Gene Database (https://www.geneimprint.com/site/genes-by-
888 species.Mus+musculus), filtering away genes with "unknown" or "not imprinted" status.

889 **Gene ontology analysis**

890 To estimate enrichment in certain gene ontology (GO) terms in RNA sequencing datasets, the
891 Webgestalt 2024 (Elizarraras et al., 2024) webtool was used, using mouse protein-coding
892 genes as background. The GO term "Biological Process" was evaluated and FDR<0.05 used
893 to estimate significance. The top 10-15 terms were reported for each analysis. For enrichment
894 of certain GO terms and KEGGs pathways in genes associated with latent factors in the NMF
895 analysis, the EnrichR (Kuleshov et al., 2016) webtool was used.

896 **Transcription factor motif enrichment analysis**

897 For identification of TF motif enrichment in DMRs and promoter regions, the Analysis for
898 Motif Enrichment (AME) tool within MEME-suite v5.5.8 (https://meme-
899 suite.org/meme/tools/ame) was used (Bailey et al., 2015; McLeay & Bailey, 2010). The top
900 30k most differentially methylated regions from the *Dnmt1*-KO were tested against a
901 background that included 30k random regions from the mm10 genome that were matched in
902 size distribution and CpG density and did not include any of the *Dnmt1*-KO DMRs

903 (**Supplementary Table 16**). For promoter regions of shared DEGs we obtained mouse
904 protein-coding promoters for canonical transcripts as described above, filtered them for the
905 shared DEGs and used all other protein-coding genes as a background. In both cases, we
906 tested against the mouse HOCOMOCO v11 core database and estimated enrichment using a
907 Fisher's exact test. Odds ratios were calculated using the following formula:

908 $$OR = \frac{\%TP \times (100 - \%FP)}{\%FP \times (100 - \%TP)}$$

909 where TP = true positives, or positive hits in the tested sequences, and FP = false positives, or
910 positive hits in the background sequences.

911 **Data visualization**

912 For scatter plots and volcano plots, the R package ggplot2 was used (Wickham, 2016). For
913 Venn diagrams, the R package VennDiagram was used (Chen & Boutros, 2011). The
914 heatmaps were generated through the *pheatmap()* function of the R package pheatmap. For
915 individual gene expression values from RNA sequencing data, and for quantification results
916 from qPCR, WB and IF experiments, Graphpad Prism (v10.4) was used. Schematic figures
917 were generated using BioRender.

918 **Data availability**

919 Illumina short-read RNA sequencing data of *Kmt2a* and *Dnmt1* KO mNPCs are deposited in
920 GEO under accession number GSE308011. ONT long-read RNA sequencing data of murine
921 neurodevelopmental samples is deposited in GEO under accession number GSE308012. ONT
922 long-read WGS data of murine neurodevelopmental samples is deposited in GEO under
923 accession number GSE308528. EM KO data is deposited in GEO under two accession
924 numbers GSE308573 and GSE308669. All data will be publicly available upon publication.

925 **Code availability**

926 Source code for samCRISPR is deposited in https://github.com/kaanokay/samCRISPR. Other
927 source codes are deposited in https://github.com/kaanokay/DNA-methylation-paper.

928 **Acknowledgements:** We would like to thank Elvar Örn Jónsson and Þorgils Árni
929 Hjálmarsson for their dedication and valuable efforts in troubleshooting data storage and
930 analysis issues within the high-performance computing (HPC) cluster at the University of
931 Iceland. We would like to thank Olafur Th. Magnusson for advice regarding ONT
932 sequencing. H.T.B. has a grant from RANNIS (Grant of Excellence, 217988-051) specifically

covering this project. J.O. has a grant from Eimskipasjóður Íslands (Doctoral grant, 1238113). K.M. has a grant from RANNIS (Postdoctoral Fellowship, 2410204-052). This work benefitted from equipment from the SAMSNID infrastructure project which brought the PromethION instrument to Iceland, and the IREI infrastructure project which supports the high-performance computer cluster at the University of Iceland. We used ChatGPT (https://chatgpt.com/) to assist in improving scripting efficiency and robustness.

Author contributions: H.T.B. and K.D.H. conceptualized project and supervised students. H.T.B. received the project funding to fund the entire project; J.O., K.M., K.W., and T.P. performed experiments. K.O., K.D.H., J.O. and K.M. analyzed the data. H.T.B., K.D.H., K.M., J.O. and K.O. wrote manuscript. K.A. K.W. and T.P. edited manuscript.

Declaration of interest: H.T.B. is a consultant for Mahzi therapeutics and founder of KALDUR therapeutics.

945 **Tables**

946 **List of supplementary tables**

947 **Supplementary Table 1.** Guide RNA and scrambled guide RNA pairs sequences for each of
948 the genes.

949 **Supplementary Table 2.** A list of genome-wide significant DMRs from *Dnmt1*-KO, genomic
950 location and presence in regulatory material.

951 **Supplementary Table 3.** Enrichment results of genomic region enrichment analysis of the
952 *Dnmt1*-KO DMRs.

953 **Supplementary Table 4.** Transcription factor motif enrichment in the top 30K *Dnmt1*-KO
954 DMRs compared to a length and CpG density matched background list.

955 **Supplementary Table 5.** Differential expression analysis results after *Dnmt1*-KO.

956 **Supplementary Table 6.** Differential expression analysis results after *Kmt2a*-KO.

957 **Supplementary Table 7.** Shared DEGs between *Dnmt1*-KO and *Kmt2a*-KO.

958 **Supplementary Table 8.** Transcription factor motif enrichment of promoters of shared DEGs
959 between *Dnmt1*-KO and *Kmt2a*-KO.

960 **Supplementary Table 9.** Differential expression analysis results after 2 days of
961 differentiation.

962 **Supplementary Table 10.** Differential expression analysis results after 6 days of
963 differentiation.

964 **Supplementary Table 11.** Genome-wide DMRs between time points of neurodevelopmental
965 model.

966 **Supplementary Table 12.** Overlap between *Dnmt1*-KO DMRs and Day6 DMRs.

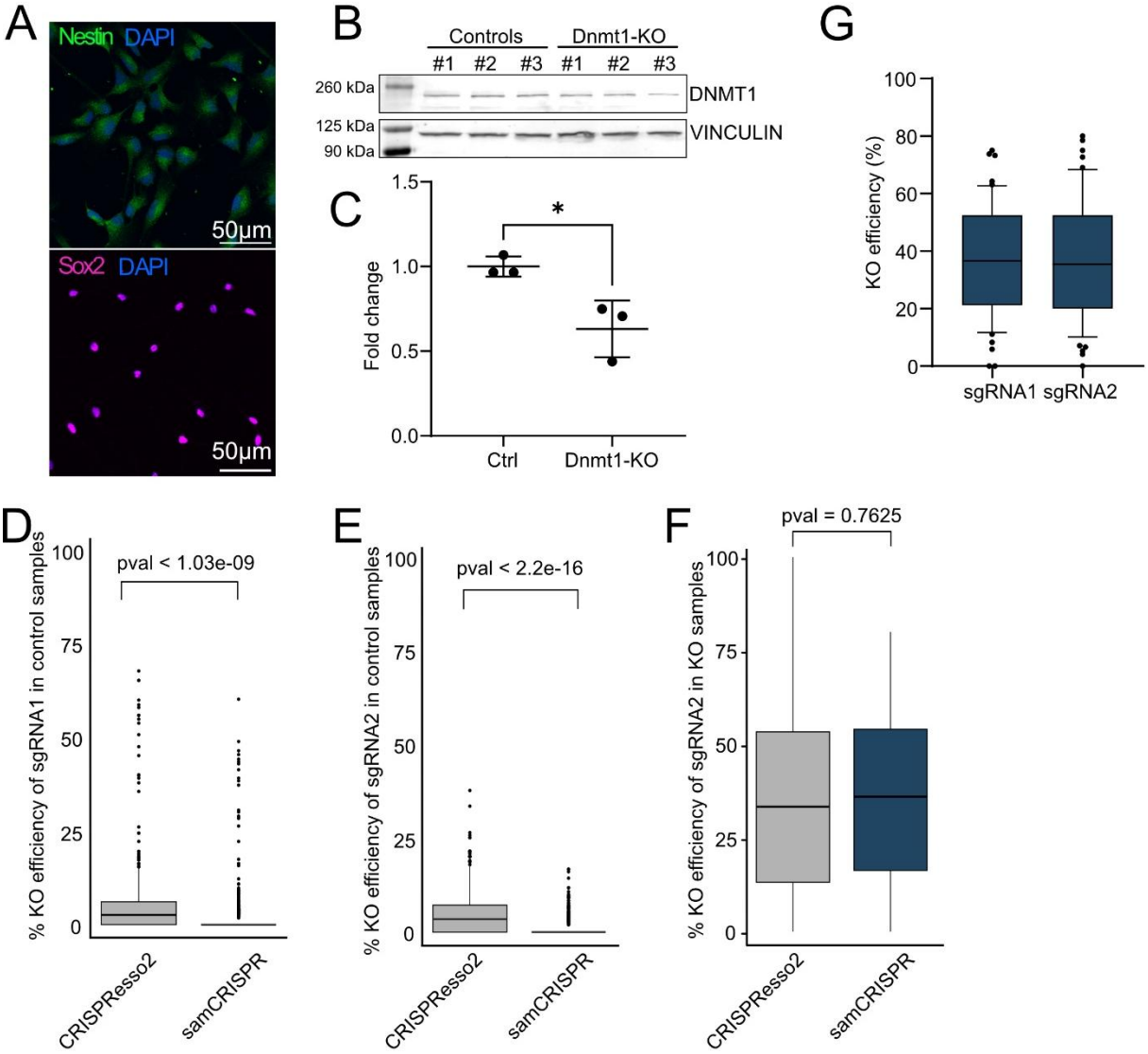
967 **Supplementary Table 13.** Haplotype specific DMRs.

968 **Supplementary Table 14.** Putative imprinting control regions (ICRs).

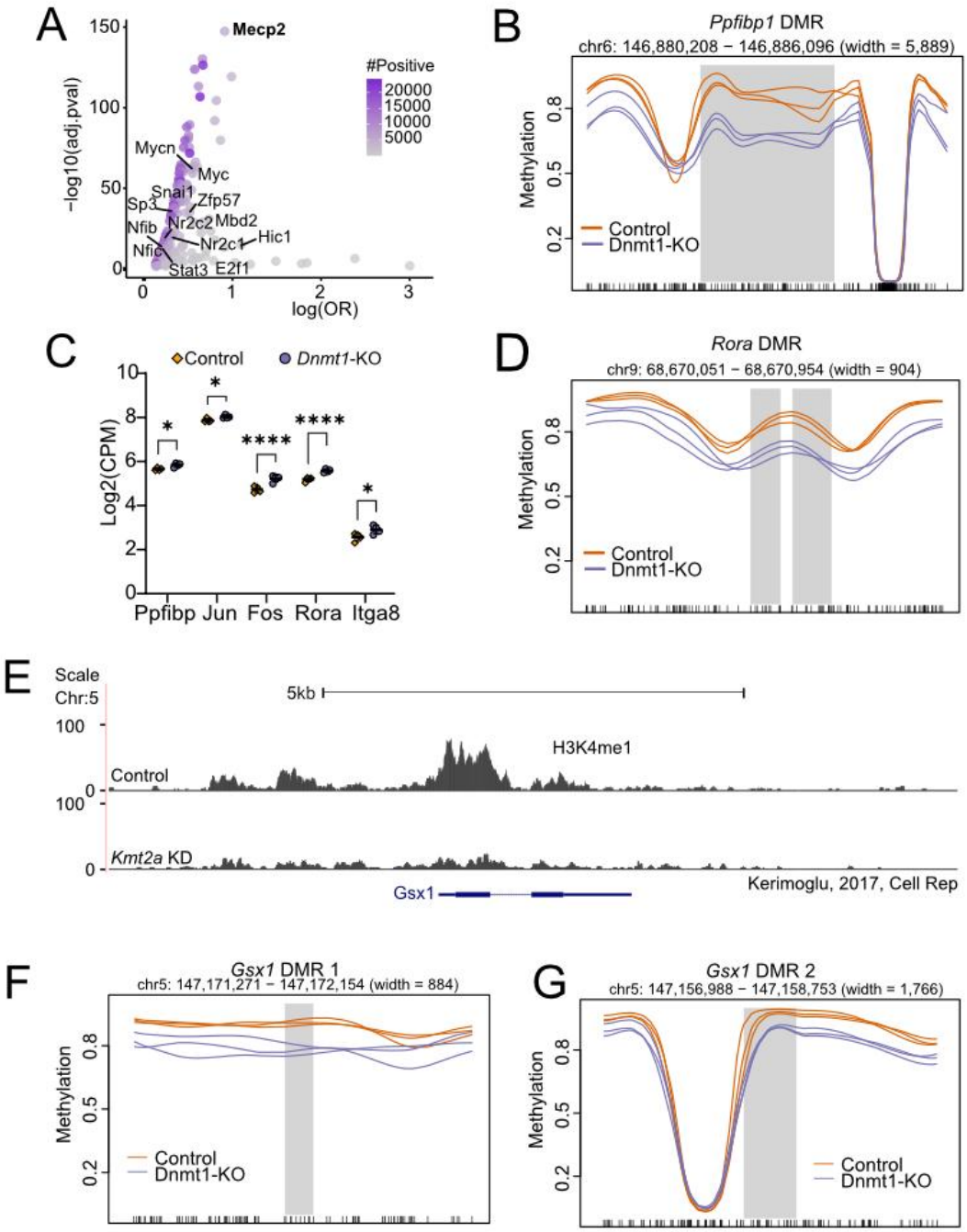
969 **Supplementary Table 15**. Primers used during the current study.

970 **Supplementary Table 16.** A list of 30k DMRs that were used as background for motif
971 enrichment analysis.

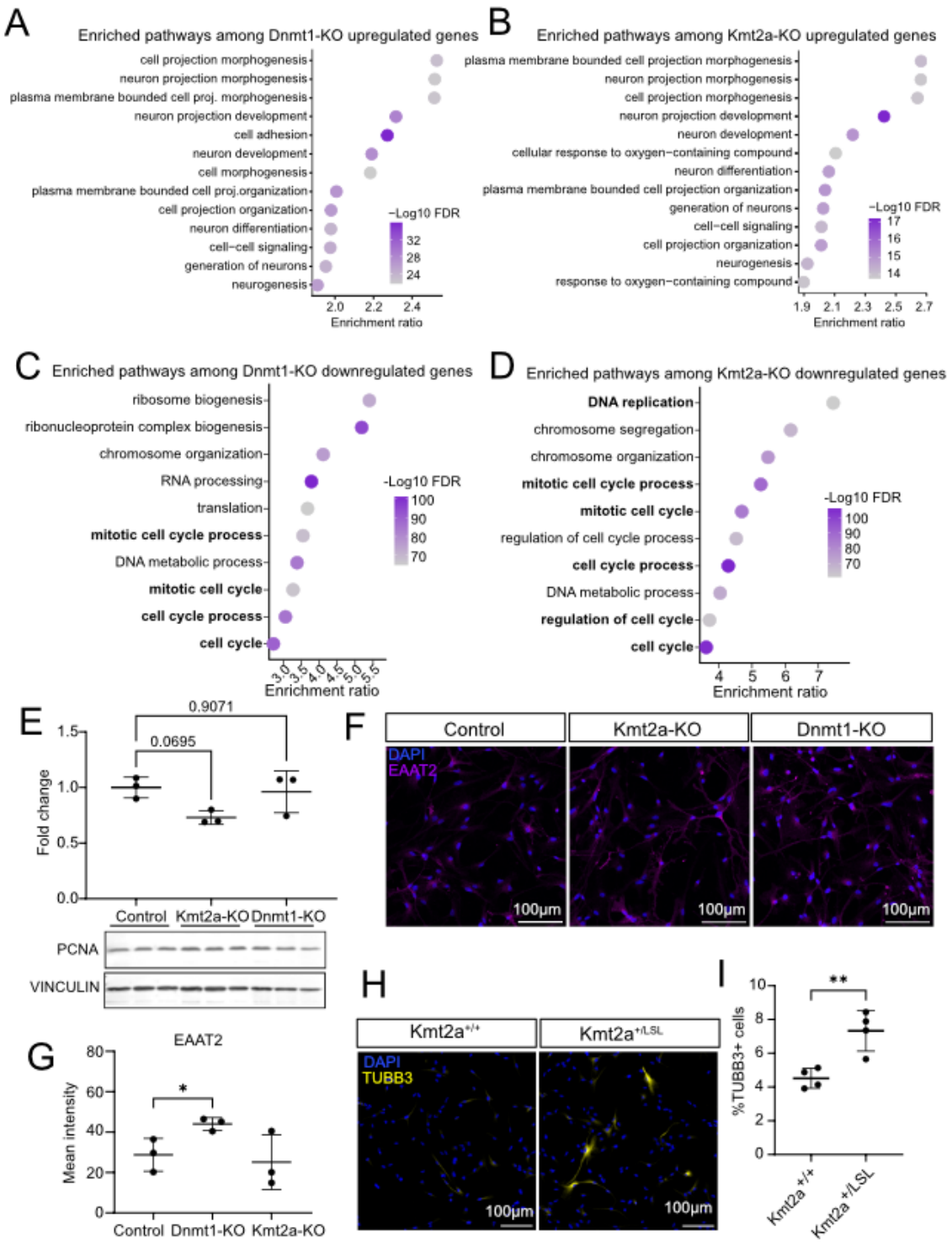
Supplemental figures



Supplementary Figure 1: CRISPR-Cas9 KO system evaluation. (A) IF staining of isolated mouse NPCs, staining for Nestin (green, upper) and Sox2 (magenta, lower) and DAPI (blue). (B-C) WB and quantification showing *Dnmt1*-KO and scrambled controls in triplicates. Results in (C) show DNMT1 (upper in B) normalized to VINCULIN (lower in B). A t-test was used for significance. (D) KO efficiency of sgRNA 1 in control samples across methods (two-tailed t-test). (E) KO efficiency of sgRNA 2 in control samples across methods (two tailed t-test). (F) KO efficiency of sgRNA 2 in EM KOs across methods (two-tailed t-test). (G) KO efficiency of sgRNA 1 and 2 in EM KOs using *samCRISPR*.



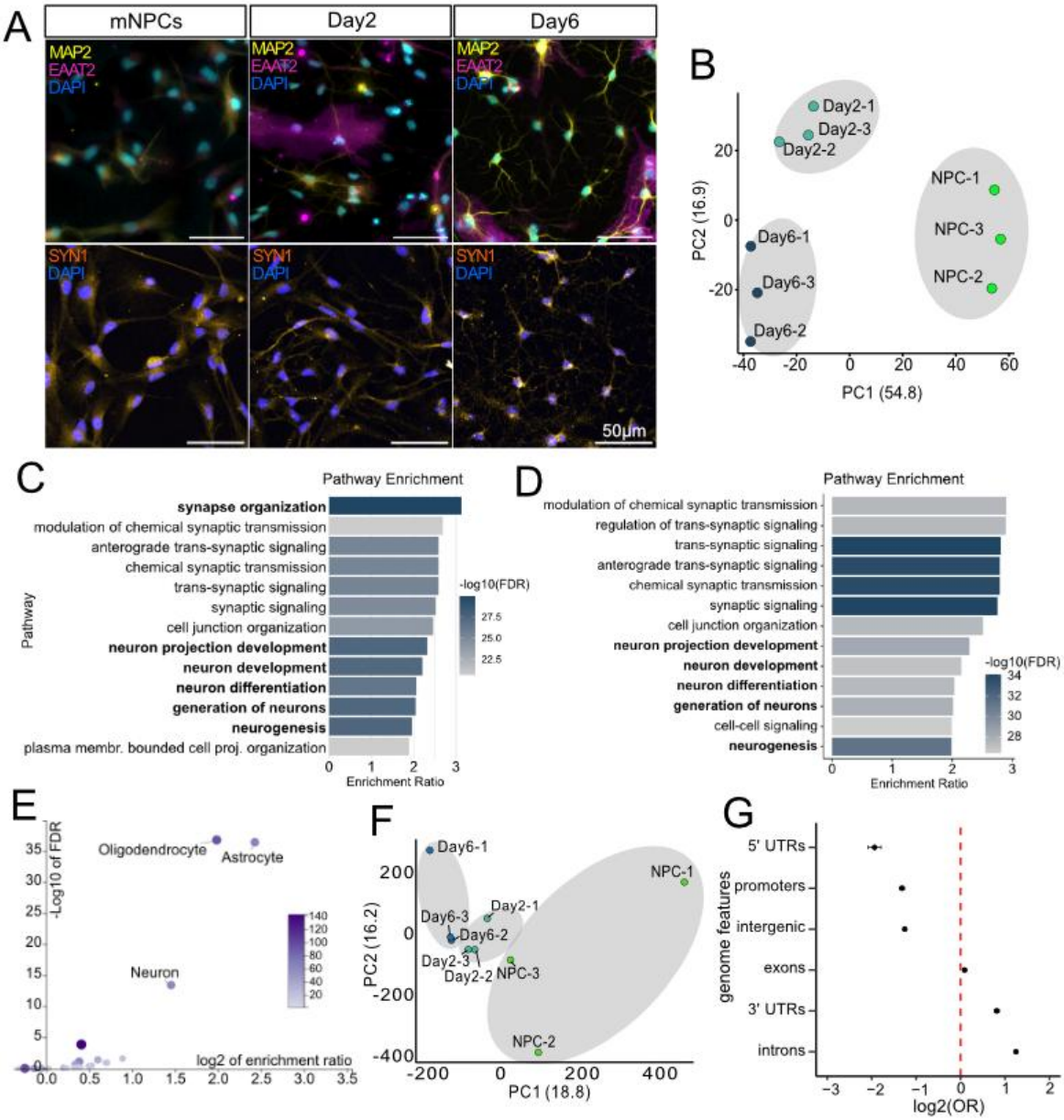
Supplementary Figure 2: *Dnmt1*-KO and *Kmt2a*-KO DMRs. (A) Scatter plot showing transcription factor's motif enrichment in *Dnmt1*-KO associated DMRs. Known binding partners of DNMT1 from BioGrid are labelled. (B,D) Examples of genome-wide significant DMRs in *Dnmt1*-KO mNPCs (blue) compared to controls (orange), located near the neuronal-related gene *Ppfibp1* (B), and a known DNMT1 target gene *Rora* (D) (Tao et al., 2022). (C) Dot plot showing normalized expression 5 examples of *Dnmt1*-KO DMR-associated DEGs. (E) Publicly available ChIP-seq data (Kerimoglu et al., 2017) showing a loss of H3K4me1 peaks at the *Gsx1* exon in *Kmt2a* knockdown cells compared to control. (F,G) Smoothed methylation plots showing two *Dnmt1*-KO DMRs located in an intergenic region near *Gsx1* gene.



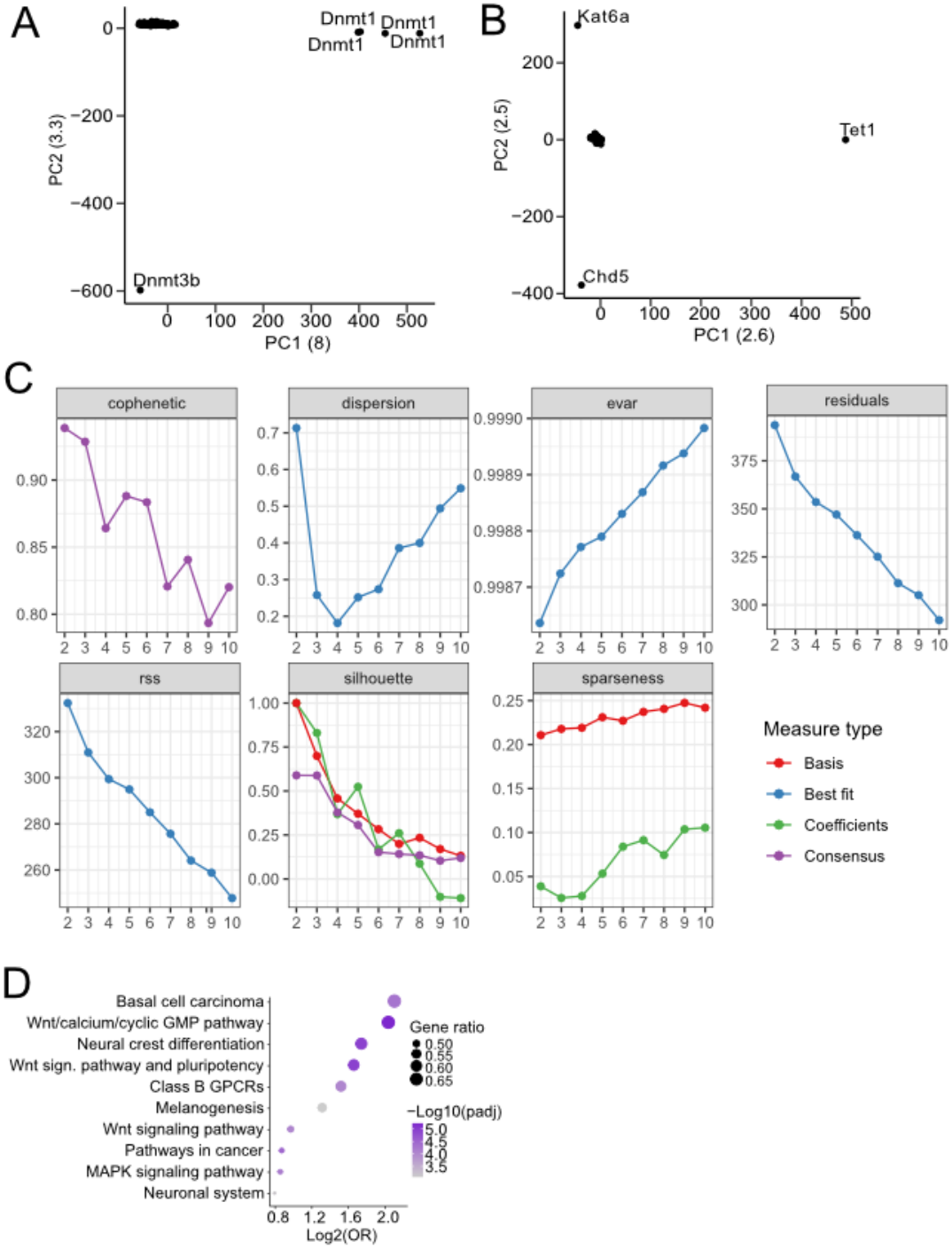
995

996 **Supplementary Figure 3: RNA abnormalities in *Kmt2a*-KO and *Dnmt1*-KO DEGs.**
997 (A-D) Gene ontology (GO) term analysis (Biological Process) for upregulated genes in (A)
998 *Dnmt1*-KOs and (B) *Kmt2a*-KOs, and downregulated genes in (C) *Dnmt1*-KOs and (D)
999 *Kmt2a*-KOs. (E) Representative immunofluorescence images of control, *Kmt2a*-KO and
1000 *Dnmt1*-KO mNPCs stained for DAPI (blue) and EAAT2 (magenta). Scale bars: 100 µm. (F)
1001 Quantification of EAAT2 mean fluorescence intensity in control, *Kmt2a*-KO and *Dnmt1*-KO
1002 mNPCs (n = 3, $p < 0.05$, a One-way Anova with Dunnett's correction was used.). (G) WB
1003 analysis of PCNA expressions in *Kmt2a*-KO (ns, p = 0.0695) and *Dnmt1*-KO (ns, p = 0.9071)
1004 compared to control mNPC (One-way ANOVA with Dunnett´s correction was used). (H)

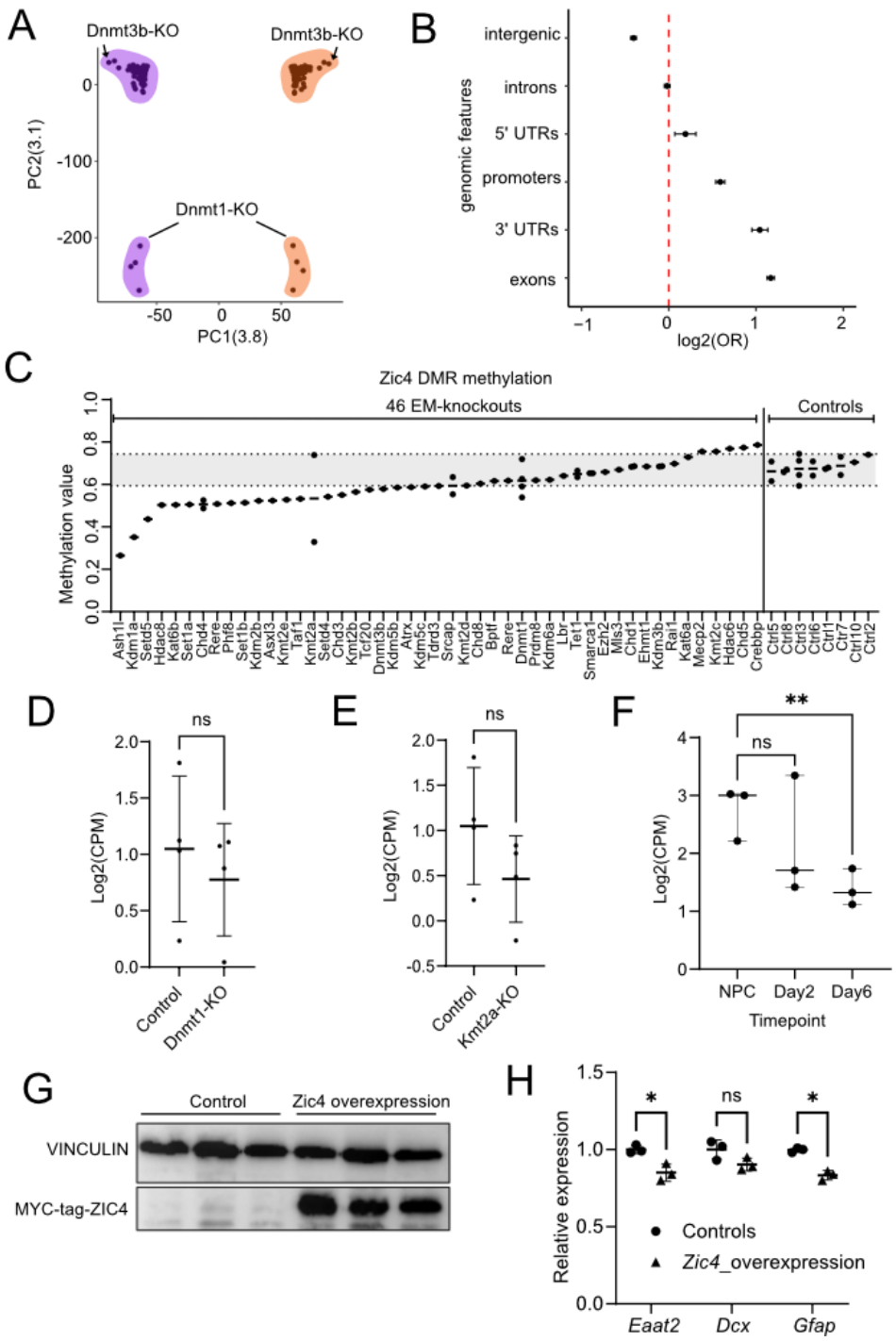
1005 Representative immunofluorescence images of B6J-Kmt2a$^{+/+}$ and B6J-Kmt2a$^{+/LSL}$ P0 mNPCs,
1006 stained for DAPI (blue) and TUBB3 (yellow). (I) Quantification of % of TUBB3 positive
1007 cells in B6J-Kmt2a$^{+/+}$ and B6J-Kmt2a$^{+/LSL}$ P0 mNPCs. A student's t-test was used. *$p<0.05$,
1008 **$p<0.01$, ns=not significant.
1009



Supplementary Figure 4: DNA methylation changes in neurodevelopment. (A) Immunofluorescence staining of mNPCs and neuronal cultures at Day2 and Day6 of differentiation. The upper panel shows staining for Neurons (MAP2: yellow) and Astrocytes (EAAT2: magenta), lower panel shows staining for synapse marker expression (SYN1: orange). DAPI is shown in blue. (B) PCA plot of mNPCs-Day2-Day6 datasets based on ONT RNA sequencing and gene expression analysis. (C-D) GO-term enrichment (Biological Process) analysis showing the top 13 enriched GO terms among DEGs at Day2 (C) and Day6 (D). (E) Gene-set enrichment analysis for cell-types, showing a significant enrichment for all three neuronal lineages in the Day6-DEGs. (F) PCA plot of mNPC-Day2-Day6 samples after ONT DNA sequencing and analysis for differentially methylated regions. (G) Bar graph showing the log2(observed/expected) enrichment ratios for genome locations of Day6-DMRs.



Supplementary Figure 5. Promoter specific analysis of EM-KO DNAm (A) PCA plot showing 46 EM KOs and 8 control samples. (B) PCA plot showing same analysis as in A) but without *Dnmt1* and *Dnmt3b* samples. (C) Diagnostic plots in NMF analysis in decision of number of latent factors (D) GO term analysis (KEGGs pathway) of genes contributing most to latent factor 2.



Supplementary Figure 6. A potential marker gene for EM disruption. (A) A PCA plot showing *Dnmt1*-KO and *Dnmt3b*-KO are two outliers. (B) Dot plot showing methylation level at the 3´UTR of *Zic4* locus across 46 EM-KOs compared to 8 controls. Two horizontal dotted lines indicate minimum and maximum values of the controls. (C) The enrichment of haplotype-specific DMRs within genomic features (Fisher's exact test used). (D-E) Normalized counts showing *Zic4* expression in *Kmt2a*-KO (not significant, p = 0.1846, D) and *Dnmt1*-KO (not significant, p = 0.5, E), compared to controls. (F) Normalized counts showing *Zic4* expression in neurodevelopment, where it is significantly downregulated at Day6 compared with the NPC state (p = 0.00197). (G) WB analysis of Myc-ZIC4 protein after overexpression. (H) qPCR results showing relative mRNA levels of *Eaat2, Dcx* and *Gfap* after ZIC4 overexpression, compared with Myc-only controls; significance calculated using a t-test.

References

Agresti, A., & Coull, B. A. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, *52*(2), 119–126. https://doi.org/10.1080/00031305.1998.10480550

Aref-Eshghi, E., Kerkhof, J., Pedro, V. P., Groupe, D. I. F., Barat-Houari, M., Ruiz-Pallares, N., Andrau, J. C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., Cormier-Daire, V., Rondeau, S., Lecoquierre, F., Saugier-Veber, P., Nicolas, G., Lesca, G., Chatron, N., Sanlaville, D.,…Sadikovic, B. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet*, *106*(3), 356–370. https://doi.org/10.1016/j.ajhg.2020.01.019

Aref-Eshghi, E., Rodenhiser, D. I., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., Pare, G., Hood, R. L., Bulman, D. E., Kernohan, K. D., Care4Rare Canada, C., Boycott, K. M., Campeau, P. M., Schwartz, C., & Sadikovic, B. (2018). Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. *Am J Hum Genet*, *102*(1), 156–174. https://doi.org/10.1016/j.ajhg.2017.12.008

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res*, *43*(W1), W39–49. https://doi.org/10.1093/nar/gkv416

Bernas, S. N., Leiter, O., Walker, T. L., & Kempermann, G. (2017). Isolation, Culture and Differentiation of Adult Hippocampal Precursor Cells. *Bio Protoc*, *7*(21), e2603. https://doi.org/10.21769/BioProtoc.2603

Bjornsson, H. T. (2015). The Mendelian disorders of the epigenetic machinery. *Genome Res*, *25*(10), 1473–1481. https://doi.org/10.1101/gr.190629.115

Boukas, L., Havrilla, J. M., Hickey, P. F., Quinlan, A. R., Bjornsson, H. T., & Hansen, K. D. (2019). Coexpression patterns define epigenetic regulators associated with neurological dysfunction. *Genome Res*, *29*(4), 532–542. https://doi.org/10.1101/gr.239442.118

Boukas, L., Luperchio, T. R., Razi, A., Hansen, K. D., & Bjornsson, H. T. (2024). Neuron-specific chromatin disruption at CpG islands and aging-related regions in Kabuki syndrome mice. *Genome Res*, *34*(5), 696–710. https://doi.org/10.1101/gr.278416.123

Carosso, G. A., Boukas, L., Augustin, J. J., Nguyen, H. N., Winer, B. L., Cannon, G. H., Robertson, J. D., Zhang, L., Hansen, K. D., Goff, L. A., & Bjornsson, H. T. (2019). Precocious neuronal differentiation and disrupted oxygen responses in Kabuki syndrome. *JCI Insight*, *4*(20). https://doi.org/10.1172/jci.insight.129375

Cemgil, A. T. (2009). Bayesian inference for nonnegative matrix factorisation models. *Comput Intell Neurosci*, *2009*, 785152. https://doi.org/10.1155/2009/785152

Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, *12*, 35. https://doi.org/10.1186/1471-2105-12-35

Choufani, S., Cytrynbaum, C., Chung, B. H., Turinsky, A. L., Grafodatskaya, D., Chen, Y. A., Cohen, A. S., Dupuis, L., Butcher, D. T., Siu, M. T., Luk, H. M., Lo, I. F., Lam, S. T., Caluseriu, O., Stavropoulos, D. J., Reardon, W., Mendoza-Londono, R., Brudno, M., Gibson, W. T.,…Weksberg, R. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. *Nat Commun*, *6*, 10207. https://doi.org/10.1038/ncomms10207

Ciptasari, U., & van Bokhoven, H. (2020). The phenomenal epigenome in neurodevelopmental disorders. *Hum Mol Genet*, *29*(R1), R42–R50. https://doi.org/10.1093/hmg/ddaa175

Coltogirone, R. A., Sherfinski, E. I., Dobler, Z. A., Peterson, S. N., Andlinger, A. R., Fadel, L. C., Patrick, R. L., & Bergeron, S. A. (2023). Gsx2, but not Gsx1, is necessary for early forebrain patterning and long-term survival in zebrafish. *Dev Dyn*, *252*(3), 377–399. https://doi.org/10.1002/dvdy.542

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

1094 Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev*, *25*(10),
1095 1010–1022. https://doi.org/10.1101/gad.2037511
1096 Do, C., Lang, C. F., Lin, J., Darbary, H., Krupska, I., Gaba, A., Petukhova, L., Vonsattel, J. P., Gallagher,
1097 M. P., Goland, R. S., Clynes, R. A., Dwork, A., Kral, J. G., Monk, C., Christiano, A. M., & Tycko,
1098 B. (2016). Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA
1099 Methylation. *Am J Hum Genet*, *98*(5), 934–955. https://doi.org/10.1016/j.ajhg.2016.03.027
1100 Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart
1101 and Bioconductor: a powerful link between biological databases and microarray data
1102 analysis. *Bioinformatics*, *21*(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525
1103 Elizarraras, J. M., Liao, Y., Shi, Z., Zhu, Q., Pico, A. R., & Zhang, B. (2024). WebGestalt 2024: faster
1104 gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Res*,
1105 *52*(W1), W415–W421. https://doi.org/10.1093/nar/gkae456
1106 Fahrner, J. A., Lin, W. Y., Riddle, R. C., Boukas, L., DeLeon, V. B., Chopra, S., Lad, S. E., Luperchio, T. R.,
1107 Hansen, K. D., & Bjornsson, H. T. (2019). Precocious chondrocyte differentiation disrupts
1108 skeletal growth in Kabuki syndrome mice. *JCI Insight*, *4*(20).
1109 https://doi.org/10.1172/jci.insight.129380
1110 Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., & Liu, H. K. (2013). The
1111 chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription
1112 factors. *Cell Stem Cell*, *13*(1), 62–72. https://doi.org/10.1016/j.stem.2013.05.002
1113 Foroutan, A., Haghshenas, S., Bhai, P., Levy, M. A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi, A.,
1114 Pedace, L., Miele, E., Genevieve, D., Heide, S., Alders, M., Zampino, G., Merla, G., Fradin, M.,
1115 Bieth, E., Bonneau, D., Dieterich, K.,…Lebre, A. S. (2022). Clinical Utility of a Unique Genome-
1116 Wide DNA Methylation Signature for KMT2A-Related Syndrome. *Int J Mol Sci*, *23*(3).
1117 https://doi.org/10.3390/ijms23031815
1118 Friedrich, G., & Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to
1119 identify and mutate developmental genes in mice. *Genes Dev*, *5*(9), 1513–1523.
1120 https://doi.org/10.1101/gad.5.9.1513
1121 Fu, M. P., Merrill, S. M., Sharma, M., Gibson, W. T., Turvey, S. E., & Kobor, M. S. (2023). Rare diseases
1122 of epigenetic origin: Challenges and opportunities. *Front Genet*, *14*, 1113086.
1123 https://doi.org/10.3389/fgene.2023.1113086
1124 Garel, S., Marin, F., Grosschedl, R., & Charnay, P. (1999). Ebf1 controls early cell differentiation in the
1125 embryonic striatum. *Development*, *126*(23), 5285–5294.
1126 https://doi.org/10.1242/dev.126.23.5285
1127 Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed,
1128 T. P., Blewitt, M. E., & Ritchie, M. E. (2019). Using long-read sequencing to detect imprinted
1129 DNA methylation. *Nucleic Acids Res*, *47*(8), e46. https://doi.org/10.1093/nar/gkz107
1130 Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W. B., & Millen, K. J. (2004). Heterozygous
1131 deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat*
1132 *Genet*, *36*(10), 1053–1055. https://doi.org/10.1038/ng1420
1133 Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite
1134 sequencing reads to differentially methylated regions. *Genome Biol*, *13*(10), R83.
1135 https://doi.org/10.1186/gb-2012-13-10-r83
1136 Heberle, E., & Bardet, A. F. (2019). Sensitivity of transcription factors to DNA methylation. *Essays*
1137 *Biochem*, *63*(6), 727–741. https://doi.org/10.1042/EBC20190033
1138 Irion, S., Luche, H., Gadue, P., Fehling, H. J., Kennedy, M., & Keller, G. (2007). Identification and
1139 targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*, *25*(12), 1477–
1140 1482. https://doi.org/10.1038/nbt1362
1141 Isles, A. R. (2022). The contribution of imprinted genes to neurodevelopmental and neuropsychiatric
1142 disorders. *Transl Psychiatry*, *12*(1), 210. https://doi.org/10.1038/s41398-022-01972-4
1143 Jonsson, M. E., Ludvik Brattas, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., Verschuere, S.,
1144 Madsen, S., Hansson, J., Larsson, J., Mansson, R., Meissner, A., & Jakobsson, J. (2019).

1145 Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human
1146 neural progenitors. *Nat Commun*, *10*(1), 3182. https://doi.org/10.1038/s41467-019-11150-8
1147 Kerimoglu, C., Sakib, M. S., Jain, G., Benito, E., Burkhardt, S., Capece, V., Kaurani, L., Halder, R., Agis-
1148 Balboa, R. C., Stilling, R., Urbanke, H., Kranz, A., Stewart, A. F., & Fischer, A. (2017). KMT2A
1149 and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Rep*,
1150 *20*(3), 538–548. https://doi.org/10.1016/j.celrep.2017.06.072
1151 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment
1152 and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, *37*(8), 907–915.
1153 https://doi.org/10.1038/s41587-019-0201-4
1154 Kim, S., Kim, H., & Um, J. W. (2018). Synapse development organized by neuronal activity-regulated
1155 immediate-early genes. *Exp Mol Med*, *50*(4), 1–7. https://doi.org/10.1038/s12276-018-0025-
1156 1
1157 Kimura, H., & Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for
1158 maintenance DNA methyltransferase, Dnmt1. *J Biol Chem*, *278*(7), 4806–4812.
1159 https://doi.org/10.1074/jbc.M209923200
1160 Korthauer, K., Chakraborty, S., Benjamini, Y., & Irizarry, R. A. (2019). Detection and accurate false
1161 discovery rate control of differentially methylated regions from whole genome bisulfite
1162 sequencing. *Biostatistics*, *20*(3), 367–383. https://doi.org/10.1093/biostatistics/kxy007
1163 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S.,
1164 Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen,
1165 G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web
1166 server 2016 update. *Nucleic Acids Res*, *44*(W1), W90–97.
1167 https://doi.org/10.1093/nar/gkw377
1168 Kutner, R. H., Zhang, X. Y., & Reiser, J. (2009). Production, concentration and titration of pseudotyped
1169 HIV-1-based lentiviral vectors. *Nat Protoc*, *4*(4), 495–505.
1170 https://doi.org/10.1038/nprot.2009.22
1171 La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth,
1172 C., Khven, I., Lederer, A. R., Dratva, L. M., Johnsson, A., Nilsson, M., Lonnerberg, P., &
1173 Linnarsson, S. (2021). Molecular architecture of the developing mouse brain. *Nature*,
1174 *596*(7870), 92–96. https://doi.org/10.1038/s41586-021-03775-x
1175 Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019).
1176 CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids*
1177 *Res*, *47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365
1178 Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey,
1179 V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol*, *9*(8),
1180 e1003118. https://doi.org/10.1371/journal.pcbi.1003118
1181 Lei, X., & Jiao, J. (2018). UTX Affects Neural Stem Cell Proliferation and Differentiation through PTEN
1182 Signaling. *Stem Cell Reports*, *10*(4), 1193–1207.
1183 https://doi.org/10.1016/j.stemcr.2018.02.008
1184 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–
1185 3100. https://doi.org/10.1093/bioinformatics/bty191
1186 Liang, H., Xiao, G., Yin, H., Hippenmeyer, S., Horowitz, J. M., & Ghashghaei, H. T. (2013). Neural
1187 development is dependent on the function of specificity protein 2 in cell cycle progression.
1188 *Development*, *140*(3), 552–561. https://doi.org/10.1242/dev.085621
1189 Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J.,
1190 Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X.,
1191 Fischer, C., Fulton, R. S.,…Paten, B. (2023). A draft human pangenome reference. *Nature*,
1192 *617*(7960), 312–324. https://doi.org/10.1038/s41586-023-05896-x
1193 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for
1194 assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.
1195 https://doi.org/10.1093/bioinformatics/btt656

1196 Liska, O., Bohar, B., Hidas, A., Korcsmaros, T., Papp, B., Fazekas, D., & Ari, E. (2022). TFLink: an
1197 integrated gateway to access transcription factor-target gene interactions for multiple
1198 species. *Database (Oxford)*, *2022*. https://doi.org/10.1093/database/baac083
1199 Lo, H. L., & Yee, J. K. (2007). Production of vesicular stomatitis virus G glycoprotein (VSV-G)
1200 pseudotyped retroviral vectors. *Curr Protoc Hum Genet*, *Chapter 12*, Unit 12 17.
1201 https://doi.org/10.1002/0471142905.hg1207s52
1202 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for
1203 RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-
1204 0550-8
1205 Luperchio, T. R., Boukas, L., Zhang, L., Pilarowski, G., Jiang, J., Kalinousky, A., Hansen, K. D., &
1206 Bjornsson, H. T. (2021). Leveraging the Mendelian disorders of the epigenetic machinery to
1207 systematically map functional epigenetic variation. *Elife*, *10*.
1208 https://doi.org/10.7554/eLife.65884
1209 Mahmood, T., & Yang, P. C. (2012). Western blot: technique, theory, and trouble shooting. *N Am J*
1210 *Med Sci*, *4*(9), 429–434. https://doi.org/10.4103/1947-2714.100998
1211 Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4:
1212 A fast and versatile genome alignment system. *PLoS Comput Biol*, *14*(1), e1005944.
1213 https://doi.org/10.1371/journal.pcbi.1005944
1214 Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of
1215 occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770.
1216 https://doi.org/10.1093/bioinformatics/btr011
1217 Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V. S.,
1218 Schramm, M., Becher, I., Steidel, M., Noh, K. M., Bergamini, G., Beck, M., Bantscheff, M., &
1219 Savitski, M. M. (2018). Systematic analysis of protein turnover in primary cells. *Nat Commun*,
1220 *9*(1), 689. https://doi.org/10.1038/s41467-018-03106-1
1221 McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an
1222 evaluation on ChIP data. *BMC Bioinformatics*, *11*, 165. https://doi.org/10.1186/1471-2105-
1223 11-165
1224 Noguchi, H., Kimura, A., Murao, N., Namihira, M., & Nakashima, K. (2016). Prenatal deletion of DNA
1225 methyltransferase 1 in neural stem cells impairs neurogenesis and causes anxiety-like
1226 behavior in adulthood. *Neurogenesis (Austin)*, *3*(1), e1232679.
1227 https://doi.org/10.1080/23262133.2016.1232679
1228 Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas,
1229 N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers,
1230 M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein,
1231 genetic, and chemical interactions. *Protein Sci*, *30*(1), 187–200.
1232 https://doi.org/10.1002/pro.3978
1233 Park, D. H., Hong, S. J., Salinas, R. D., Liu, S. J., Sun, S. W., Sgualdino, J., Testa, G., Matzuk, M. M.,
1234 Iwamori, N., & Lim, D. A. (2014). Activation of neuronal gene expression by the JMJD3
1235 demethylase is required for postnatal and adult brain neurogenesis. *Cell Rep*, *8*(5), 1290–
1236 1299. https://doi.org/10.1016/j.celrep.2014.07.060
1237 Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G. W., & Schonhuth, A. (2015).
1238 WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J*
1239 *Comput Biol*, *22*(6), 498–509. https://doi.org/10.1089/cmb.2014.0157
1240 Pei, Z., Wang, B., Chen, G., Nagao, M., Nakafuku, M., & Campbell, K. (2011). Homeobox genes Gsx1
1241 and Gsx2 differentially regulate telencephalic progenitor maturation. *Proc Natl Acad Sci U S*
1242 *A*, *108*(4), 1675–1680. https://doi.org/10.1073/pnas.1008824108
1243 Pereira, J. D., Sansom, S. N., Smith, J., Dobenecker, M. W., Tarakhovsky, A., & Livesey, F. J. (2010).
1244 Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal
1245 and differentiation in the cerebral cortex. *Proc Natl Acad Sci U S A*, *107*(36), 15957–15962.
1246 https://doi.org/10.1073/pnas.1002530107

1247 Persengiev, S. P., Kondova, II, & Kilpatrick, D. L. (1999). E2F4 actively promotes the initiation and
1248 maintenance of nerve growth factor-induced cell differentiation. *Mol Cell Biol*, *19*(9), 6048–
1249 6056. https://doi.org/10.1128/MCB.19.9.6048
1250 Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016).
1251 Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*, *34*(7), 695–
1252 697. https://doi.org/10.1038/nbt.3583
1253 Plaisier, C. L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C. M., Ding, Y., Reiss, D. J.,
1254 Paddison, P. J., & Baliga, N. S. (2016). Causal Mechanistic Regulatory Network for
1255 Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Syst*, *3*(2), 172–186.
1256 https://doi.org/10.1016/j.cels.2016.06.006
1257 Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M.,
1258 Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen,
1259 T. S., Weissman, J. S., & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by
1260 direct guide RNA capture and targeted sequencing. *Nat Biotechnol*, *38*(8), 954–961.
1261 https://doi.org/10.1038/s41587-020-0470-y
1262 Reynisdottir, T., Anderson, K. J., Brinn, A., Franklin, K., Ouyang, J., Snorradottir, A. O., Lutz, C. M.,
1263 Zuberi, A. R., DeLeon, V. B., & Bjornsson, H. T. (2024). In-utero rescue of
1264 neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome. *bioRxiv*,
1265 2024.2007.2019.604339. https://doi.org/10.1101/2024.07.19.604339
1266 Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P.
1267 (2011). Integrative genomics viewer. *Nat Biotechnol*, *29*(1), 24–26.
1268 https://doi.org/10.1038/nbt.1754
1269 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for
1270 differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–
1271 140. https://doi.org/10.1093/bioinformatics/btp616
1272 Rothbart, S. B., & Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications.
1273 *Biochim Biophys Acta*, *1839*(8), 627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001
1274 Rybachuk, O., Kopach, O., Pivneva, T., & Kyryk, V. (2019). Isolation of Neural Stem Cells from the
1275 embryonic mouse hippocampus for in vitro growth or engraftment into a host tissue. *Bio*
1276 *Protoc*, *9*(4), e3165. https://doi.org/10.21769/BioProtoc.3165
1277 Schoof, M., Launspach, M., Holdhof, D., Nguyen, L., Engel, V., Filser, S., Peters, F., Immenschuh, J.,
1278 Hellwig, M., Niesen, J., Mall, V., Ertl-Wagner, B., Hagel, C., Spohn, M., Lutz, B., Sedlacik, J.,
1279 Indenbirken, D., Merk, D. J., & Schuller, U. (2019). The transcriptional coactivator and histone
1280 acetyltransferase CBP regulates neural precursor cell development and migration. *Acta*
1281 *Neuropathol Commun*, *7*(1), 199. https://doi.org/10.1186/s40478-019-0849-5
1282 Sheu, J. R., Hsieh, C. Y., Jayakumar, T., Lin, G. Y., Lee, H. N., Huang, S. W., & Yang, C. H. (2019). HDAC6
1283 dysfunction contributes to impaired maturation of adult neurogenesis in vivo: vital role on
1284 functional recovery after ischemic stroke. *J Biomed Sci*, *26*(1), 27.
1285 https://doi.org/10.1186/s12929-019-0521-1
1286 Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting
1287 DNA cytosine methylation using nanopore sequencing. *Nat Methods*, *14*(4), 407–410.
1288 https://doi.org/10.1038/nmeth.4184
1289 Sun, Z., Xu, X., He, J., Murray, A., Sun, M. A., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., Li, L., Zhu,
1290 J., Chen, J., Morozov, A., Pickrell, A. M., Theus, M. H., & Xie, H. (2019). EGR1 recruits TET1 to
1291 shape the brain methylome during development and upon neuronal activity. *Nat Commun*,
1292 *10*(1), 3892. https://doi.org/10.1038/s41467-019-11905-3
1293 Syal, C., Seegobin, M., Sarma, S. N., Gouveia, A., Hsu, K., Niibori, Y., He, L., Wondisford, F. E.,
1294 Frankland, P. W., & Wang, J. (2018). Ectopic expression of aPKC-mediated phosphorylation in
1295 p300 modulates hippocampal neurogenesis, CREB binding and fear memory differently with
1296 age. *Sci Rep*, *8*(1), 13489. https://doi.org/10.1038/s41598-018-31657-2
1297 Tao, C., Liu, J., Li, Z., Lai, P., Zhang, S., Qu, J., Tang, Y., Liu, A., Zou, Z., Bai, X., & Li, J. (2022). DNMT1 is
1298 a negative regulator of osteogenesis. *Biol Open*, *11*(3). https://doi.org/10.1242/bio.058534

1299 Tycko, B. (2010). Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet*, *19*(R2), R210–
1300 220. https://doi.org/10.1093/hmg/ddq376
1301 Veyrac, A., Besnard, A., Caboche, J., Davis, S., & Laroche, S. (2014). The transcription factor
1302 Zif268/Egr1, brain plasticity, and memory. *Prog Mol Biol Transl Sci*, *122*, 89–129.
1303 https://doi.org/10.1016/B978-0-12-420170-5.00004-0
1304 Wang, J., Weaver, I. C., Gauthier-Fisher, A., Wang, H., He, L., Yeomans, J., Wondisford, F., Kaplan, D.
1305 R., & Miller, F. D. (2010). CBP histone acetyltransferase activity regulates embryonic neural
1306 differentiation in the normal and Rubinstein-Taybi syndrome brain. *Dev Cell*, *18*(1), 114–125.
1307 https://doi.org/10.1016/j.devcel.2009.10.023
1308 Wei, Z., Zheng, S., Spangler, S. A., Yu, C., Hoogenraad, C. C., & Zhang, M. (2011). Liprin-mediated large
1309 signaling complex organization revealed by the liprin-alpha/CASK and liprin-alpha/liprin-beta
1310 complex structures. *Mol Cell*, *43*(4), 586–598. https://doi.org/10.1016/j.molcel.2011.07.021
1311 Wickham, H. (2016). ggplot2. In: Springer International Publishing:.
1312 Wulfridge, P., Langmead, B., Feinberg, A. P., & Hansen, K. D. (2019). Analyzing whole genome
1313 bisulfite sequencing data from highly divergent genotypes. *Nucleic Acids Res*, *47*(19), e117.
1314 https://doi.org/10.1093/nar/gkz674
1315 Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., & Ren, B. (2012). Base-
1316 resolution analyses of sequence and parent-of-origin dependent DNA methylation in the
1317 mouse genome. *Cell*, *148*(4), 816–831. https://doi.org/10.1016/j.cell.2011.12.035
1318 You, L., Yan, K., Zou, J., Zhao, H., Bertos, N. R., Park, M., Wang, E., & Yang, X. J. (2015). The lysine
1319 acetyltransferase activator Brpf1 governs dentate gyrus development through neural stem
1320 cells and progenitors. *PLoS Genet*, *11*(3), e1005034.
1321 https://doi.org/10.1371/journal.pgen.1005034
1322 Zhang, J., Ji, F., Liu, Y., Lei, X., Li, H., Ji, G., Yuan, Z., & Jiao, J. (2014). Ezh2 regulates adult hippocampal
1323 neurogenesis and memory. *J Neurosci*, *34*(15), 5184–5199.
1324 https://doi.org/10.1523/JNEUROSCI.4129-13.2014
1325 Zhang, W., Macias-Velasco, J. F., Zhuo, X., Belter, E. A., Jr., Tomlinson, C., Garza, J., Tekkey, N., Li, D.,
1326 & Wang, T. (2025). methylGrapher: genome-graph-based processing of DNA methylation
1327 data from whole genome bisulfite sequencing. *Nucleic Acids Res*, *53*(3).
1328 https://doi.org/10.1093/nar/gkaf028
1329 Zhou, Q., Obana, E. A., Radomski, K. L., Sukumar, G., Wynder, C., Dalgard, C. L., & Doughty, M. L.
1330 (2016). Inhibition of the histone demethylase Kdm5b promotes neurogenesis and
1331 derepresses Reln (reelin) in neural stem cells from the adult subventricular zone of mice. *Mol*
1332 *Biol Cell*, *27*(4), 627–639. https://doi.org/10.1091/mbc.E15-07-0513

1333

Figure 1

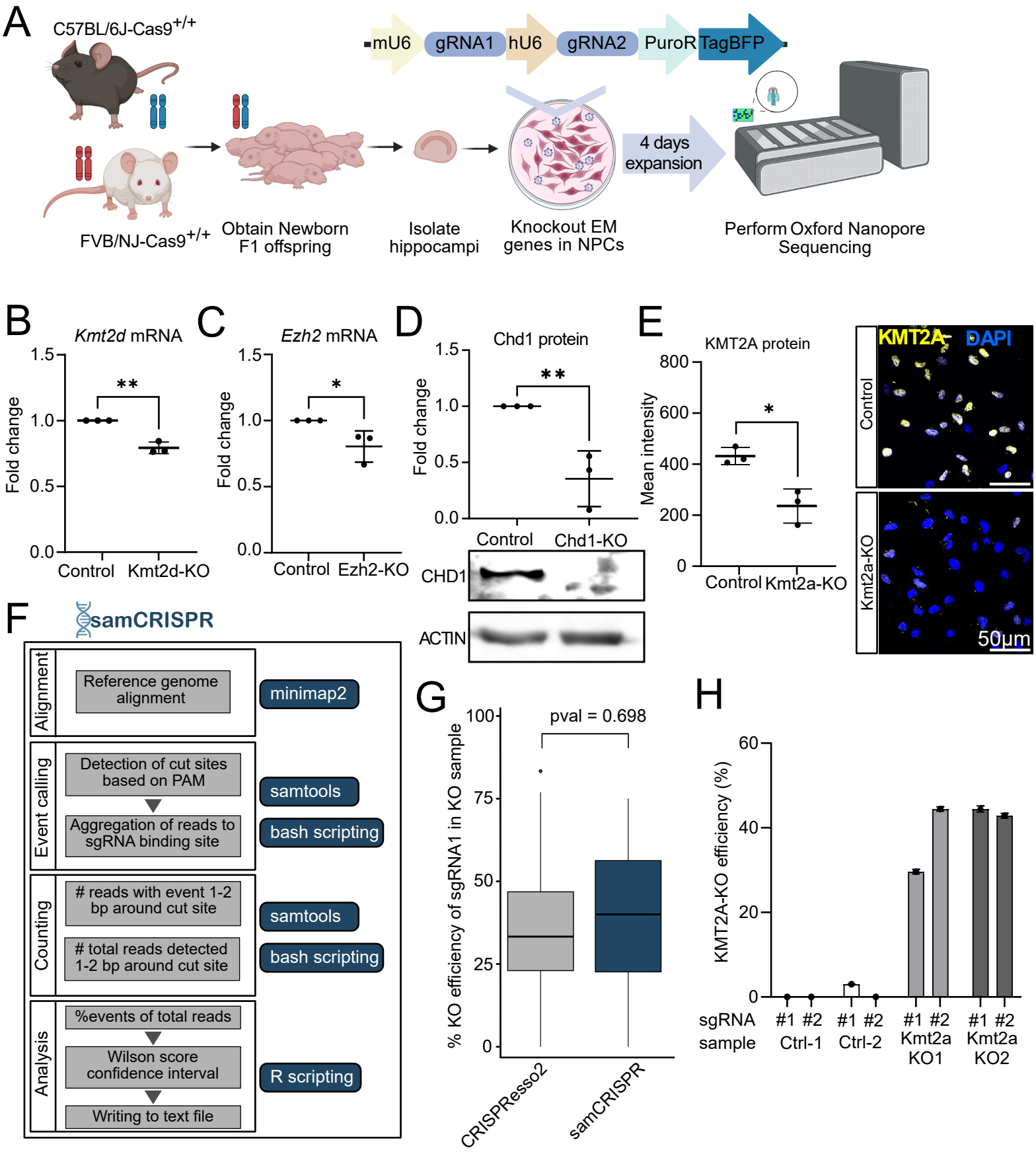

A
C57BL/6J-Cas9+/+
FVB/NJ-Cas9+/+
Obtain Newborn F1 offspring
Isolate hippocampi
mU6
gRNA1
hU6
gRNA2
PuroR
TagBFP
Knockout EM genes in NPCs
4 days expansion
Perform Oxford Nanopore Sequencing
B
Kmt2d mRNA
Fold change
Control
Kmt2d-KO
C
Ezh2 mRNA
Control
Ezh2-KO
D
Chd1 protein
Control
Chd1-KO
CHD1
ACTIN
E
KMT2A protein
Mean intensity
Control
Kmt2a-KO
KMT2A
DAPI
50µm
F
samCRISPR
Alignment
Reference genome alignment
minimap2
Event calling
Detection of cut sites based on PAM
Aggregation of reads to sgRNA binding site
samtools
bash scripting
Counting
reads with event 1-2 bp around cut site
total reads detected 1-2 bp around cut site
samtools
bash scripting
Analysis
%events of total reads
Wilson score confidence interval
Writing to text file
R scripting
G
pval = 0.698
% KO efficiency of sgRNA1 in KO sample
CRISPResso2
samCRISPR
H
KMT2A-KO efficiency (%)
sgRNA
sample
Ctrl-1
Ctrl-2
Kmt2a KO1
Kmt2a KO2

Figure 2

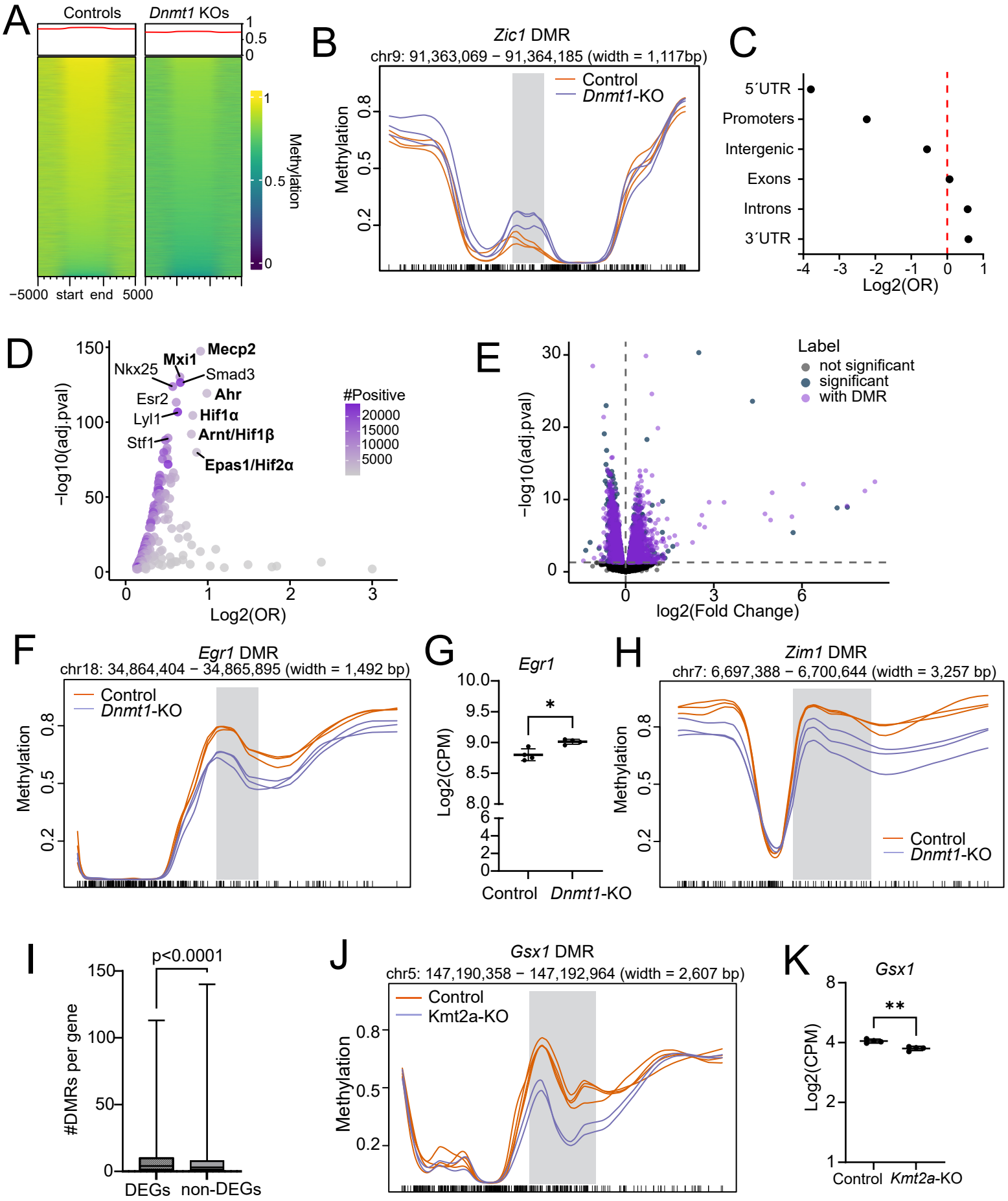

A
Controls
Dnmt1 KOs
Methylation
−5000 start end 5000
B
Zic1 DMR
chr9: 91,363,069 − 91,364,185 (width = 1,117bp)
Control
Dnmt1-KO
Methylation
C
5´UTR
Promoters
Intergenic
Exons
Introns
3´UTR
Log2(OR)
D
Mecp2
Mxi1
Nkx25
Smad3
Ahr
Esr2
Hif1α
Lyl1
Arnt/Hif1β
Stf1
Epas1/Hif2α
#Positive
−log10(adj.pval)
Log2(OR)
E
Label
not significant
significant
with DMR
−log10(adj.pval)
log2(Fold Change)
F
Egr1 DMR
chr18: 34,864,404 − 34,865,895 (width = 1,492 bp)
Control
Dnmt1-KO
Methylation
G
Egr1
Log2(CPM)
Control
Dnmt1-KO
H
Zim1 DMR
chr7: 6,697,388 − 6,700,644 (width = 3,257 bp)
Methylation
Control
Dnmt1-KO
I
p<0.0001
#DMRs per gene
DEGs
non-DEGs
J
Gsx1 DMR
chr5: 147,190,358 − 147,192,964 (width = 2,607 bp)
Control
Kmt2a-KO
Methylation
K
Gsx1
Log2(CPM)
Control
Kmt2a-KO

Figure 3

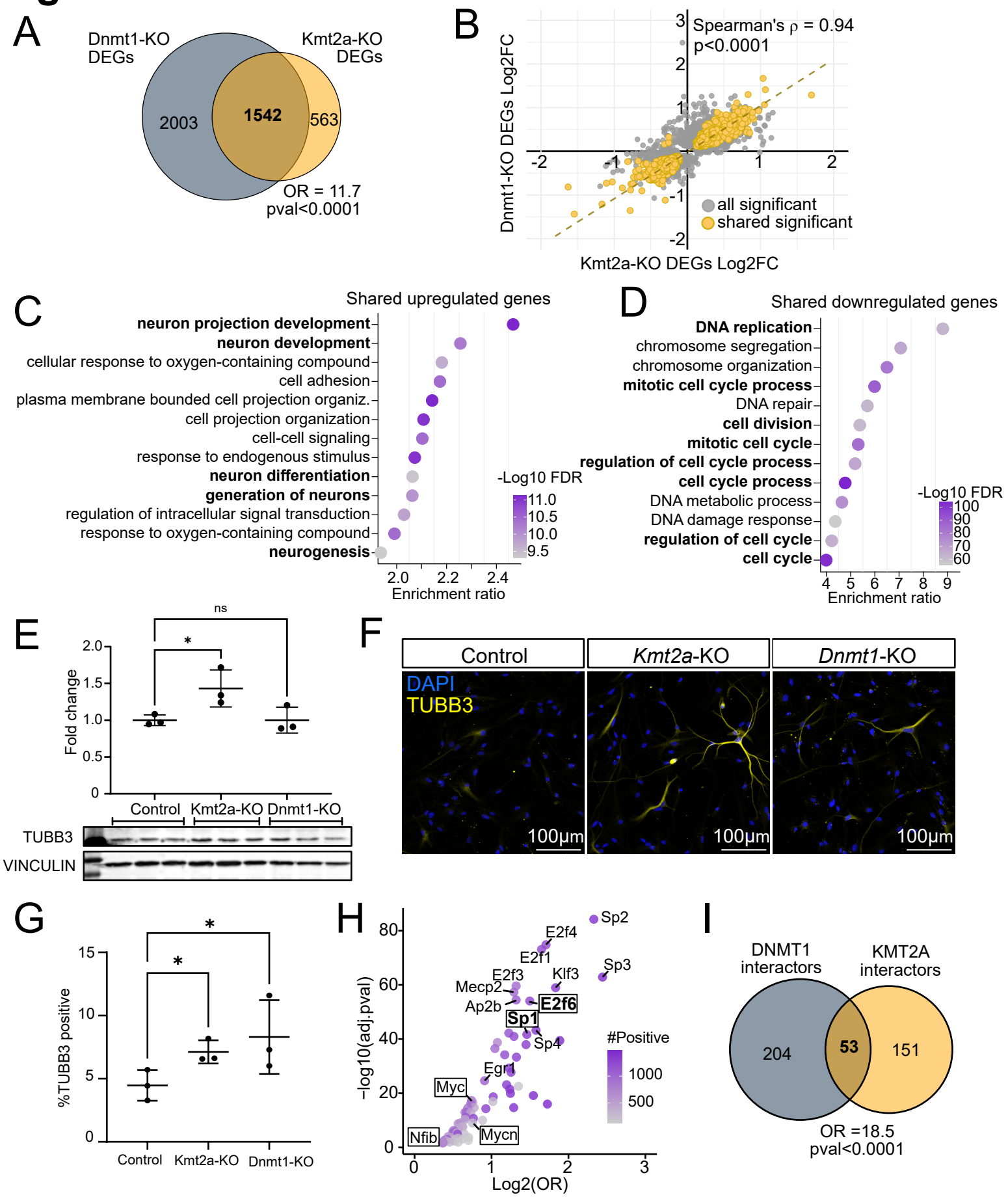

A
Dnmt1-KO DEGs
Kmt2a-KO DEGs
2003
1542
563
OR = 11.7
pval<0.0001
B
Spearman's ρ = 0.94
p<0.0001
Dnmt1-KO DEGs Log2FC
Kmt2a-KO DEGs Log2FC
all significant
shared significant
C
Shared upregulated genes
neuron projection development
neuron development
cellular response to oxygen-containing compound
cell adhesion
plasma membrane bounded cell projection organiz.
cell projection organization
cell-cell signaling
response to endogenous stimulus
neuron differentiation
generation of neurons
regulation of intracellular signal transduction
response to oxygen-containing compound
neurogenesis
-Log10 FDR
11.0
10.5
10.0
9.5
Enrichment ratio
D
Shared downregulated genes
DNA replication
chromosome segregation
chromosome organization
mitotic cell cycle process
DNA repair
cell division
mitotic cell cycle
regulation of cell cycle process
cell cycle process
DNA metabolic process
DNA damage response
regulation of cell cycle
cell cycle
-Log10 FDR
100
90
80
70
60
Enrichment ratio
E
ns
*
Fold change
Control
Kmt2a-KO
Dnmt1-KO
TUBB3
VINCULIN
F
Control
Kmt2a-KO
Dnmt1-KO
DAPI
TUBB3
100µm
G
%TUBB3 positive
Control
Kmt2a-KO
Dnmt1-KO
H
Sp2
E2f4
E2f1
Sp3
E2f3
Klf3
Mecp2
Ap2b
E2f6
Sp1
Sp4
Egr1
Myc
Nfib
Mycn
#Positive
1000
500
−log10(adj.pval)
Log2(OR)
I
DNMT1 interactors
KMT2A interactors
204
53
151
OR =18.5
pval<0.0001

Figure 4

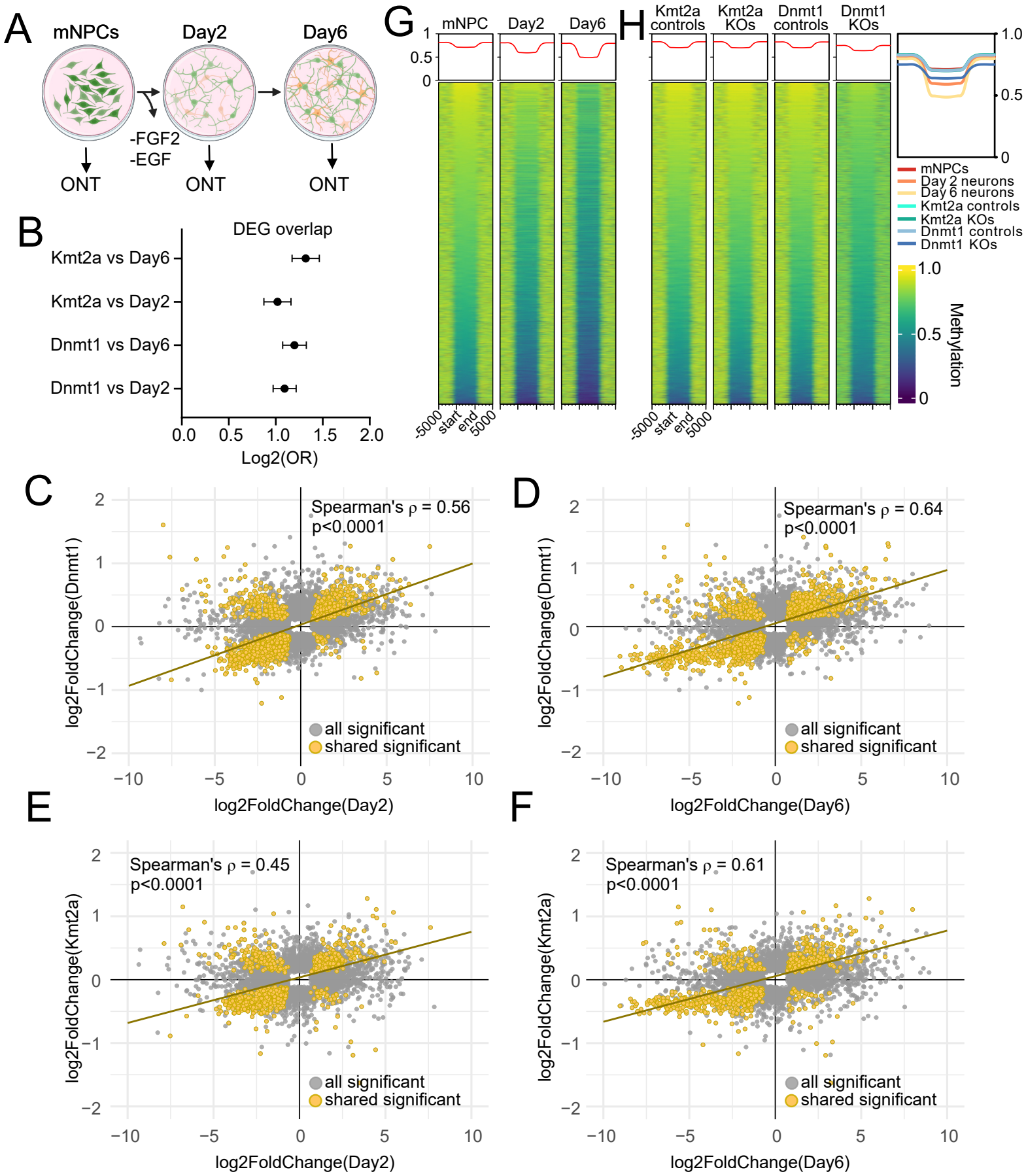

A
mNPCs
Day2
Day6
-FGF2
-EGF
ONT
ONT
ONT
B
DEG overlap
Kmt2a vs Day6
Kmt2a vs Day2
Dnmt1 vs Day6
Dnmt1 vs Day2
0.0
0.5
1.0
1.5
2.0
Log2(OR)
G
mNPC
Day2
Day6
1
0.5
0
-5000
start
end
5000
H
Kmt2a controls
Kmt2a KOs
Dnmt1 controls
Dnmt1 KOs
1.0
0.5
0
mNPCs
Day 2 neurons
Day 6 neurons
Kmt2a controls
Kmt2a KOs
Dnmt1 controls
Dnmt1 KOs
Methylation
C
Spearman's ρ = 0.56
p<0.0001
log2FoldChange(Dnmt1)
log2FoldChange(Day2)
all significant
shared significant
D
Spearman's ρ = 0.64
p<0.0001
log2FoldChange(Dnmt1)
log2FoldChange(Day6)
all significant
shared significant
E
Spearman's ρ = 0.45
p<0.0001
log2FoldChange(Kmt2a)
log2FoldChange(Day2)
all significant
shared significant
F
Spearman's ρ = 0.61
p<0.0001
log2FoldChange(Kmt2a)
log2FoldChange(Day6)
all significant
shared significant

Figure 5

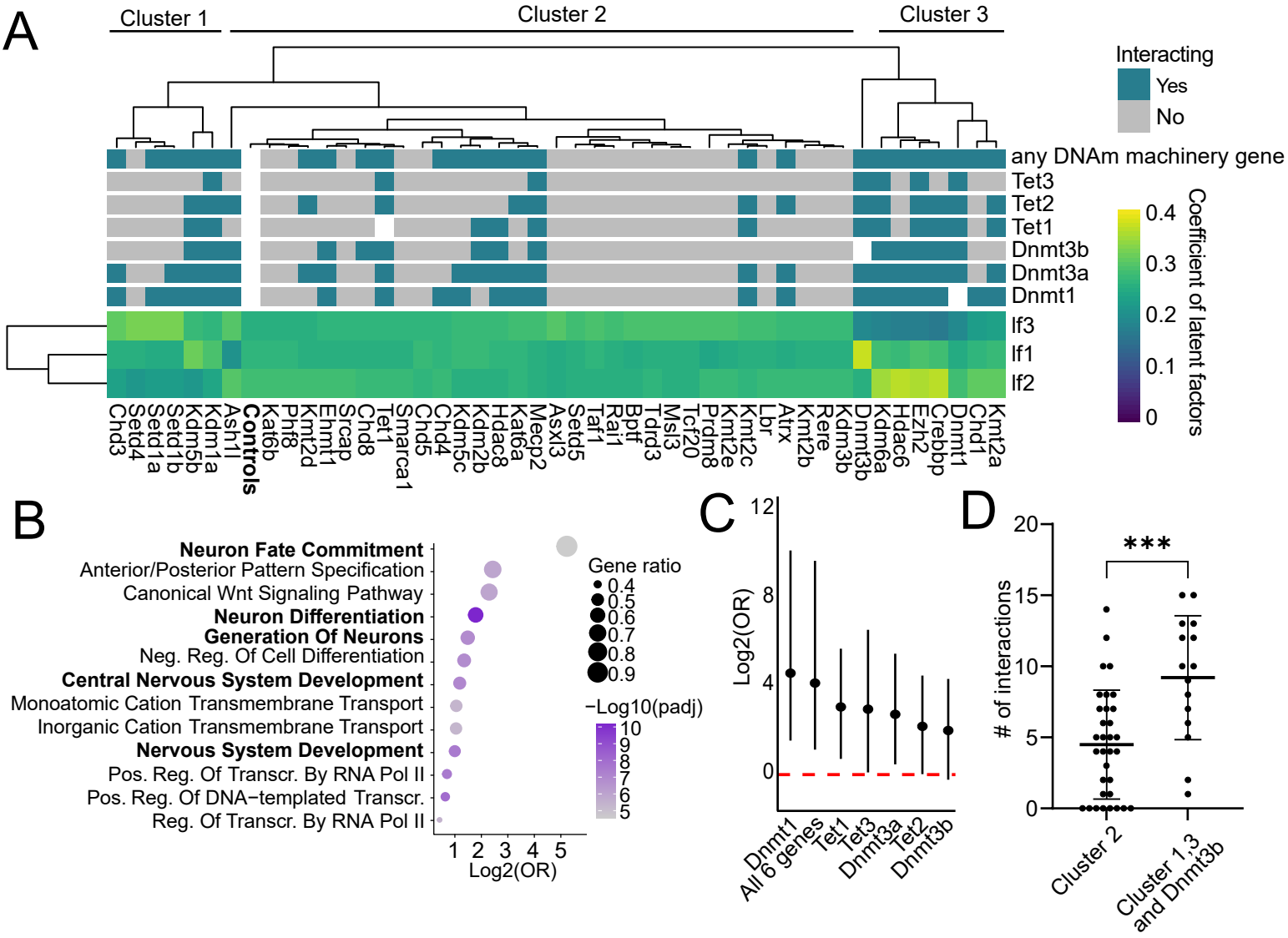
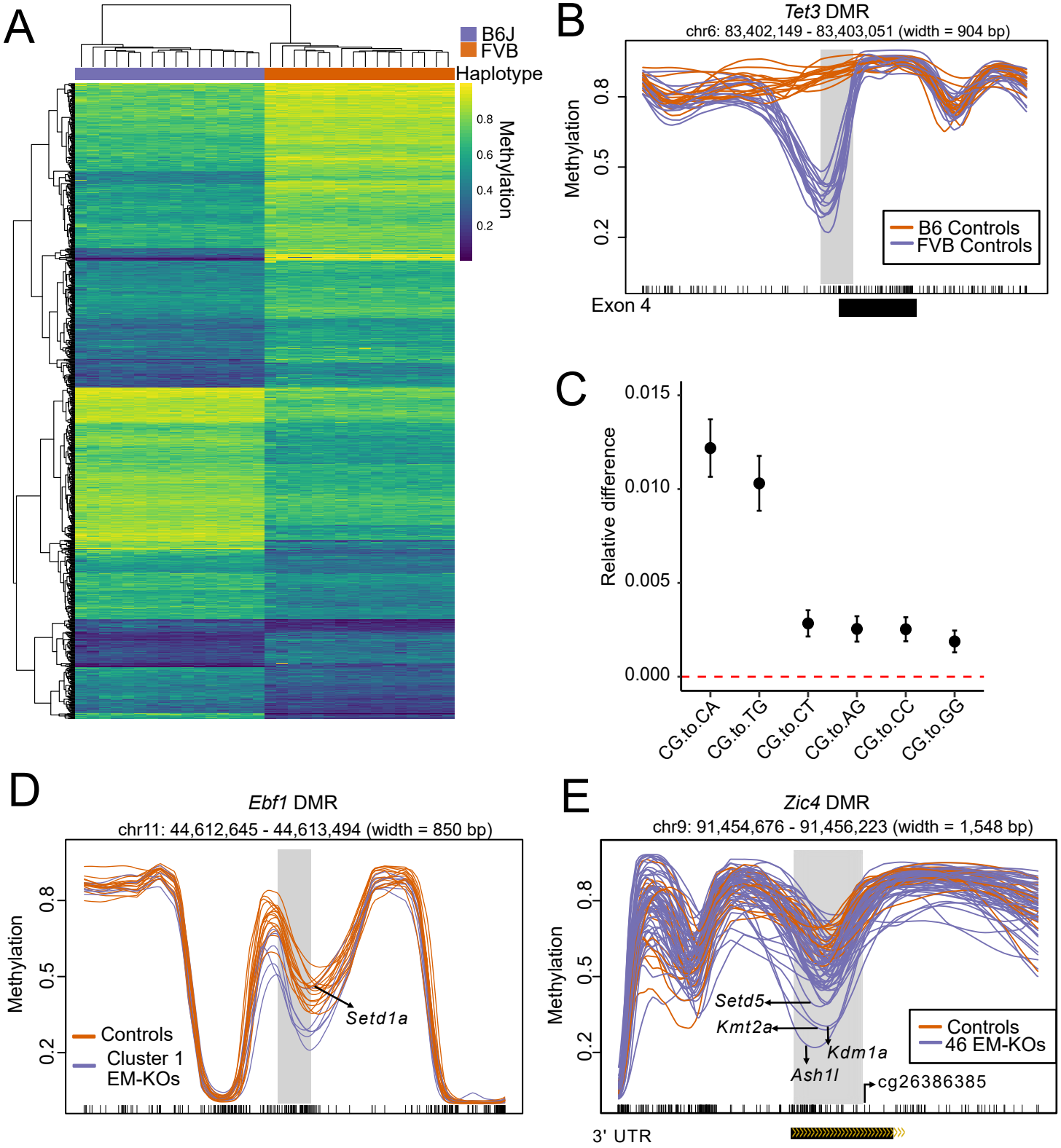
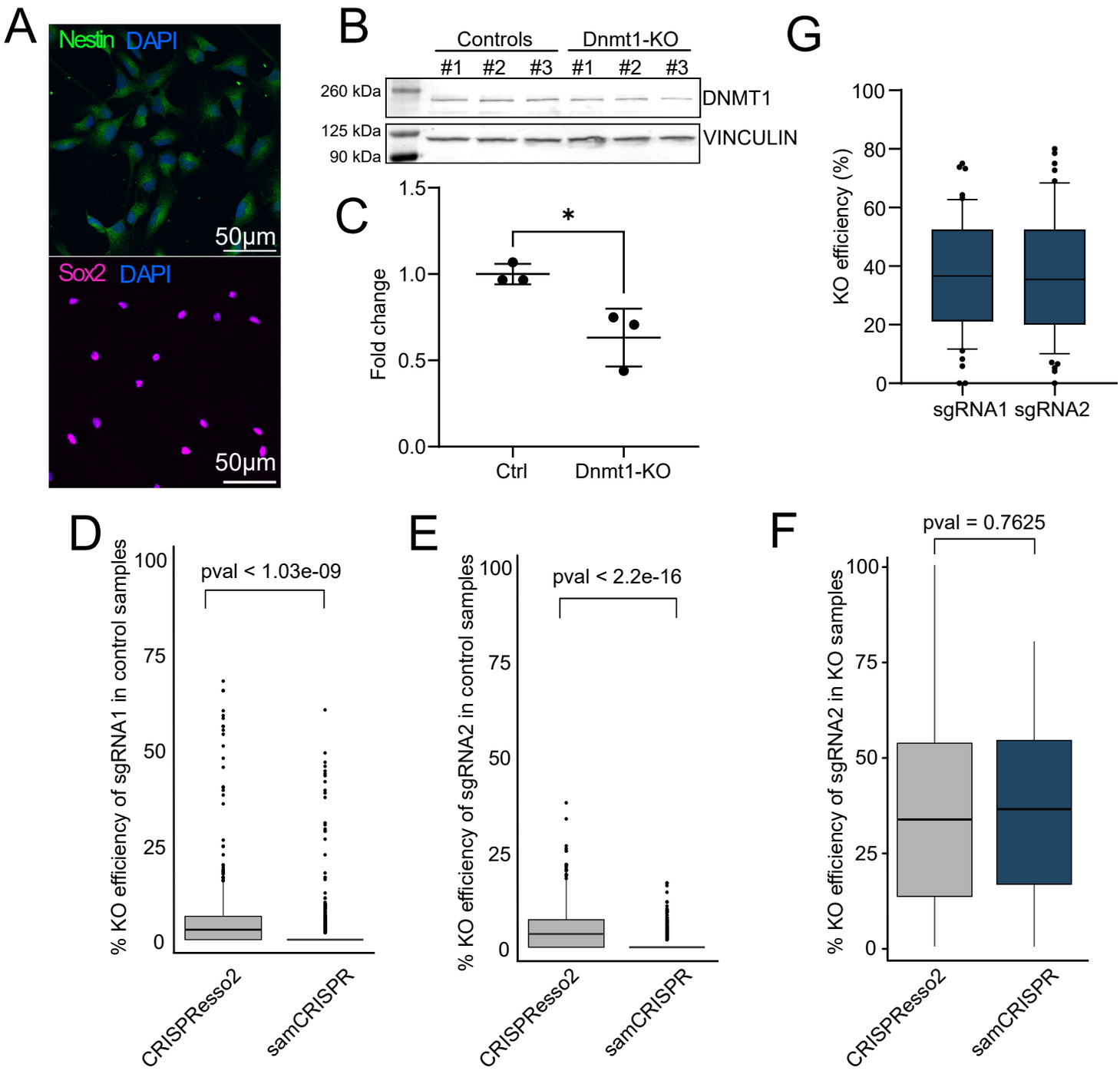

A
Cluster 1
Cluster 2
Cluster 3
Interacting
Yes
No
any DNAm machinery gene
Tet3
Tet2
Tet1
Dnmt3b
Dnmt3a
Dnmt1
If3
If1
If2
Coefficient of latent factors
0.4
0.3
0.2
0.1
0
Chd3
Setd4
Setd1a
Setd1b
Kdm5b
Kdm1a
Ash1l
Controls
Kat6b
Phf8
Kmt2d
Ehmt1
Srcap
Chd8
Tet1
Smarca1
Chd5
Chd4
Kdm5c
Kdm2b
Hdac8
Kat6a
Mecp2
Asxl3
Setd5
Taf1
Rai1
Bptf
Tdrd3
Msl3
Tcf20
Prdm8
Kmt2e
Kmt2c
Lbr
Atrx
Kmt2b
Rere
Kdm3b
Dnmt3b
Kdm6a
Hdac6
Ezh2
Crebbp
Dnmt1
Chd1
Kmt2a
B
Neuron Fate Commitment
Anterior/Posterior Pattern Specification
Canonical Wnt Signaling Pathway
Neuron Differentiation
Generation Of Neurons
Neg. Reg. Of Cell Differentiation
Central Nervous System Development
Monoatomic Cation Transmembrane Transport
Inorganic Cation Transmembrane Transport
Nervous System Development
Pos. Reg. Of Transcr. By RNA Pol II
Pos. Reg. Of DNA−templated Transcr.
Reg. Of Transcr. By RNA Pol II
1 2 3 4 5
Log2(OR)
Gene ratio
0.4
0.5
0.6
0.7
0.8
0.9
−Log10(padj)
10
9
8
7
6
5
C
12
8
4
0
Log2(OR)
Dnmt1
All 6 genes
Tet1
Tet3
Dnmt3a
Tet2
Dnmt3b
D
20
15
10
5
0
# of interactions
***
Cluster 2
Cluster 1,3 and Dnmt3b

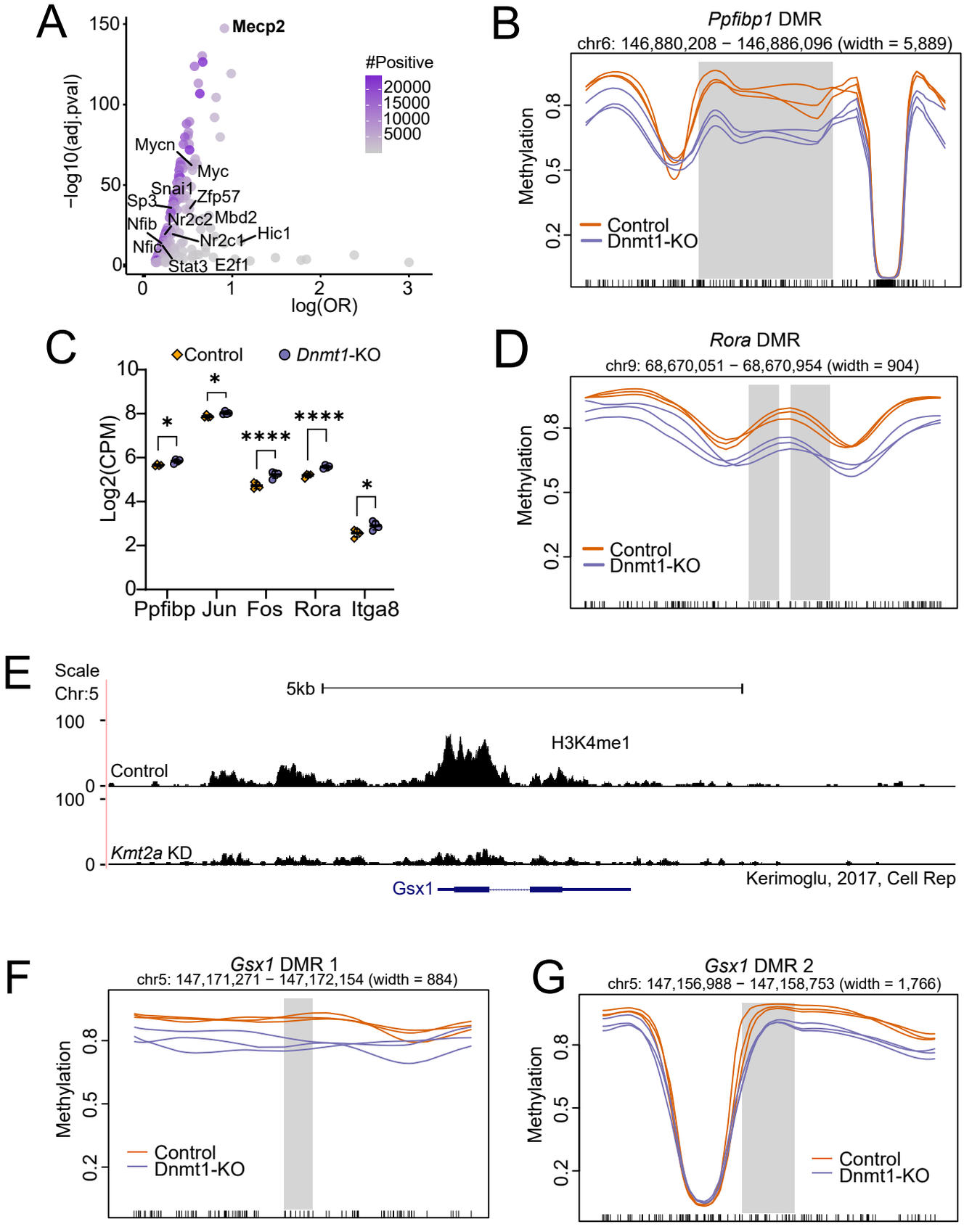
Figure 6


A
B6J
FVB
Haplotype
Methylation
0.8
0.6
0.4
0.2
B
Tet3 DMR
chr6: 83,402,149 - 83,403,051 (width = 904 bp)
Methylation
0.8
0.5
0.2
B6 Controls
FVB Controls
Exon 4
C
Relative difference
0.015
0.010
0.005
0.000
CG.to.CA
CG.to.TG
CG.to.CT
CG.to.AG
CG.to.CC
CG.to.GG
D
Ebf1 DMR
chr11: 44,612,645 - 44,613,494 (width = 850 bp)
Methylation
0.8
0.5
0.2
Controls
Cluster 1
EM-KOs
Setd1a
E
Zic4 DMR
chr9: 91,454,676 - 91,456,223 (width = 1,548 bp)
Methylation
0.8
0.5
0.2
Setd5
Kmt2a
Kdm1a
Ash1l
cg26386385
Controls
46 EM-KOs
3' UTR

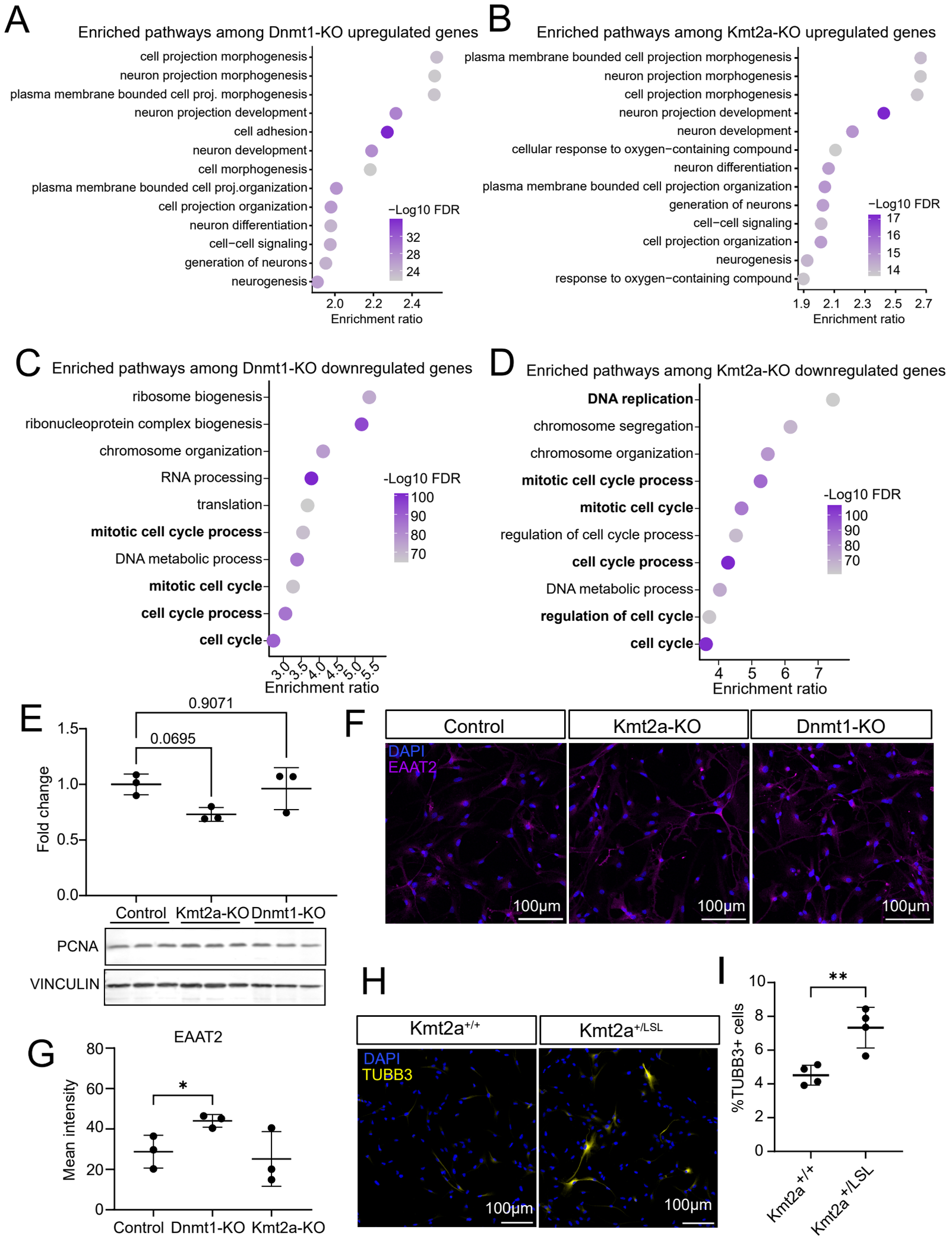
Supplementary Figure 1



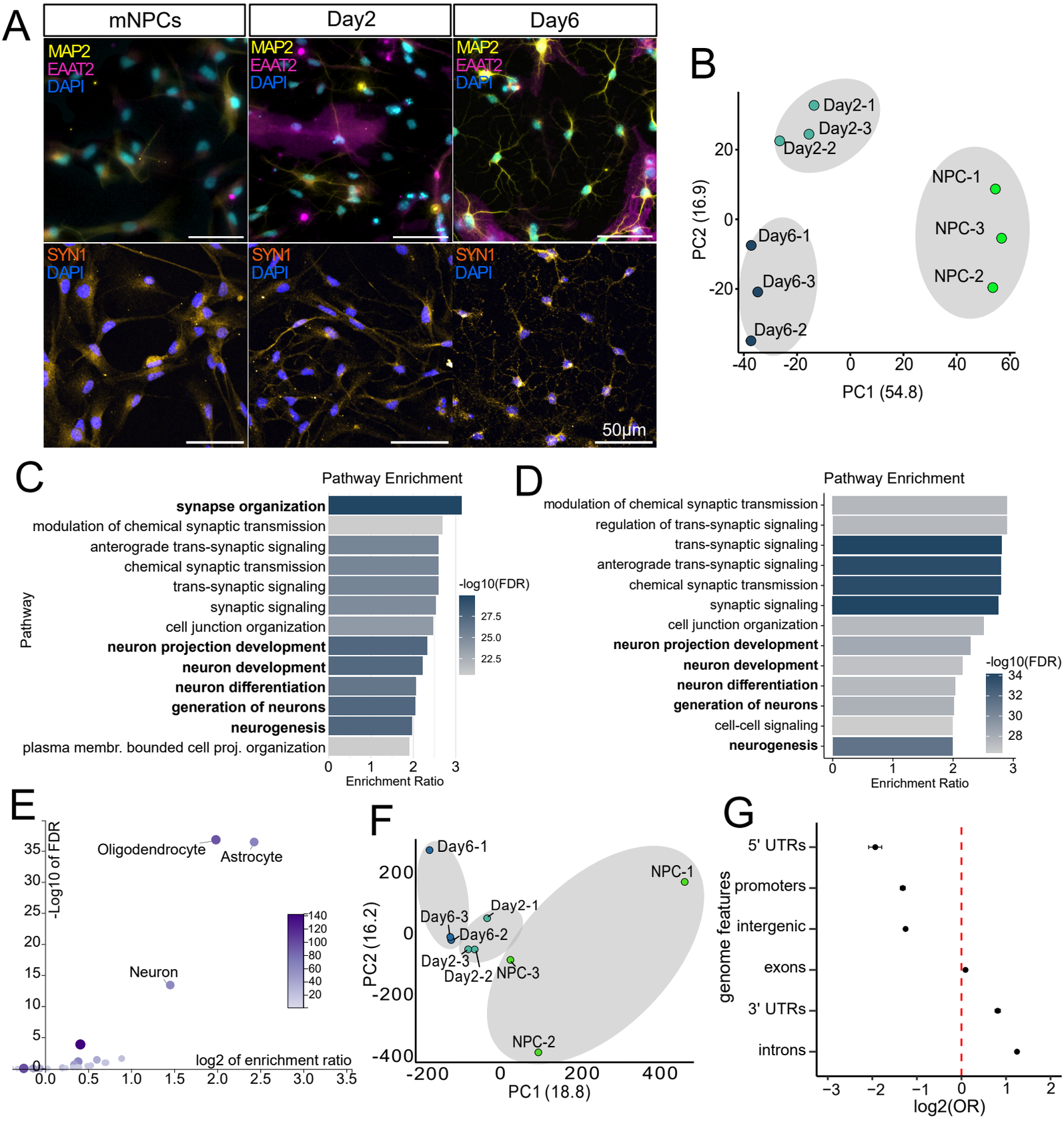
Supplementary Figure 2



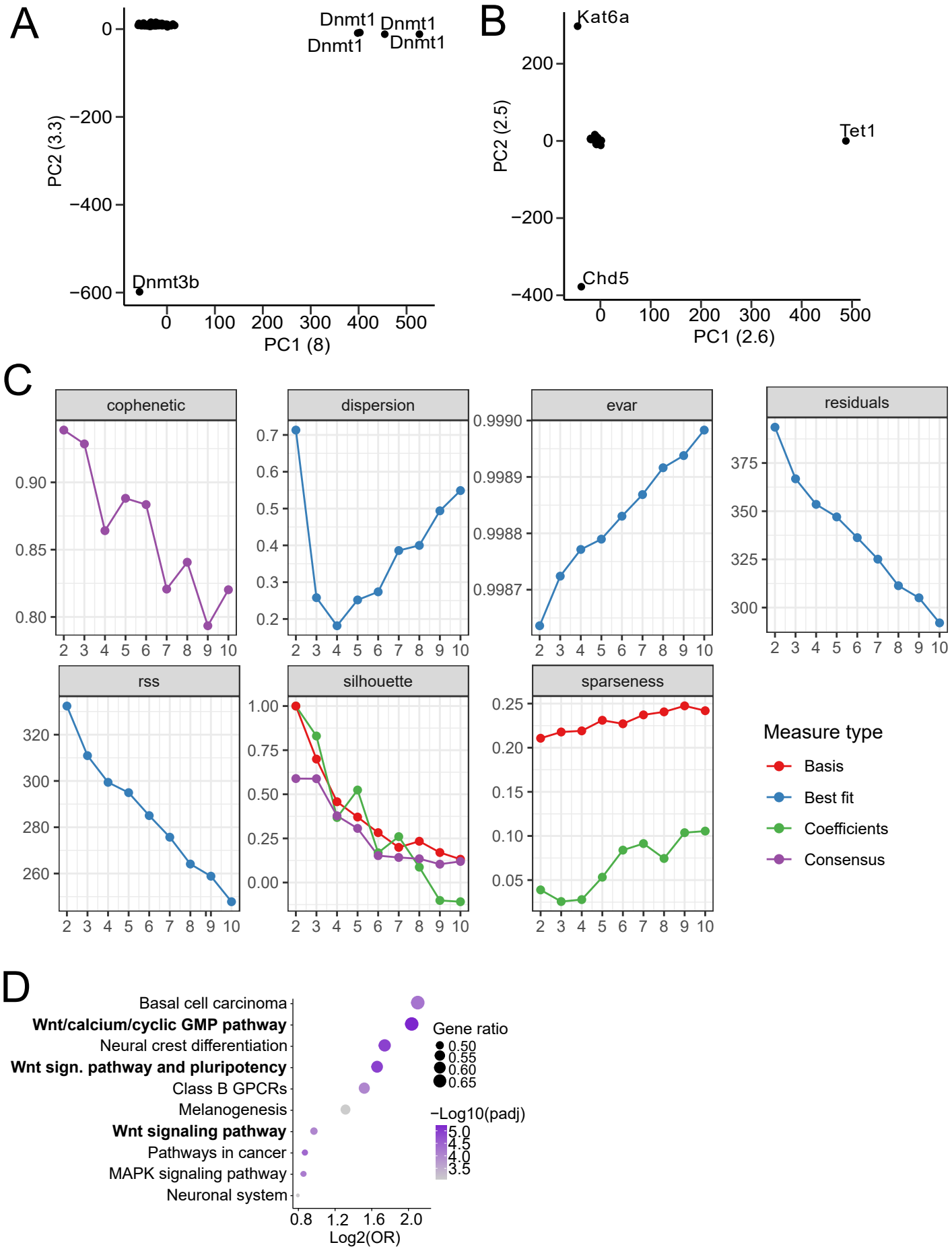
Supplementary Figure 3



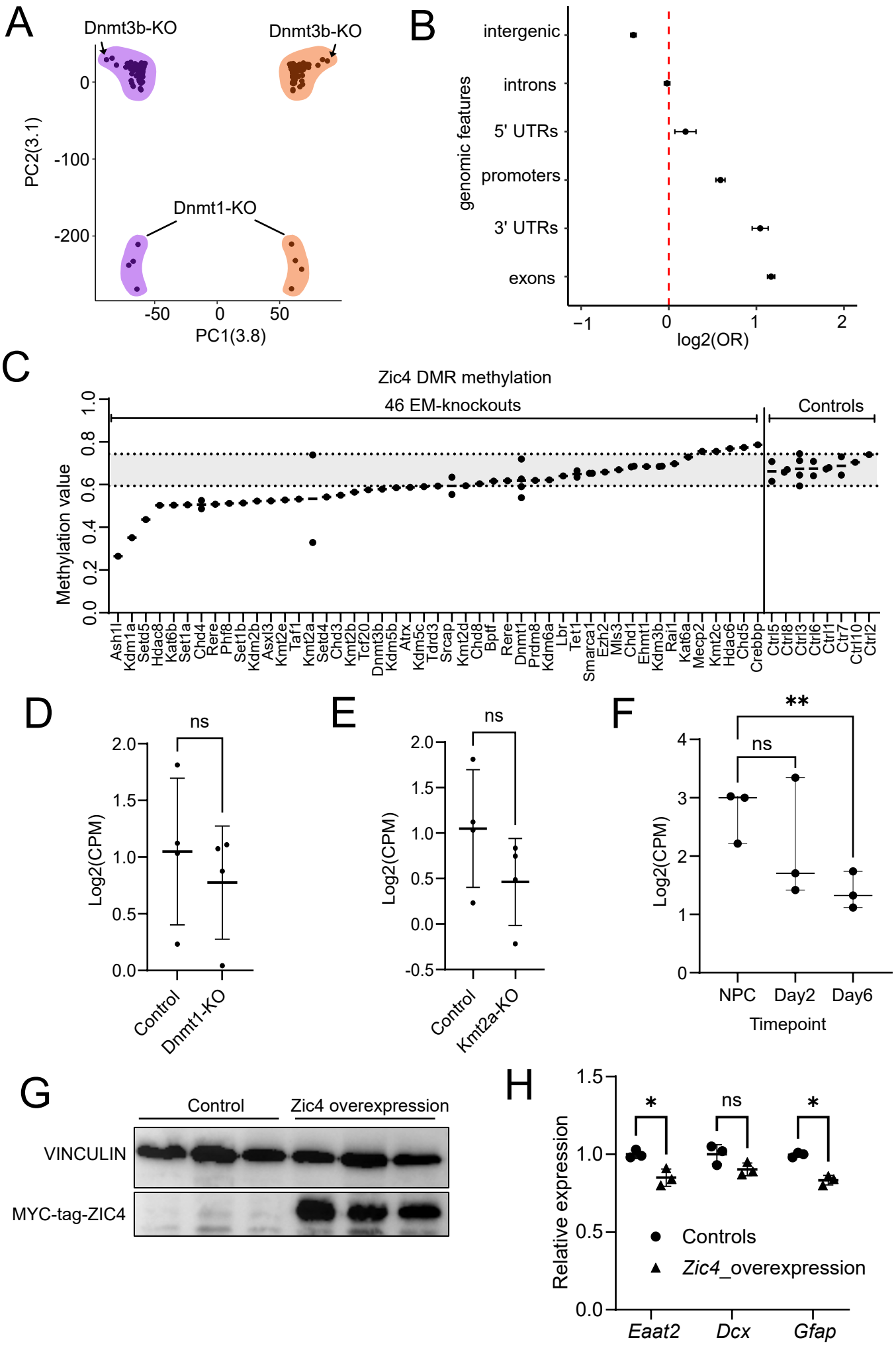
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Supplementary Methods

Dissections and NPCs dissociation

On the day of birth (P0), pups were euthanized by decapitation and the brain carefully dissected out of the skull. The brain was transferred onto a tissue culture dish containing phosphate-buffered saline (PBS, pH 7.4, ThermoFisher, 10010015). Hemispheres were separated and hippocampi isolated from underlying cortex and transferred to ice-cold PBS in a 1.5 ml Eppendorf, keeping each pair separate. Each hippocampal pair was enzymatically dissociated using 0.5 ml 1% TrypLE Select Enzyme (ThermoFisher, A1217701) diluted in PBS followed by a 5 minute (m) incubation at 37°C. Hippocampi were further mechanically dissociated by pipetting 15-20x with P1000 tips (low-adhesion) and incubated for another 5m at 37°C to ensure complete dissociation. TrypLE was inactivated by adding 1 ml Neurobasal Medium. Hippocampi were dissociated further into single-cell suspension by pipetting 15-20x with P1000 then P200 tips sequentially. Cell suspension was passed through a 70 µm cell filter (Miltenyi Biotec, 130-110-916) pre-moistened with 1 ml Neurobasal Medium (ThermoFisher, 21103049) into a 15 ml Falcon tube, centrifuged at 350xg for 5 m at room temperature (RT) to isolate a cell pellet. Cell pellet was resuspended in 1 ml neuronal growth medium (NGM), which included Neurobasal medium supplemented with 1% B27 supplement (ThermoFisher, 17504044), 1% Glutamax (ThermoFisher, 35050061), 0.25% Penicillin-streptomycin (ThermoFisher, 15140122) 20 ng/ml EGF (PeproTech, AF-100-15), 20 ng/ml FGF2 (PeproTech, 100-18B) and 2 µg/ml Heparin (MP Biomedicals, 0210193125), and plated in 1 well each pair on a Matrigel-coated 12-well plate (BD Biosciences, 354234, diluted 1:100 in DMEM/F12). In addition, 10 µM ROCK inhibitor (Tocris, 1245) was added to the NGM to enhance survival for the first 24 hours. The following day, cell debris was washed off 2x with PBS and replaced with 1 ml fresh NGM. After culturing for 48 h (day 3), cells were fed by half medium exchange with replacing 0.5 ml neuronal growth medium, then followed by complete media changes every 48 h, until cells were 75-85% confluent. NPCs (passage 0) were dissociated with Accutase (Sigma-Aldrich, A6964) and seeded on pre-coated plate, including 5 µg/ml PDL (Sigma-Aldrich, P7280) overnight at RT and 5-10 µg/ml Laminin (Roche Diagnostics, 11243217001) for at least 4 h at 37°C. NPCs (passage 2) were frozen down in freezing medium (NGM supplemented with 10% DMSO (Sigma-Aldrich, D8418)), kept in -80°C freezer for a few days then moved to liquid nitrogen for long-termstorage.

ONT-DNA sequencing library construction

For sequencing we used 1µg sheared genomic HMW DNA in 47 µl nuclease-free water (ThermoFisher, Am9937) and followed the instructions from the ONT Ligation Sequencing kit V14 (ONT, SQK-LSK114) exactly. The sample was then mixed with 3.5 µl NEBnext FFPE DNA Repair Buffer (NEB, M6630), 2 µl NEBnext FFPE DNA Repair Mix (NEB, M6630), 3.5 µl Ultra II End-prep reaction buffer (NEB, E7546), and 3 µl Ultra II End-prep Enzyme Mix (NEB, E7546) in a 0.2 ml thin-walled PCR tube (Azenta, 4ti-0792). The reaction was thoroughly mixed by pipetting 15-20 times with P200 tips and incubated at 20°C for 5 m and 65°C for 5 m in a thermal cycler (ThermoFisher Scientific). Samples were transferred into 1.5 ml Eppendorf DNA LoBind tubes (Eppendorf, 30108051), and mixed with 60 µl AMPure XP Beads (AXP) to clean and size select DNA, followed by incubating on a rotator mixer at 10 rpm for 15 m at RT. Samples were kept on the magnet, supernatant was carefully pipetted off and then washed with 500 µl of fresh 80% ethanol twice for 0.5-1 m without disturbing the pellet. Samples were spun down, placed on the magnet and gently pipetted off the residual ethanol and dried for 30 s. Samples were removed from magnet, resuspended in 61 µl nuclease-free water, gently mixed by flicking the tube, incubated for 10 m at RT. Samples were pelleted on a magnet for 2 m. Then 61 µl of clear and colorless samples was obtained and moved into a clean Eppendorf DNA LoBind tube (Eppendorf, 30108051). The ligation reaction was prepared by adding reagents in the following order: 5 µl Ligation Adapter, 25 µl Ligation buffer, 60 µl samples from previous step and 10 µl NEBNext Quick T4 DNA Ligase (NEB, M0202M), followed by gently mixing and a 10 m incubation at RT. Then samples were added 40 µl AXP with gently flicking to mix and incubated on a rotator mixer at 10 rpm for 10 m at RT, followed by spinning down, pelleting on a magnetic rack for 5 m and pipetting off the supernatant. Samples next were washed twice with 250 µl Long Fragment buffer by gently flicking to resuspend the pellet, returned to the magnetic rack to form pellet for 2 m, and pipetted off supernatant. Samples were spun down, placed on the magnet, pipetted off any residual supernatant, dried for 30 s and then removed from the magnetic rack and resuspended the pellet in 25 µl Elution buffer, followed by 30 m incubation on a rotator at RT. Samples were pelleted on a magnet for 2 m and DNA samples in supernatant were transferred into clean Eppendorf DNA LoBind tubes. The concentration of samples was measured with Qubit dsDNA HS assay. The total amount of samples (10 fmol) for making library was determined by the size of samples based on an online calculator of Weight to Molar Quantity for nucleic acids (bioline). Samples for loading were prepared as follows: 100 µl sequencing buffer and 68 µl library solution and 32 µl samples (10 fmol) diluted in elution buffer.

Dissections and primary NPCs isolation

Cells were cultured in neuronal growth medium (NGM), including Neurobasal medium supplemented with 1% B27 supplement (ThermoFisher, 17504044), 1% Glutamax (ThermoFisher, 35050061), 0.25% Penicillin-streptomycin (ThermoFisher, 15140122) 20 ng/ml EGF (PeproTech, AF-100-15), 20 ng/ml FGF_2 (PeproTech, 100-18B) and 2 µg/ml Heparin (MP Biomedicals, 0210193125), and initially plated on a Matrigel-coated 12-well plate (BD Biosciences, 354234, diluted 1:100 in DMEM/F12). In addition, 10 µM ROCK inhibitor (Tocris, 1245) was used for the first 24 hours. After that medium was replaced every two days and cells grown to 60-80% confluency, before passaging using Accutase (Sigma-Aldrich, A6964). After first passage, cells were kept on pre-coated PDL/Laminin plates. After 2 passages, NPCs were frozen down in freezing medium (NGM supplemented with 10% DMSO (Sigma-Aldrich, D8418)), kept in a -80°C freezer for a few days then moved to liquid nitrogen for long-term storage. To minimize variation relating to different sex chromosome complements, we only used female NPCs. Sex was verified by isolating DNA from tails and performing sex-determining PCR (McFarlane et al., 2013) (**Supplementary Table 11**). The isolated cells express typical NPC markers (NESTIN, SOX2, **Supplementary Figure 1A**) and were cultured in B-27 medium with FGF2 and EGF which maintained them in a progenitor state until actively differentiated.

References

McFarlane, L., Truong, V., Palmer, J. S., & Wilhelm, D. (2013). Novel PCR assay for determining the genetic sex of mice. *Sex Dev*, *7*(4), 207–211. https://doi.org/10.1159/000348677