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]Caltech, 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 have been repeatedly found to be an effected cell type for these disorders. To directly probe the effect of disruption of the epigenetic machinery, we established a murine neuronal hippocampal differentiation model which allows rapid and systemic knockout (KO) of diverse epigenetic machinery (EM) factors. This system allows us to interrogate the consequences of each KO using ONT DNA methylation sequencing and RNA sequencing. Upon KO of KMT2A (a histone methyltransferase) or DNMT1 (a DNA methyltransferase), we observed a striking convergence in gene expression. In contrast, KO of DNMT1 lead to extensive changes in DNA methylation whereas KO of KMT2A had little-to-no effect on DNA methylation. In both models, KO accelerated neuronal differentiation, and the convergence in gene expression appeared linked to this differentiation state. The transcription factor GSX1 was differentially methylated, differentially expressed and ~20% of its known down-stream targets were differentially expressed genes. Extending our methylation analysis to 46 EM genes, we found that loss of DNA methyltransferases caused the most profound DNAm alterations, whereas other genes induced more subtle effects. Clustering of promoter DNAm changes revealed two EM subgroups that collectively showed a marked enrichment of interactions with the DNAm machinery. Through allele-specific phasing, we identified a differentially methylated region (DMR) at *Zic4* that was shared across the entire group of EM KOs, as well as 1074 haplotype-specific DMRs. Together, our results highlight shared mechanisms across diverse EM mutations and provide new insight into the molecular etiology of MDEMs.

Keywords: Epigenetics, chromatinopathies, histone machinery, intellectual disability, NPCs

Introduction:

The past 15 years have witnessed the rapid discovery of a group of at least 90 distinct disorders, mostly caused 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., 2023).

A defining feature of these disorders is that the majority of these (42) have been shown to have distinct blood cell 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 four 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 &

77 Strahl, 2014) and DNAm can interfere with transcription factor binding (Heberle & Bardet,
78 2019). Therefore, it may not be surprising that systematic KO of EM factors may reveal key
79 TFs whose role in neuronal development are modulated by DNA methylation dynamics. Our
80 approach, which allows phasing of the genomes, also highlights the considerable effect that
81 genetic variation has on the extent of the observed abnormalities in either strain genome.
82 Together, our work lays the groundwork for understanding how the epigenetic machinery
83 helps orchestrate hippocampal neurogenesis and how disruption of this process through EM
84 loss uniformly leads to learning and neurological defects.

85

86 **Results:**

87 **Systematic KO of EM factors in a single neuronal system**

88 To comprehensively study the role of DNAm abnormalities in EM-deficient neurons,
89 we utilized CRISPR-Cas9 technology to knock-out (KO) EM genes in mNPCs. To obtain
90 homogenous and efficient CRISPR-based gene KOs, we used F1 offspring from two fully
91 sequenced mouse-strains (C57BL/6J and FVB) both expressing transgenic Cas9-GFP protein
92 from the ROSA locus, a well-known epigenetic safe harbor region (Friedrich & Soriano,
93 1991; Irion et al., 2007). We isolated hippocampal mNPCs at postnatal day 0 (P0) from this
94 F1-outcross and grew in culture media that helped maintain the mNPC state. We verified cell
95 identities (**Supplementary Figure 1A**) prior to using these for individual EM-KO
96 experiments using 46 dual guide-RNA (sgRNA) lentiviral plasmids (**Supplementary Table**
97 **1**), each containing two sgRNA sequences targeting the same gene. We also created 8 sgRNA
98 plasmids with random sequences (scramble gRNAs, **Supplementary Table 1**). These guide-
99 vectors were individually packaged and transduced using lentiviral delivery into the Cas9-
100 expressing mNPCs (**Figure 1A**). Cells were selected for stable integration of a construct and
101 four days after transduction, harvested for Oxford Nanopore Technologies (ONT) sequencing
102 or RNA sequencing (see methods).

103 To confirm the efficiency of the KO in our system, we performed qPCR, western blot
104 (WB), and immunofluorescence (IF) assays on individual EM-KO mNPCs. As expected, four
105 days after transduction we routinely saw a decrease in mRNA and protein levels for
106 individual EM KOs. As an example, we observed significant ($p < 0.05$) downregulation of
107 *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 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 and 1.4% vs 4.5% for sgRNA2, $p < 0.01$; **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**). This is expected to underestimate the actual KO efficiency in both methods, 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 the current evaluation of long range sequencing.

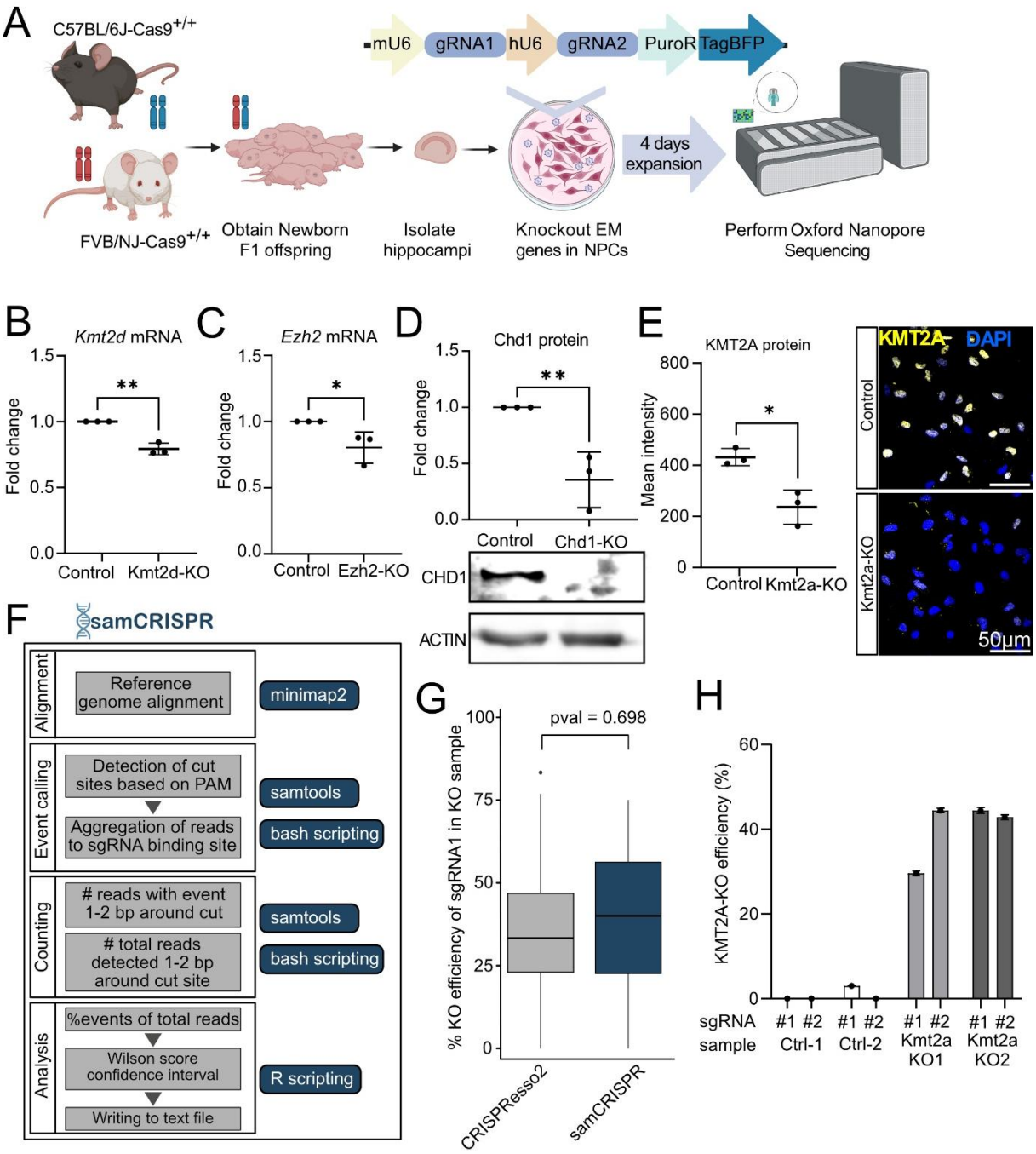


Figure 1: Effective CRISPR-Cas9 targeting in a disease relevant cellular system. (A) Schematic workflow of EM-KO in mNPCs using lentiviral vectors followed by evaluation of DNAm with ONT sequencing. (B) Gene expression changes of *Kmt2d* (p=0.0013) and (C) *Ezh2* (p=0.0229) in mNPCs after knocking out *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 each sgRNA

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

143 **DNAm changes in *Dnmt1*- and *Kmt2a*-KO mNPCs**

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

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

172 To investigate whether methylation changes in *Dnmt1*-KOs coincide with changes in
173 transcription, we performed RNA-seq on *Dnmt1*-KO cells which resulted in 3,454
174 differentially expressed genes (DEGs) compared with controls (**Supplementary Table 5**). Of
175 these, 2,713 DEGs were in close proximity to 26,909 DMRs (**Figure 2E, purple**). Among
176 these were the gene *Ppfibp1* (**Supplementary Figure 2B**), encoding a factor involved in
177 synaptic protein-protein interactions (PPIs) (Wei et al., 2011), the immediate early genes *Jun*,
178 *Fos* and *Egr1* (**Figure 2F-G, Supplementary Figure 2C)** (Kim et al., 2018; Veyrac et al.,
179 2014), as well as *Rora* and *Itga8* (**Supplementary Figure 2C,D**), known DNMT1 target
180 genes (Tao et al., 2022). We also found that mouse imprinted genes (Isles, 2022) were
181 significantly overrepresented in the *Dnmt1*-KO associated DEGs (Fisher's exact test, OR =
182 1.83, p=0.0118. and this enrichment was even higher when we focused only on the subset of
183 DMR-associated DEGs (Fisher's exact test, OR = 2.13, p=0.003). This included gene such as
184 *Zim1* (**Figure 2H**). While DEGs were not overlapping with DMRs more frequently than other
185 detected genes (Fischer's exact test, OR = 1.07, p=0.069), DEGs carried on average more
186 DMRs than non-DEGs (7.6 vs 6.4, respectively, Wilcoxon rank sum test, $p=3.57x10^{-06}$ **Figure**
187 **2I**). Together, this suggests that the wide-spread methylation changes we observe after
188 *Dnmt1*-KO have relatively little impact on gene expression at this time point in our system
189 other than perhaps a subtle effect in intronic and intergenic regions.

190 We next examined KMT2A, a histone methyltransferase which is expected to have a more
191 indirect influence on DNAm. Notably, pathogenic variants in *KMT2A* are the cause of
192 Wiedemann-Steiner syndrome (OMIM, 605130) which has a distinct DNAm signature in
193 peripheral blood (Foroutan et al., 2022). To investigate whether *Kmt2a* loss similarly perturbs
194 methylation in neurons, we performed ONT sequencing on *Kmt2a*-KO NPCs in replicates,
195 which revealed only three genome-wide significant DMRs compared to controls. One DMRs
196 are located in the 3'UTR of the *Gsx1* gene (mm10 coordinates, meanDiff = 0.001121656,
197 qval = 0.042, **Figure 2J**). We also performed RNA-Seq after *Kmt2a* KO and found 2105
198 differentially expressed genes compared with controls (**Supplementary table 6**). Among the
199 previously identified DMRs, only one had significant changes in gene expression, *Gsx1*
200 (**Figure 2K**). GSX1 is a TF expressed early in NPCs and is known to promote maturation in
201 specific neuronal subtypes (Coltogirone et al., 2023; Pei et al., 2011). Notably, a prior study
202 reported reduced H3K4me1 near this DMR in *Kmt2a* knockdown neurons derived from a
203 mouse model (Kerimoglu et al., 2017) (**Supplementary Figure 2E**), supporting that *Gsx1*
204 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 significant (Fischer's exact test, OR = 1.18 and p = 0.055). For the *Dnmt1*-KO experiment, GSX1 was not differentially expressed, but its downstream targets were significantly enriched amongst differentially expressed genes (Fisher's exact test, OR=1.16, p=0.044). Finally, unlike what was found for DNMT1, 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.

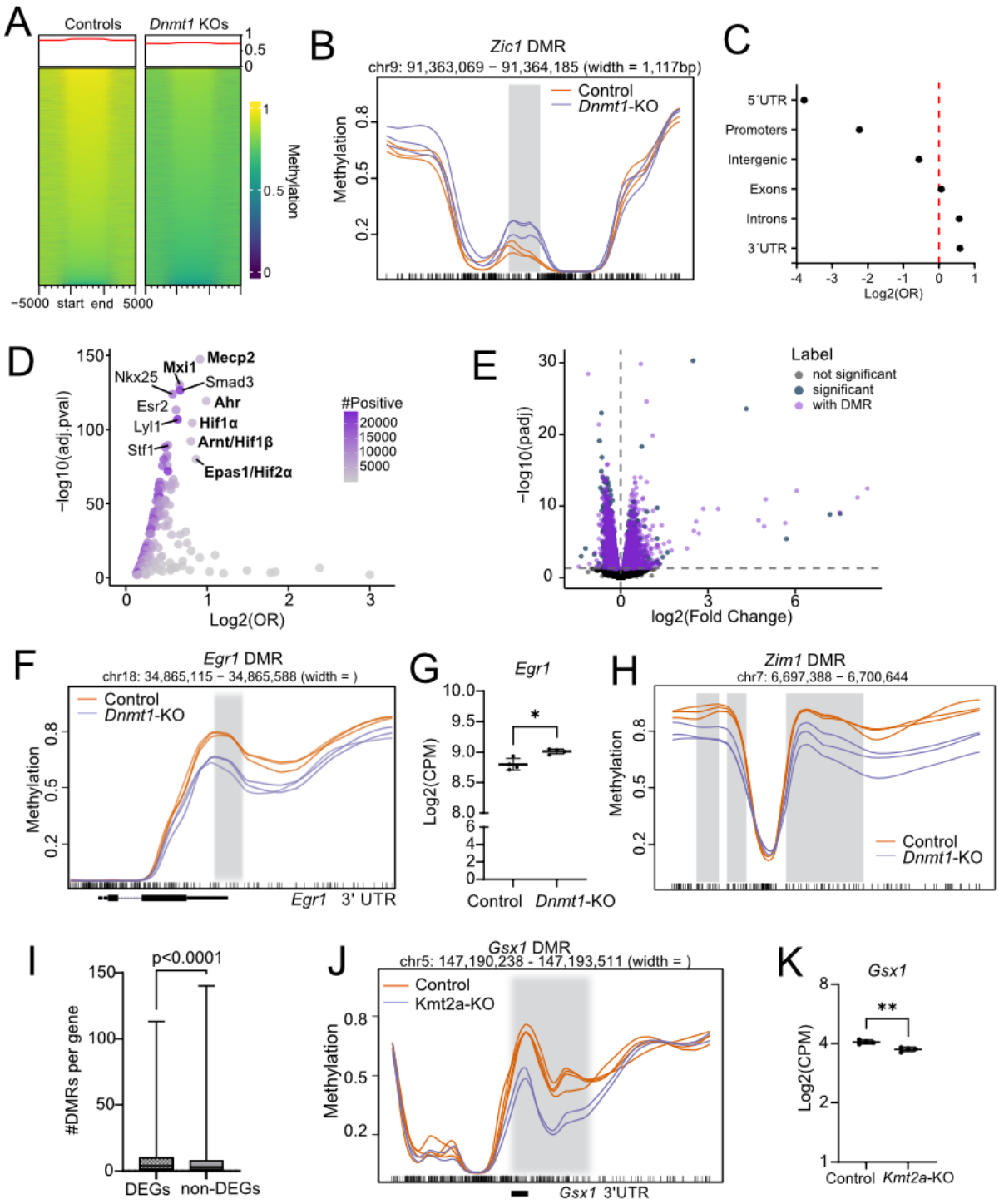


Figure 2: Extent and impact of DNA methylation changes in *Dnmt1*-KO and *Kmt2a*-KO mNPCs. (A) Enrichment heatmaps of smoothed CpG methylation across 141,722 *Dnmt1*-associated DMRs. Each horizontal line represents the average smoothed methylation profile across CpG sites within a DMR and ±5kb flanking regions. Yellow shows higher smoothed methylation but blue shows lower smoothed methylation. (B) The *Zic1* DMR (qval < 0.05) in *Dnmt1*-KOs (purple) compared to control NPCs (orange) highlighted in gray. 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

225 listed in **Supplementary Table 3**. (D) Scatter plot showing enriched TF motifs in *Dnmt1*-KO
226 DMRs (padj < 0.05). (E) A volcano plot showing the distribution of *Dnmt1*-KO DEGs, with
227 significant genes (padj<0.05) shown in dark blue, and DEGs with DMRs associated shown in
228 purple. (F) *Egr1* DMR (qval < 0.05) in *Dnmt1*-KOs (purple) compared to control NPCs
229 (orange) highlighted in gray. Flanking regions of DMR are ±5kb. In x-axis, each vertical
230 black line represents a CpG in mouse genome. (G) Significant upregulation of *Egr1* in
231 *Dnmt1*-KO compared to controls, p < 0.05. (H) Three *Zim1* DMRs (qval < 0.05) in *Dnmt1*-
232 KOs (purple) compared to control NPCs (orange) highlighted in gray. Flanking regions of
233 middle DMR are ±5kb. In x-axis, each vertical black line represents a CpG in mouse genome.
234 (I) Plot showing mean DMR numbers per gene with error bars representing standard error.
235 Wilxocon-rank test shows statistical significance, p<0.0001. (J) *Gsx1* DMR (FWER < 0.05)
236 found in *Kmt2a*-KO NPCs (orange) compared to control NPCs (purple), highlighted in gray.
237 Flanking regions of DMR are ±5kb. In x-axis, each vertical black line represents a CpG in
238 mouse genome. (K) RNA sequencing results showing of normalized *Gsx1* counts in *Kmt2a*-
239 KOs and controls. *p<0.05, **p<0.01, ns=not significant.

240

241 **Substantial convergence of gene expression abnormalities**

242 Given the phenotypic overlap between multiple MDEMs, we next investigated the
243 overlap in molecular consequences between *Kmt2a*- and *Dnmt1*-KOs. When we compared the
244 DEGs from both *Dnmt1*-KO and *Kmt2a*-KO cells, we found a substantial overlap of 1543
245 DEGs, including 679 upregulated and 861 downregulated genes (Fischer's exact test,
246 OR=11.6, p < 0.0001, **Figure 3A, Supplementary Table 7**). Moreover, gene expression
247 changes were highly correlated between the two datasets (Spearman's ρ = 0.94, p<0.0001,
248 **Figure 3B**), and almost all differentially expressed genes were concordant in the direction of
249 change. In both datasets, the upregulated genes were enriched in neuronal differentiation and
250 maturation GO terms (**Figure 3C, Supplementary Figure 3A-B**), while the downregulated
251 genes were enriched in GO terms linked to cell cycle and chromosome organization (**Figure**
252 **3D, Supplementary Figure C-D**). These patterns mirror the transcriptional profile of
253 differentiated cells, suggesting that premature differentiation may underlie at least part of the
254 observed gene expression changes.

255 To determine whether *Kmt2a*-KO and *Dnmt1*-KO mNPCs undergo premature
256 differentiation, we explored markers for neuronal and glial differentiation in both KOs and
257 scrambled control mNPCs, using WB and IF staining. WB analysis revealed a significant
258 increase in the neuronal marker TUBB3 in *Kmt2a*-KO mNPCs compared to controls (**Figure**
259 **3E**), whereas *Dnmt1*-KO mNPCs did not exhibit a similar change at the protein level (**Figure**
260 **3E**). No notable differences in the proliferation marker PCNA were detected in either KO
261 model (**Supplementary Figure 3G**). Interestingly, IF staining demonstrated a greater number

262 of TUBB3-positive cells in both *Kmt2a*-KO and *Dnmt1*-KO cultures, compared to scrambled
263 controls, along with distinct morphological changes indicative of differentiation, such as
264 smaller and elongated cell-bodies (**Figure 3F-G**). Increased expression of the astrocyte
265 marker EAAT2 was also observed in *Dnmt1*-KO cultures, whereas no significant change was
266 detected in *Kmt2a*-KO cells (**Supplementary Figure 3E-F**). Prior work has shown
267 prematurely poised neuronal enhancers in primary mNPCs from a *Kmt2a*-haploinsufficient
268 model, a phenotype which normalized after genetic rescue of *Kmt2a* (Reynisdottir et al.,
269 2024). Additionally, NPCs from this model exhibit a similar premature neuronal
270 differentiation phenotype as in our full KO (**Supplementary Figure 3H-I**), supporting that
271 this is a feature of *Kmt2a* disruption. These findings support that loss of both factors
272 contributes to premature differentiation of mNPCs but they differ in impact. Specifically, loss
273 of *Kmt2a* leads to a robust and uniform neuronal differentiation phenotype, whereas *Dnmt1*
274 deficiency promotes both neuronal and glial differentiation, but potentially in a more variable
275 manner.

276 For DNA methylation, we identified two intergenic DMRs near *Gsx1* in the *Dnmt1*-
277 KO (**Supplementary Figure 2G-H**). However, unlike *Kmt2a*-KO, *Gsx1* did not exhibit
278 corresponding transcriptional changes in *Dnmt1*-KO. We next asked whether the shared
279 DEGs were driven by common regulatory programs, using motif enrichment analysis of their
280 promoter regions. Analysis of Motif Enrichment (AME; MEME-suite) of promoter regions of
281 the shared DEGs identified common TFs in the promoter regions, including members of the
282 specificity protein (SP) and E2F families, notably SP2 and E2F4, both of which are known
283 regulators of cell cycle progression and differentiation in neuronal cells (**Figure 3H,**
284 **Supplementary Table 8)** (Liang et al., 2013; Persengiev et al., 1999).

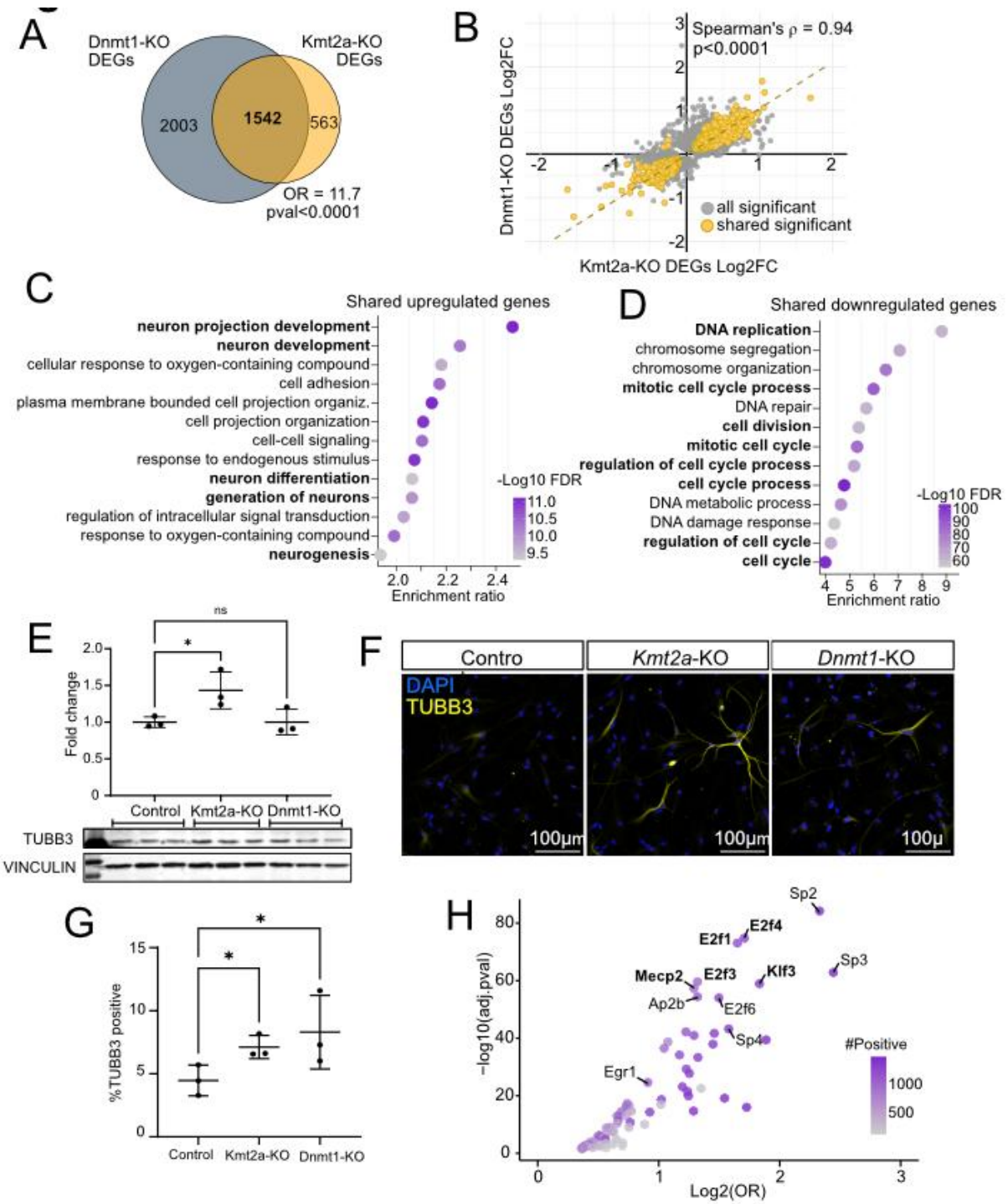


Figure 3: Premature differentiation in *Kmt2a*-KO and *Dnmt1*-KO. (A) Venn diagram showing extensive overlap of DEGs in *Dnmt1*-KO and *Kmt2a*-KO cells compared to control NPCs, with results from Fisher's exact test (OR=11,7, p<0.001). (B) Scatter plot showing Log2FC of significant DEGs from both *Dnmt1*-KOs 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

295 enrichment ratio for each GO-term on the x-axis and color depicting the FDR. **(**E) WB
296 analysis of TUBB3 protein after knocking out *Kmt2a* (padj=0.0474) and *Dnmt1* (padj>0.05,
297 ns) with CRISPR/Cas9. A one-way ANOVA with Dunnett´s multiple comparisons test was
298 used to calculate significance. (F) Immunofluorescence staining of mNPCs after *Kmt2a*-KO
299 and *Dnmt1*-KO compared with controls (blue = DAPI, yellow = TUBB3, magenta = EAAT2).
300 (G) Quantification of the percentage of TUBB3 positive cells in *Kmt2a*-KO, *Dnmt1*-KO and
301 controls. A one-way ANOVA with an FDR correction was used to calculate significance. (H)
302 Scatter plot of TFs ranked by the number of promoters containing their binding motifs. The
303 odds ratio indicates motif overrepresentation in promoters of shared DEGs.*p<0.05, ns=not
304 significant.

305

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

307 Given the premature differentiation observed in EM KOs, some of the expression
308 changes may reflect differences in cell state composition across genotypes. To explore the
309 extent of this in our EM KOs, we differentiated wild-type (WT) mNPCs into mixed neuronal
310 cultures (see methods) and demonstrated that these cultures express typical markers for
311 neurons and astrocytes, as well as synaptic markers at Day 6 (**Supplementary Figure 4A**).
312 We then performed ONT long-read RNA-Seq on mNPCs, Day 2, and Day 6 differentiating
313 cells (schematic in **Figure 4A**), which yielded 3,706 and 3,441 DEGs at Day 2 and Day 6,
314 respectively, compared with mNPCs (**Supplementary Table 9-10, Supplementary Figure**
315 **4B**). Both developmental timepoints were enriched in neurogenic and synaptic terms,
316 supporting neural differentiation and maturity of the culture (**Supplementary Figure 4C-D**).
317 Additionally, a gene-set enrichment analysis of cell-types showed significant enrichment in
318 cell-type markers for the three main brain-cell lineages (neurons, astrocytes and
319 oligodendrocytes, **Supplementary Figure 4E**) as would be expected.

320 To understand the extent of transcriptional changes that may be secondary to
321 premature differentiation in *Dnmt1* and *Kmt2a* KOs, we compared DEGs from each KO to
322 those identified during normal neuronal differentiation (mNPCs vs. Day 2 and Day 6).
323 Interestingly, DEGs from both KOs showed strong and consistent overlap with normal
324 differentiation associated gene expression changes (odds ratios > 4 in all comparisons, $p <$
325 0.0001, Fisher's exact test; **Figure 4B**) and displayed a linear correlation (Spearman's $\rho >$
326 0.45 for all comparisons, $p < 0.0001$; **Figure 4C-F**). These results suggest substantial
327 convergence between KO-induced and differentiation-related transcriptional programs,
328 consistent with a premature differentiation phenotype in EM-KOs.

329 To map the temporal landscape of DNAm during neuronal differentiation, we profiled
330 genome-wide DNAm using ONT in mNPCs at Day 0, Day 2, and Day 6 of differentiation
331 (schematic in **Figure 4A**). A principal component analysis (PCA) of DMRs showed that the
332 developmental stage was the main driver of methylation variance, as captured by PC1
333 (**Supplementary Figure 4F**). While differential methylation analysis identified no significant
334 changes at Day 2 it revealed 8,627 genome-wide significant DMRs (qval < 0.05) at Day 6
335 compared to Day 0 mNPCs, indicating that most DNAm remodeling occurs relatively late in
336 differentiation (**Figure 4G, Supplementary Table 11**). Genic-context analysis indicated that,
337 similar to *Dnmt1*-KO DMRs, on Day 6 DMRs were enriched in introns, 3' UTRs, and exons,
338 but they were depleted in promoters, 5' UTRs, and intergenic regions (**Supplementary**
339 **Figure 4G**). To assess whether the methylation dynamics observed during normal
340 differentiation are recapitulated in the KOs, we examined the methylation state of the Day 6
341 DMR regions in *Dnmt1*-KO and *Kmt2a*-KO cells. *Kmt2a*-KO cells maintained a methylation
342 profile similar to undifferentiated mNPCs at these loci, while *Dnmt1*-KO cells exhibited a
343 methylation pattern resembling that of Day 2 neurons (**Figure 4H**). Methylation in the areas
344 surrounding these Day 6 DMRs were also lower in *Dnmt1*-KO cells. Additionally, 1,154 of
345 the Day 6 DMRs were also identified in *Dnmt1*-KO cells (odds ratio = 1.48, $p < 2.2e-16$,
346 **Supplementary Table 12**). We conclude that the methylation changes over these Day 6
347 DMRs are driven by global hypomethylation in *Dnmt1*-KO cells but is also consistent with
348 aspects of premature differentiation. Together, these results suggest that transcriptional
349 changes precede widespread DNAm remodeling, supporting a model in which DNAm acts as
350 a secondary layer that consolidates, rather than initiates, neurodevelopmental programs.

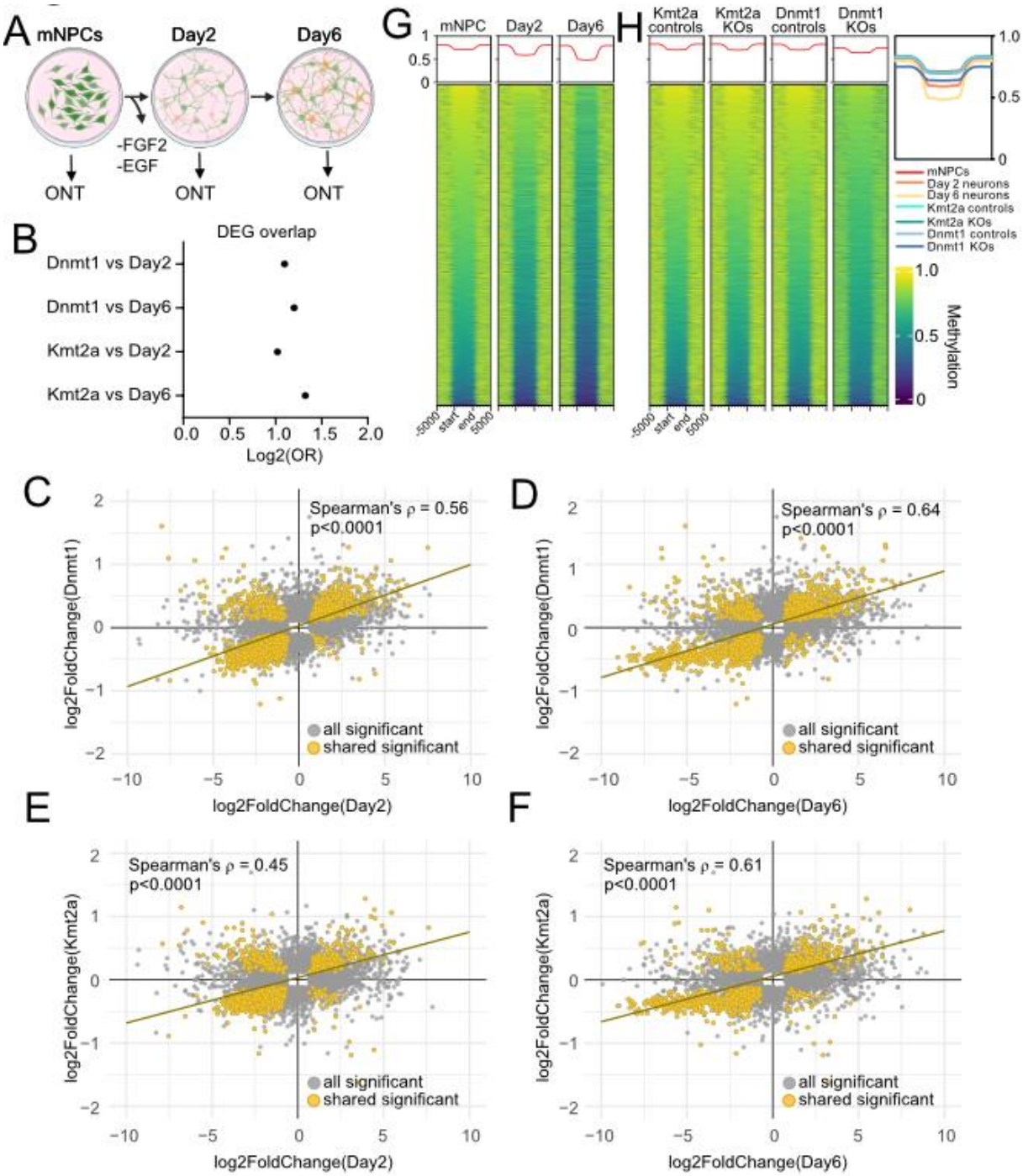


Figure 4: Changes observed in DNAm and gene expression during differentiation of NPCs. (A) A schematic of the neurodevelopmental model. (B) Results of Fischer's exact test shown for the overlap between *Kmt2a*-KO and *Dnmt1*-KO DEGs with NPC-Day 2 and NPC-Day 6 DEGs. (C-F) Significant Spearman correlations 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.

363 **An expanded cohort of EM KOs reveals subtle DNAm changes**

364 We next broadened our analysis to assess DNAm abnormalities across the wider
365 epigenetic machinery. Using ONT long-read whole genome sequencing, we profiled 44
366 additional EM-KO lines alongside matched scrambled controls. These data were pooled with
367 previously generated data from *Dnmt1*-KO and *Kmt2a*-KO, resulting in a total of 46 EM-KOs
368 and 16 control samples. To systematically evaluate genome-wide DNAm changes, we
369 partitioned the mouse genome into ~2.7 million 1 kb tiles (each containing >2 CpGs) and
370 compared methylation levels across groups. A PCA plot of these tiles revealed that two
371 known DNAm machinery components, *Dnmt1* and *Dnmt3b*, showed the most dramatic
372 DNAm abnormalities (**Supplementary Figure 5A**). When these two were excluded from the
373 analysis, *Kat6a*, *Chd5*, and *Tet1* emerged as outliers, clustering separately from the remaining
374 EM-KOs (**Supplementary Figure 5B**). Despite this variation, no shared DMRs were
375 identified across the 44 EM-KOs. Given the diverse roles of EM genes in modulating DNAm,
376 we hypothesized that shared DMRs may exist within specific subgroups of EM-KOs with
377 related molecular functions. Therefore, classifying EM genes based on regional DNAm
378 patterns could help uncover convergence.

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

380 Promoter regions are well-established regulatory elements that are easily assigned to
381 the corresponding genes and involved in transcriptional control (Deaton & Bird, 2011). To
382 investigate whether a subset of EM-KOs exhibit similar DNAm changes at gene promoters,
383 we applied non-negative matrix factorization (NMF) to cluster the 46 EM-KOs based on
384 DNAm patterns at the promoters of 20,482 mouse protein-coding genes based on the
385 cophenetic distance and dispersion of latent factors (**Supplementary Figure 5C)**.
386 Hierarchical clustering of EM-KOs using the NMF-derived coefficients further characterized
387 three latent factors (**Figure 5A**). Specifically, cluster 1 included *Chd3, Setd4, Setd1a, Setd1b,*
388 *Kdm5b*, and *Kdm1a*. In contrast cluster 2 included the controls and showed low coefficients in
389 all three latent factors. Finally, cluster 3 included the genes *Kmt2a*, *Dnmt1*, *Chd1*, *Kdm6a*,
390 *Hdac6*, *Ezh2*, and *Crebbp* and showed the highest coefficient in latent factor 2 (**Figure 5A**).
391 *Dnmt3b*-KO showed the highest coefficient in latent factor 1 and did not cluster with other
392 EM genes (**Figure 5A**). Interestingly, genes with the highest weight in latent factor 2 were
393 significantly enriched in GO terms related to generation of neurons and neurodifferentiation
394 (**Figure 5B**) as well as the Wnt signaling pathway (**Supplementary Figure 5D**). In contrast,
395 no significant enrichment in 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, there was a significant enrichment in interactions with 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 factors (Clusters 1&2 and DNMT3B), we also observed 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 = 0.001367). These results suggest that EM factors outside the core DNAm machinery may influence promoter DNAm through interactions with the maintenance methyltransferase.

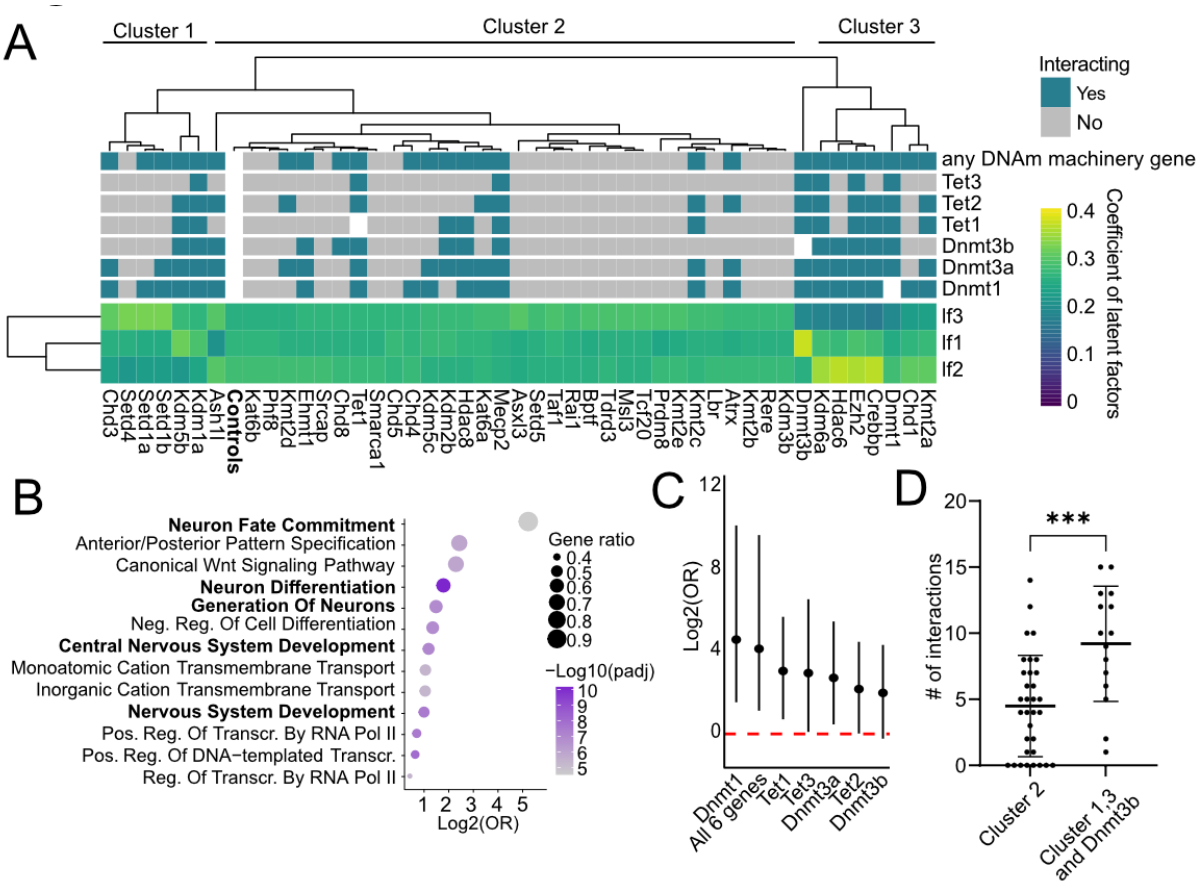


Figure 5: DNAm analysis of 46 EM KO genes in NPCs reveals shared promoter effects and haplotype specific effects. (A) A composite graph showing number of interactions with DNAm enzymatic genes on top, the heatmap showing coefficient of latent factors for each EM across three latent factors. Color scale 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 DNAm machinery-interacting EMs in two relevant clusters compared to rest. Significance testing was done with a Fisher's exact test. Dots represent Fisher's exact test odds ratios ($\log_2$ scale), with lines indicating 95% confidence intervals. Horizontal dotted red line represents statistical significance. (D) Comparison of number of STRING interactions with 25 DNAm machinery components between the clusters. A Wilcoxon rank sum test was used for significance. lf1 = latent factor 1, lf2 = latent factor 2, lf3 = latent factor 3. ***<0.001, ns=not significant.

Haplotype-specific DNAm within the context of EM KOs

We used mNPCs derived from F1 offspring of the two strains (FVB maternal × B6J paternal, **Figure 1A**). Because the C57BL/6J (B6J) and FVB mouse strains differ genetically across roughly half the genome (~50–60%), it is possible to phase this part of the genome and perform an allele-specific methylation analysis. PCA of the phased data shows the same pattern between KOs and controls as for the unphased data (**Supplementary Figure 6A**) but supports the presence of allelic-specific methylation between the two strains.

Using the control samples only, we found 1,074 haplotype-specific DMRs (FWER < 0.05) (**Figure 6A, Supplementary Table 13**) (Do et al., 2016; Tycko, 2010), including one located in the 3′ UTR of *Tet3* (**Figure 6B**). Because of the consistent parental origin of the two genomes, these DMRs should include imprinted loci. Indeed, our DMRs include 21 out of 57 known imprinted loci (**Supplementary Table 14**) (Gigante et al., 2019; Xie et al., 2012); note that imprinting is partly cell type specific. We next examined the genetic variation at these loci. As expected, we found that 95% of the haplotype-specific DMRs (n = 1,018) contained haplotype-specific genetic variants, including insertions/deletions (indels), SNPs and CpG changes (**Figure 6C**). This variation appeared enriched in exons, 5' and 3' UTRs, and promoters but depleted in intergenic regions (**Figure 6D**). Within these DMRs we found 11,285 single nucleotide polymorphisms, 7836 indels, 70 insertions and 64 breakpoints within the 1074 DMRs (i.e. 17,92/DMR). In contrast, only about half of the ICRs defined DMRs (52,35%, n=11) harbored genetic variants and these regions carried fewer variants on average (7.57 per DMR). This pattern is consistent with imprinted regions being primarily regulated epigenetically, whereas other allele-specific DMRs are more often mediated by underlying genetic variation. This pattern is consistent with imprinted regions being primarily regulated epigenetically, whereas other allele-specific DMRs are more often mediated by underlying genetic variation.

Finally, we searched for DMRs associated with our 46 EM KOs, separately on each haplotype. We found a shared hypomethylated DMR on the C57BL/6J haplotype

451 (FWER<0.05) and none on FVB. The DMR is located in the 3' UTR of *Zic4* (**Figure 6E,**
452 **Supplementary Figure 6B**), a TF that plays a role in cerebellar development and
453 malformations (Grinberg et al., 2004). Although *Zic4* was not differentially expressed in
454 *Dnmt1*-KO nor *Kmt2a*-KO mNPCs compared to scrambled controls (**Supplementary Figure**
455 **6C-D**), it was significantly downregulated (log2FC = -1.978, adjusted p-val = 0.00197) at
456 Day 6 compared to Day 0 mNPCs in our neurodevelopmental model (**Supplementary Figure**
457 **6E**). Similarly, prior studies (Choufani et al., 2015) have reported alterations of the 3' UTR of
458 *ZIC4* (CpG cg26386385). Overexpression of *Zic4* in NPCs reduced expression of neuronal
459 maturation markers (**Supplementary Figure 6F-G**), implicating ZIC4 in maintaining the
460 NPC state and raising the possibility that it may act as a brake on neuronal maturation.

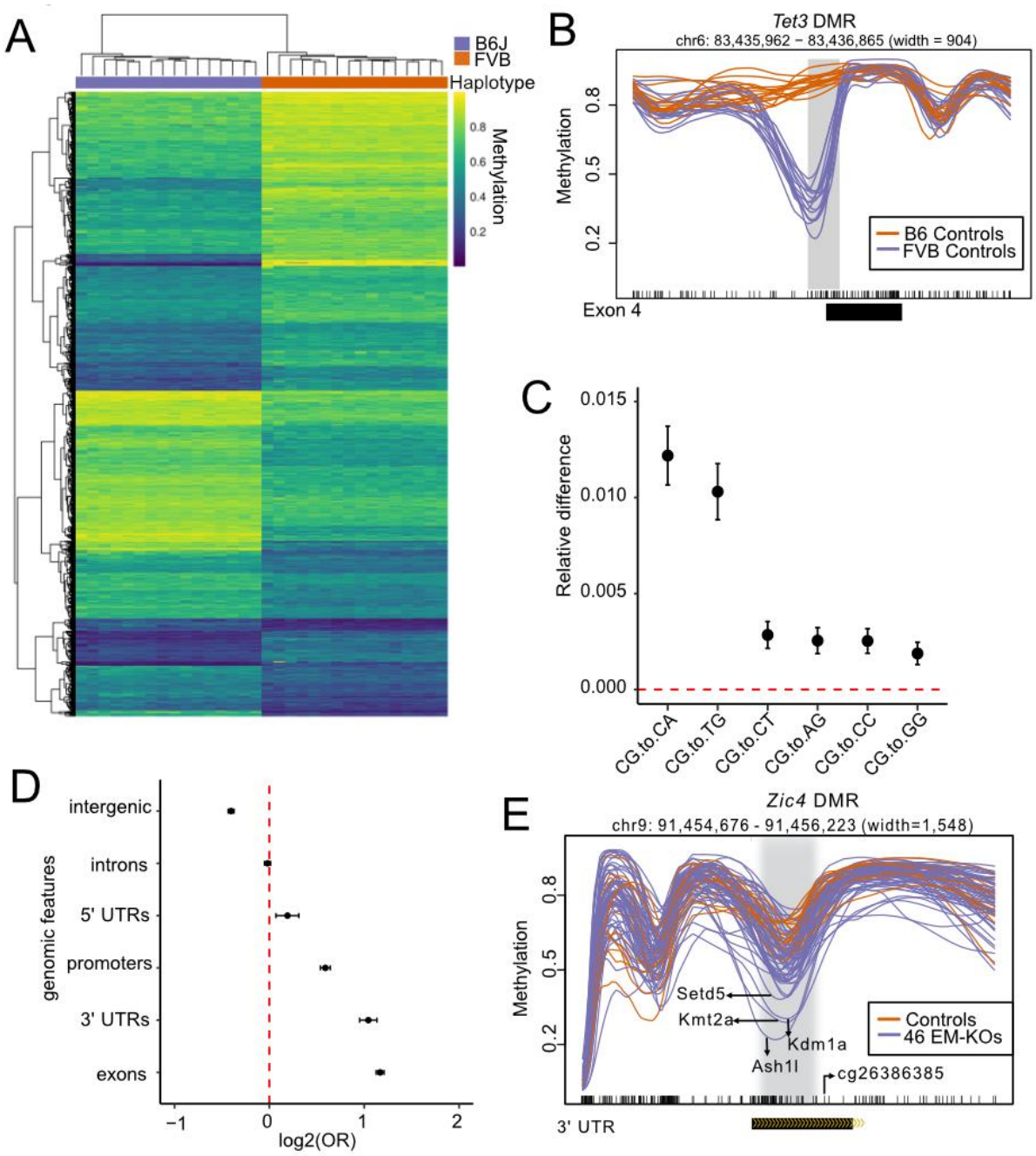


Figure 6: Haplotype specific changes in DNA methylation. (A) A heatmap plot showing smoothed methylation of 1074 haplotype-specific DMRs across B6J and FVB haplotypes. Yellow is higher methylation but blue is lower methylation. Each line in y-axis represents a single DMR across haplotyped controls but x-axis represents haplotyped controls. B) A haplotype specific DMR in the *Tet3* gene. C) An overrepresentation of CpG-related changes within the haplotype specific DMRs supporting an ongoing role for DNAm in this region. D) The localization of genetic variation within the haplotype specific DMRs.

(E) A line plot showing a DMR in 3' UTR of *Zic4* between B6 haplotyped control mNPC (orange) and B6 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.

Discussion

Our study presents a comprehensive investigation into how EM disruption affects DNAm and gene expression in a disease-relevant neuronal system, offering insights into the presence and functional impact of DNAm changes beyond peripheral blood. By combining a mNPC model with ONT sequencing, haplotype phasing, and transcriptomic profiling, we characterized the functional consequences of 46 EM gene KOs. Disruption of DNMT1 leads as expected to extensive changes in the DNAm. Our PCA suggests that the next largest effect on DNAm is from DNMT3B and, interestingly, KAT6A, although more work will be needed to map these changes genome wide. Across the 46 EM genes we profile, we observe little to no changes. This could be because of the relatively short time lapse between knockout and methylation profiling in our work. Our data suggests that TFs such as GSX1, EGR1, EBF1, and E2F4 may mediate or interpret the EM-related epigenetic disruption. For instance, *Gsx1* was differentially expressed in *Kmt2a*-deficient cells, and its target genes were enriched among DEGs in both *Kmt2a*- and *Dnmt1*-KOs. Similarly, EGR1 and SP1 binding motifs were enriched in DMRs or associated with downstream gene expression changes. These findings point to a model in which the EM indirectly modulates transcriptional networks through TFs, either by altering DNAm at TF binding sites or affecting expression of the TFs themselves. Importantly, some TFs such as EGR1 are known to recruit TET1, further linking transcriptional networks to active demethylation pathways. Similarly, our clustering of promoter DNAm reveals that at least in some EM contexts, the physical interaction with the DNAm machinery plays a role.

Interestingly, we uncovered a hypomethylated DMR in the 3' UTR of *Zic4* which was shared across all 46 EM-KOs. This gene is implicated both murine and human neurodevelopment and DNA methylation changes in the gene has been identified in peripheral blood from individuals with Sotos syndrome, a neurodevelopmental disorder caused by mutations in the epigenetic regulator *NSD1* (Choufani et al., 2015). However, the methylation changes are modest, and the gene is not differentially expressed in our system. It

501 is not clear whether this gene will be important in explaining the phenotypic overlap in the
502 MDEM group, but it could have potential utility as a biomarker.

503 We have established a premature differentiation phenotype arising as a result of KO of
504 either DNMT1 or KMT2A. The recent study of KMT2A-deficient mice revealed a H3K4me1
505 enrichment at neuronal enhancers in undifferentiated NPCs (Reynisdottir et al., 2024),
506 suggesting that these enhancers show premature poising. The premature differentiation we
507 observed is also in line with prior work on Kabuki syndrome, another MDEM, where the
508 same phenotype has been reported in two disease relevant contexts (Carosso et al., 2019;
509 Fahrner et al., 2019). Moreover, studies of Kabuki syndrome neurons from a mouse model
510 have shown an overrepresentation of aging-related loci (Boukas et al., 2023), further
511 suggesting that premature differentiation may represent a unifying phenotype across a subset
512 of MDEMs.

513 Despite the modest DNAm changes, we observed an impressive convergence of
514 transcriptional dysregulation between DNMT1 and KMT2A, along with the aforementioned
515 shared cellular phenotype of premature differentiation. Comparing the transcriptional changes
516 to changes during normal neuronal differentiation, we conclude that part of the overlap is a
517 consequence of the premature differentiation, although we note that the overlap between
518 DNMT1 and KMT2A appears stronger than either of the overlaps with normal differentiation.
519 Combined with the dearth of methylation changes in KMT2A, we have shown that the
520 expression changes are not a consequence of DNA methylation changes. It will be interesting
521 and important to understand how the shared expression changes arise, and especially to
522 identify which epigenetic changes if any are causal. In any case, differentiation state should
523 be carefully considered when analyzing gene expression, since premature or altered
524 differentiation can be a major source of the observed changes.

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

528

529 **Materials and methods:**

530 **Animals**

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

542 **Dissections and primary NPCs isolation**

543 The NPCs were isolated from an F1 incross of homozygous Cas9 mice (B6J and FVB). The
544 isolation procedure was based on published protocols (Bernas et al., 2017; Rybachuk et al.,
545 2019) with minor modifications, see Supplementary Materials for details.

546 **Neuronal progenitor differentiation**

547 NPCs (passage 3-4) were differentiated into mixed neuronal cultures by collecting them as
548 previously described, counting and seeding 150.000 cells/well on a PDL/Laminin coated 24-
549 well plate in 0.5ml N2-Neuronal Induction Medium (N2NIM) consisting of Neurobasal Plus
550 Medium (ThermoFisher, A3582901) with 1x B-27 Plus Supplement (ThermoFisher,
551 A3582801), 1x CultureOne Supplement (ThermoFisher, A3320201), 1x N2 Supplement
552 (ThermoFisher, 17502048), 0.25x Penicillin-streptomycin and 0.25x GlutaMax. This was
553 supplemented with 20ng/ml FGF2 and 10ng/ml BDNF (Peprotech, 450-02). At day 2 and 4,
554 medium was changed to N2NIM supplemented with 10ng/ml FGF2, 10ng/ml BDNF and
555 10ng/ml GDNF (ORF Genetics, IK0900). At Day2 and Day6 of differentiation, neuronal
556 cultures were collected and DNA and RNA isolated as described below.

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

558 CRISPR/Cas9 guide RNAs (gRNAs) were designed with an online tool
559 (https://chopchop.cbu.uib.no/). Two gRNAs, with the highest predicted efficiency and
560 minimal mismatches per gene, were selected to ensure out-of-frame upon Cas9-mediated
561 cleavage. All sgRNAs were verified with BLAT to ensure that none target any other

562 sequences in either of the two genomes. Eight scrambled control sgRNA pairs (random
563 sequence not binding in the murine genome) were designed for controls. Lentiviral CRISPR
564 vectors (from (Replogle et al., 2020)), each including one pair of gRNAs (**Supplementary**
565 **Table 1**) were synthetized by Vectorbuilder. Bacterial stabs were streaked onto LB plates and
566 plasmids isolated using ZymoPURE II plasmid Midiprep kit (Zymo, D4201) and stored at -
567 20°C until needed. All 54 dual guide-RNA (sgRNA) lentiviral plasmids were Sanger
568 sequenced to confirm correct gRNA pairs after amplification.

569 **Cell culture**

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

574 **Virus production and transduction on NPCs**

575 Lentivirus was generated by transfecting HEK293T cells with the CRISPR plasmid, psPAX2
576 (Addgene, 12260), pMD2.G (Addgene, 12259), at a 2:1:1 weight ratio, using Optimem
577 (ThermoFisher, 31985047) and Lipofectamine 3000 (ThermoFisher, L3000015). The
578 transfected HEK293T cells were cultured for 72 hours, and viral particles collected from the
579 medium after 48 and 72hours, passing the supernatant through a 0.45 µm syringe filter (Santa
580 Cruz, sc-358814) to remove cell debris. After the second collection, 4x PEG8000 virus
581 concentrator (Kutner et al., 2009; Lo & Yee, 2007) was used to precipitate viral particles.
582 Fifty µl viral aliquots were stored in -80°C until use. Transduction efficiency for each virus
583 was tested in NPCs at several concentrations, using serial dilutions of the virus together with
584 3.2ug/ml polybrene (Santa Cruz, sc-134220). Transduction efficiency was determined using
585 fluorescence-activated cell sorting (FACS) of BFP fluorescent cells in NPC cells, 48 hours
586 after transduction, and an efficiency of ~30% was used to conduct all experiments.

587 **EM KO in NPCs and validation**

588 For EM KO experiments, the NPCs were seeded 24 h prior to transduction. Each lentivirus,
589 targeting an individual EM gene, was transduced using 3.2 µg/ml polybrene. Non-targeting
590 controls were used as negative controls for viral infection and culturing effects. After 24 h,
591 cells were split and selected by 0.5 µg/ml puromycin for 2 days. Cells surviving puromycin
592 treatment were cultured for another day without puromycin to allow them to recover, then

593 harvested for experiments using Accutase as described above. Timing was selected based on
594 the fact that average protein half-life in mouse primary neurons is ~86 hours (Mathieson et al.,
595 2018) and for EM proteins it is considered even less (~40 hours), thus most of the EM
596 proteins would be expected to be degraded at the time of our evaluation. The KO efficiency
597 was determined by PCR, WB and immunofluorescence staining.

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

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

607 **Protein isolation and WB**

608 NPCs were collected using Accutase and counted with a Countess II FL Automated Cell
609 Counter (ThermoFisher). Cells were lysed at 15000-20000 cells/µl in 1x sample buffer
610 (60mM Tris HCl pH 6.8, 2% SDS, 10% Glycerol, 0.01% Bromophenol blue, 1.25% Beta-me)
611 supplemented with 1:100 Halt™ Protease Inhibitor Cocktail (ThermoFisher, 78437), 1:100
612 phosphatase inhibitor Cocktail (Cell signaling, 5870) and 1:100 PMSF Protease Inhibitor
613 (ThermoFisher, 36978). The lysate was heated at 95°C for 5 min and 1:1000 Benzonase
614 nuclease (Sigma-Aldrich, sc-202391) added to degrade genomic DNA. The lysate was
615 aliquoted and stored in -80°C. Western blotting was performed as has previously been
616 described (Mahmood & Yang, 2012), where 15-25 µg of protein was loaded per sample,
617 blotted onto PVDF membrane (ThermoFisher, 88520) and stained overnight using the
618 following primary antibodies: rabbit anti-CHD1, 1:1000 (Cell signaling, 4351), mouse anti-
619 ACTIN, 1:1000 (Abcam, ab8224), rabbit anti-EAAT2, 1:1200 (Abcam, ab41621), mouse
620 anti- beta III Tubulin, 1:800 (Abcam, ab78078), rabbit anti-DNMT1, 1:1000 (Proteintech,
621 24206-1-AP), mouse anti-VINCULIN, 1:1000 (Sigma, V9131), rabbit anti-Myc-Tag, 1:1000
622 (Cell signaling, 2272). Mouse anti-PCNA, 1:7000 (Proteintech, 60097-1-Ig) was incubated at
623 RT for 1.5 h. Blots were stained with secondary antibodies: Donkey anti-Mouse IgG, 1:15000
624 (LICOR, 926-68072), donkey anti-abbit IgG, 1:15000 (LICOR, 926-32213) for 1 h at RT.

625 After washing, blots were visualized with Odyssey® CLx Imager and protein quantification
626 was evaluated by Image J (PMC3855844).

627 **Immunofluorescence**

628 NPCs were cultured on PDL/Laminin-coated 8-well culture slide (Falcon, 354118) for 24-48
629 hours, fixed with 4% paraformaldehyde (PFA), blocked with 5% NGS in 0.1% PBS-T and
630 stained using primary antibodies against rabbit anti-Kmt2a, 1:200 (Cell signaling, 14197),
631 rabbit anti-EAAT2 1:400 (Abcam, ab41621) and mouse beta III Tubulin 1:200 (Abcam,
632 ab78078), overnight. The following day, cells were stained using secondary antibodies: Alexa
633 goat anti-rabbit 647, 1:1000 (ThermoFisher, A-21244) and Alexa goat anti-mouse 555,
634 1:1000 (ThermoFisher, A-32727). Cells were mounted with Fluoromount-G™ Mounting
635 Medium, with DAPI (ThermoFisher, 00-4959-52). Images were captured with Olympus
636 Confocal microscopy (Olympus Fluoroview, FV4000), processed with Fiji and quantified
637 with CellProfiler. Statistical analysis was done using GraphPad Prism v10.4.

638 **Overexpression**

639 NPCs were seeded at a density of approximately 250,000 cells per well in PDL and Laminin
640 coated 6-well plates, 24 hours prior to transfection. The following day, cells were co-
641 transfected with 2.5 µg of either the *Zic4* overexpression plasmid or the corresponding
642 plasmid backbone (control), along with 7.5 µl of Lipofectamine 3000, following the
643 manufacturer's protocol. Cells were harvested 24 hours post-transfection for downstream
644 analysis.

645 **DNA sample isolation and preparation**

646 Genomic DNA isolated from cells using Monarch® HMW DNA Extraction Kit for Cells &
647 Blood (NEB, T3050) followed the manufacturers protocol for DNA extraction from cells. The
648 concentration and quality of DNA was measured by Nanodrop (ThermoFisher) and Qubit to
649 quantify the precise amount of double strand DNA via Qubit dsDNA HS assay kit
650 (ThermoFisher, Q33231). High molecular weight (HMW) DNA was sheared to 20-30 kb to
651 achieve consistent size and to optimize ONT sequencing yield. HMW DNA was loaded to a
652 hydro-tube assembled with hydropore-syringe (Diagenode, E07010003), inserted into the
653 cassette, installed on a base unit and then sheared by Megaruptor® 3 (Diagenode) using the
654 following settings: a volume of 70 µl, a speed of 30, and a concentration of 50 ng/µl. The size
655 of sheared DNA was evaluated with HS Large Fragment 50 kb kit (Agilent, DNF-464-0500)

656 on a Fragment Analyzer, proceeded by High Sensitivity Large Fragment 50 kb Analysis
657 software.

658 **ONT-DNA sequencing library construction**

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

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

671 We developed a computational tool, *samCRISPR*, to estimate CRISPR/Cas9 KO efficiency
672 using ONT long-read whole-genome sequencing (WGS) data. The name "samCRISPR"
673 references both SAMtools, from which the mpileup format is used in KO efficiency
674 calculations, and CRISPR-Cas9 genome editing technology. SAMtools is used to sort, index,
675 and filter aligned reads, and to generate mpileup output at the target site, which provides base-
676 level resolution of read alignments. This output is parsed by our tool, *samCRISPR*, to identify
677 substitutions, insertions, and deletions based on mpileup letter codes, allowing precise
678 quantification of editing events and KO efficiency. The tool identifies sequencing reads that
679 carry CRISPR-induced events within designated target regions using the mpileup output from
680 SAMtools (Danecek et al., 2021) (**Figure 1F**). To eliminate the chance of counting
681 sequencing errors in the vicinity of target regions as actual CRISPR events, the counting of
682 CRISPR events was restricted to 1bp upstream and 1bp downstream of Cas9 cut sites.
683 CRISPR events are defined as mutations occurring within ±1 base pairs of the predicted cut
684 site, located 3 bp upstream of the PAM sequence. This narrow window was chosen to
685 minimize false positives caused by sequencing or alignment errors near Cas9 cleavage sites.
686 KO efficiency is calculated as the ratio of reads with CRISPR events in the target region to
687 the total number of reads mapped to that region. To validate samCRISPR results, KO

efficiencies for selected EM KOs were manually inspected using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011). The script accepts a BAM file, a reference FASTA, and BED-formatted sgRNA coordinates as input, and supports batch analysis of multiple BAM files and sgRNAs. It requires a Linux environment with SAMtools for mpileup generation and R for computing Wilson score confidence intervals (Agresti & Coull, 1998). The custom bash script is available at: https://github.com/kaanokay/samCRISPR. It will be made publicly accessible upon publication of the manuscript.

Phasing of ONT data

We implemented a custom graph genome-based approach integrating both FVB and B6 genomic sequences into a unified non-linear reference. A genome graph, a structure composed of nodes and edges, captures alternative alleles, structural variants, and indels across multiple genomes (Liao et al., 2023; Zhang et al., 2025). Unlike linear references, genome graphs support the simultaneous representation of multiple haplotypes, making them particularly suited for admixed or hybrid genomes. Initially, reads were aligned to the linear B6 and FVB references to generate haplotype-specific BAM files. The genome graph was then used to resolve homologous CpG coordinates between strains, for instance, reconciling differences where a CpG site appears at position 6 in B6 but at position 20 in FVB. Whole-genome alignment of hippocampal mNPC WGS data from F1 B6 × FVB crosses was performed using MUMmer4 v4.0 (Marcais et al., 2018) to identify heterozygous variants. Next, Jellyfish (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 both strains' diversity. Dorado v0.2.4 (https://software-docs.nanoporetech.com/dorado/latest/) was used to base-call canonical and modified bases (5mC, 5hmC) from ONT reads. 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. Methylation calling was carried out using modkit v0.1.2 (https://github.com/nanoporetech/modkit), focusing exclusively on CpG methylation and omitting strand-specific and non-CpG signals. Importantly, all DNAm data were projected

721 onto the pseudo-hybrid coordinate system, minimizing reference bias during allele-specific
722 methylation analysis.

723 **Identification of differentially methylated regions**

724 Methylation data for each sample were extracted from bedMethyl files generated by modkit,
725 which contain per-CpG entries detailing genomic coordinates and read counts. Each file
726 included approximately 21 million CpG sites. These datasets were imported into R to create
727 BSseq objects for downstream differentially methylated region (DMR) analysis. To ensure
728 data reliability, unphased datasets retained only CpGs with ≥3× coverage in at least 80% of
729 samples. For phased data, where coverage was inherently lower, CpGs with ≥1× coverage
730 across all samples were included. DMRs were identified in phased data using the bsseq 1.39.4
731 (Hansen et al., 2012) and unphased data using dmrseq 1.22.1 (Korthauer et al., 2019) R
732 packages. Dmrseq controls the false discovery rate (FDR) which bsseq controls the family-
733 wise error rate (FWER). For the phased data smoothing was performed separately for each
734 allele. Strain-specific CpGs were included during smoothing but were removed prior to DMR
735 identification, following previous work (Wulfridge et al., 2019). This strategy enhanced
736 statistical power while reducing spurious DMR detection due to strain-specific sequence
737 differences. DMRs were filtered using the following criteria phased analysis: ≥3 CpGs per
738 region, >5% mean methylation difference, and inverse CpG density ≤300 (defined as the
739 number of CpGs divided by the region's genomic width), unphased analysis >10% mean
740 methylation difference, ≥5 CpGs, and inverse CpG density ≤120. **Annotation of**
741 **differentially methylated regions**

742 We used annotatePeaks.pl program to annotate DMRs by using UCSC mm10 mouse
743 reference fasta and UCSC mm10 refGene gtf files. That program calls a separate program,
744 assignGenomeAnnotation, to efficiently assign DMRs to possible annotations genome wide.
745 By default HOMER, in case some annotation overlap, HOMER assigns a priority based on
746 the following: TSS (by default defined from -1kb to +100bp), TTS (by default defined from -
747 100 bp to +1kb), CDS Exons, 5' UTR Exons, 3' UTR Exons, introns, and intergenic.
748 Intergenic regions were considered either upstream (TSS) or downstream (TTS) of genes.
749
750 To annotate DMRs in phased data, a custom pipeline was employed: The UCSC mm10 GTF
751 file was converted to BED format using a Bash script. A Python script mapped coordinates
752 between the mm10 genome, and the pseudo-hybrid intermediate genome used for methylation

753 analysis. This allowed accurate functional annotation of DMRs in a unified coordinate space
754 representing both B6 and FVB haplotypes.

755 **Enrichment analysis of DMRs in genomic features**

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

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

774 **DNA methylation overlap across individual datasets**

775 An R-based pipeline was developed to visualize DMRs between mNPCs and Day 6 neurons,
776 incorporating additional conditions including *Dnmt1*- and *Kmt2a*-KOs and Day 2 neurons.
777 Unphased DMR coordinates (from mNPC vs. Day 6 comparisons) were used to extract
778 methylation data from the two experiments. CpGs with zero coverage in any sample were
779 excluded. Remaining CpGs were smoothed with the *BSmooth()* function from the bsseq
780 v1.39.4 package. The *normalizeToMatrix()* function from the EnrichedHeatmap 1.32 package
781 was then used to map smoothed average methylation values from each group to a fixed-width
782 matrix centered on the extended DMR intervals (±5kb) with a 100 bp bin size, and heatmaps
783 were generated using the same package.

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

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

790

791 **RNA sequencing**

792 **RNA isolation and sequencing**

793 RNA was isolated as before using TRI Reagent and purified using ZYMO Direct-zol™ RNA
794 Microprep. For *Kmt2a*- and *Dnmt1*-KOs, the two genes were knocked out individually with a
795 scrambled control in 4 biological replicates each, RNA isolated and sent to Novogene Europe
796 for mRNA Illumina sequencing. For wild-type neurodevelopmental cells, 3 biological
797 replicates were isolated for each timepoint, RNA extracted and purified as before and 200ng
798 total RNA from each sample processed using the ONT PCR-cDNA Barcoding kit (ONT,
799 SQK-PCB111.24), according to the manufacturer's instructions (version v111-revG). Samples
800 were barcoded and pooled at equimolar ratios, then loaded onto 3x R9.4.1 flow-cells (FLO-
801 PRO002) and sequenced on a PromethION 24 instrument.

802 **RNA sequencing analysis**

803 Short-read RNA-seq used paired-end reads (150bp) produced by Illumina were sequenced at
804 ~78 million reads/sample. Reads were aligned to the UCSC mm10 reference genome using
805 HISAT2 (Kim et al., 2019). Alignments with >50 mapping quality were used in featureCounts
806 (Liao et al., 2014) to create a gene-level count matrix. This count matrix was normalized in R
807 using the cpm() function from the edgeR (Robinson et al., 2010), accounting for gene lengths
808 and library size. Protein coding genes with >1 Counts Per Million (CPM) in at least one
809 sample were retained for differential expression (DE) analysis with DESeq2 (Love et al.,
810 2014), considering genes with an adjusted p-value <0.05 as differentially expressed. For long-
811 read ONT RNA-seq, libraries were sequenced with an average of ~10 million reads and a
812 median read length of ~1.1kb. Reads were aligned to the UCSC mm10 reference genome with
813 Minimap2 (Li, 2018) using -uf and -ax splice arguments. featureCounts was utilized to
814 generate the gene-level count matrix with its -L argument for long reads. DE analysis
815 followed the same procedure as for short-read data. To ensure reliability, a histogram plot of
816 unadjusted p-values for all genes with >1 CPM was created in R. To stabilize variance,

817 regularized log transformation (rlog) was applied to raw expression values of retained genes
818 using the DESeq2 package. PCA was performed on the transposed rlog-transformed
819 expression matrix using *prcomp()*, and the first two principal components (PC1 and PC2)
820 were visualized using ggplot2.

821 **Obtaining promoters of mouse protein-coding genes**

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

827 **Clustering of epigenetic machinery KOs based on DNAm promoter abnormalities**

828 We used NMF on smoothed DNAm levels over all promoters of protein-coding genes
829 (Cemgil, 2009). CpGs with more than 3x coverage across at least more than 80% of samples
830 were kept in our analysis. The promoters with no retained CpGs were discarded. The number
831 of latent factors was chosen to be 3 (Boukas et al., 2019). The brunet model in the R package
832 nmf v0.28 was employed in decomposition.

833 **Inspection of STRING interactions of EM genes**

834 These interactions consisted of three categories: known interactions derived from curated
835 databases and experimental data; predicted interactions based on gene neighborhood, gene
836 fusions, and co-occurrence; and other interactions inferred from text mining, co-expression,
837 and protein homology. To determine whether the shared impact on promoter DNAm is
838 associated with PPI-nets among EM) factors, and to establish a robust ranking threshold
839 (parameter k in NMF analysis). We examined interactions reported in the STRING database
840 (Bernas et al., 2017) using a confidence score cutoff of >0.400.

841 **Enrichment analysis and visualization**

842 To test whether overlapping genes between different RNA sequencing datasets was higher
843 than expected, Fisher's exact tests were performed using the function fisher.test() in R, using
844 shared detected genes as background. To test if transcription factor targets or imprinted genes
845 were enriched in DEGs lists, Fisher's exact tests were performed using the function
846 fisher.test() in R, using detected genes as background. TF target genes were obtained from
847 **https://tfbsdb.systemsbiology.net/** (Plaisier et al., 2016) and the TFlink (https://tflink.net/

848 (Liska et al., 2022) databases. A list of mouse imprinted genes was obtained from the
849 Imprinted Gene Database (https://www.geneimprint.com/site/genes-by-
850 species.Mus+musculus), filtering away genes with "unknown" or "not imprinted" status.

851 **Gene ontology analysis**

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

858 **Transcription factor motif enrichment analysis**

859 For identification of TF motif enrichment in DMRs and promoter regions, the Analysis for
860 Motif Enrichment (AME) tool within MEME-suite v5.5.8 was used. The top 30k most
861 differentially methylated regions were tested against a background that included random
862 regions from the mm10 that were matched in size distribution and CpG density and did not
863 include any of the *Dnmt1*-KO DMRs. For promoter regions of shared DEGs we obtained
864 mouse protein-coding promoters for canonical transcripts as described above, filtered them
865 for the shared DEGs and used all other protein-coding genes as a background. In both cases,
866 we tested against the mouse HOCOMOCO v11 core database and estimated enrichment using
867 a Fisher's exact test.

868 **Data visualization**

869 For scatter plots and volcano plots, the R package ggplot2 was used (Wickham, 2016). The
870 heatmaps were generated through the *pheatmap()* function of the R package pheatmap. For
871 individual gene expression values from RNA sequencing data, and for quantification results
872 from qPCR, WB and IF experiments, Graphpad Prism (v10.4) was used. Some figures were
873 generated using BioRender.com.

874

875 **Acknowledgements:** We would like to thank Elvar Örn Jónsson and Þorgils Árni
876 Hjálmarsson for their dedication and valuable efforts in troubleshooting data storage and
877 analysis issues within the high-performance computing cluster at the University of Iceland.
878 We would like to thank Olafur Th. Magnusson for advice regarding ONT 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.

892 **Tables**

893 **List of supplementary tables**

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

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

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

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

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

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

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

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

907 **Supplementary Table 9.** Differential expression analysis results after 2 days of
908 differentiation.

909 **Supplementary Table 10.** Differential expression analysis results after 6 days of
910 differentiation.

911 **Supplementary Table 11.** Genome-wide DMRs between time points of neurodevelopmental
912 model.

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

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

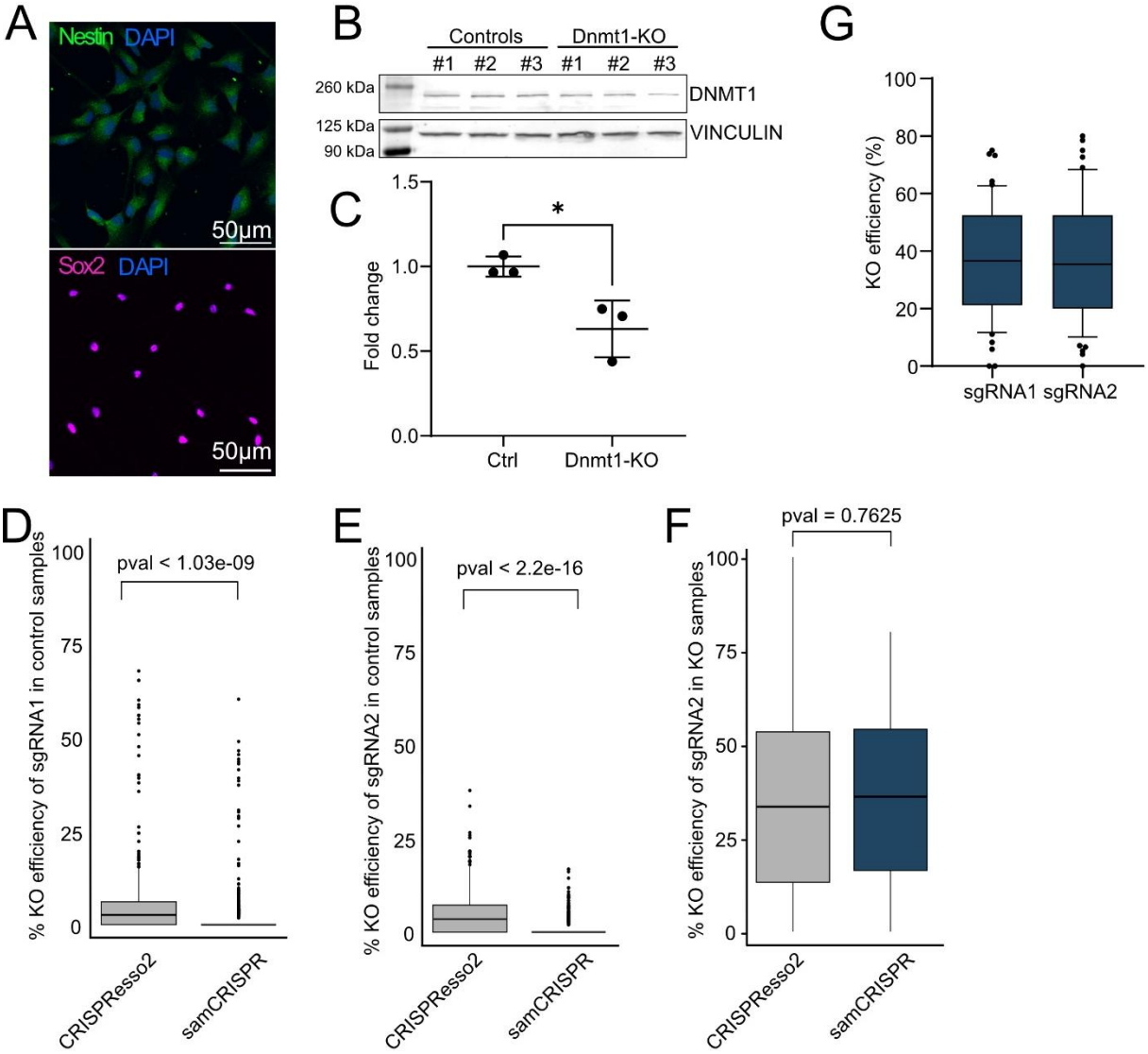
915 **Supplementary Table 14.** Putative imprinting control regions (ICRs).

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

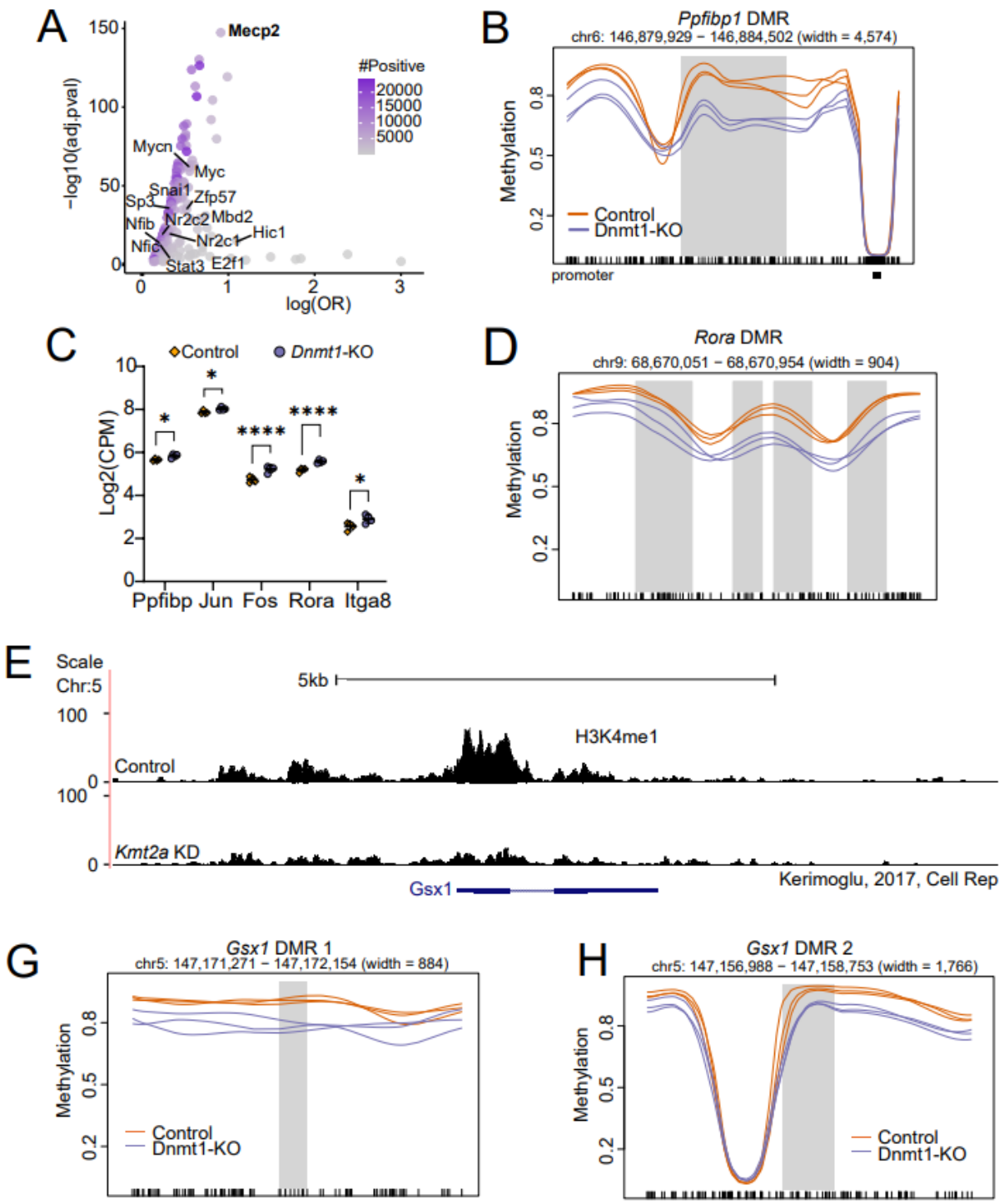
917

918

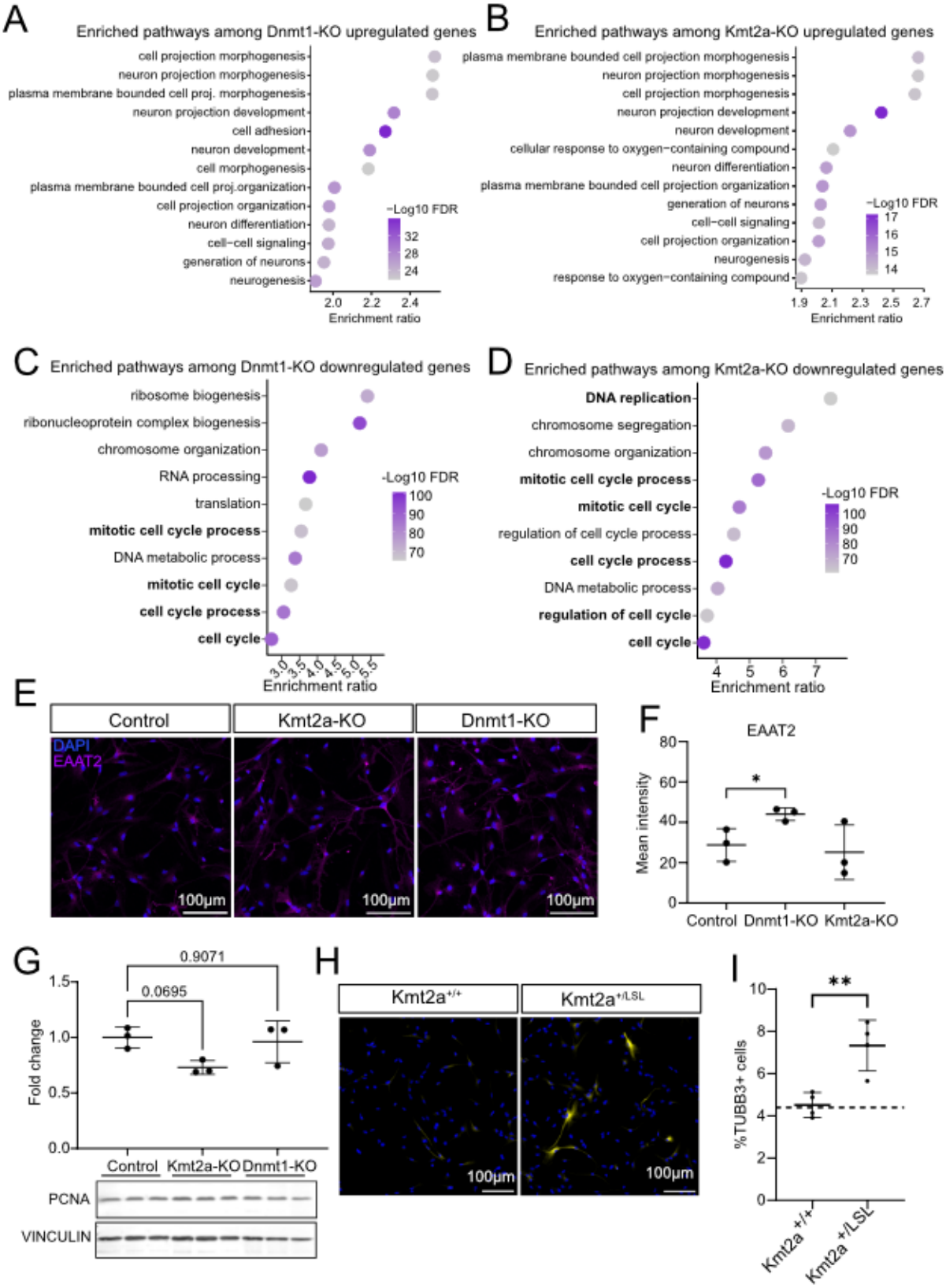
Supplemental figures



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*.

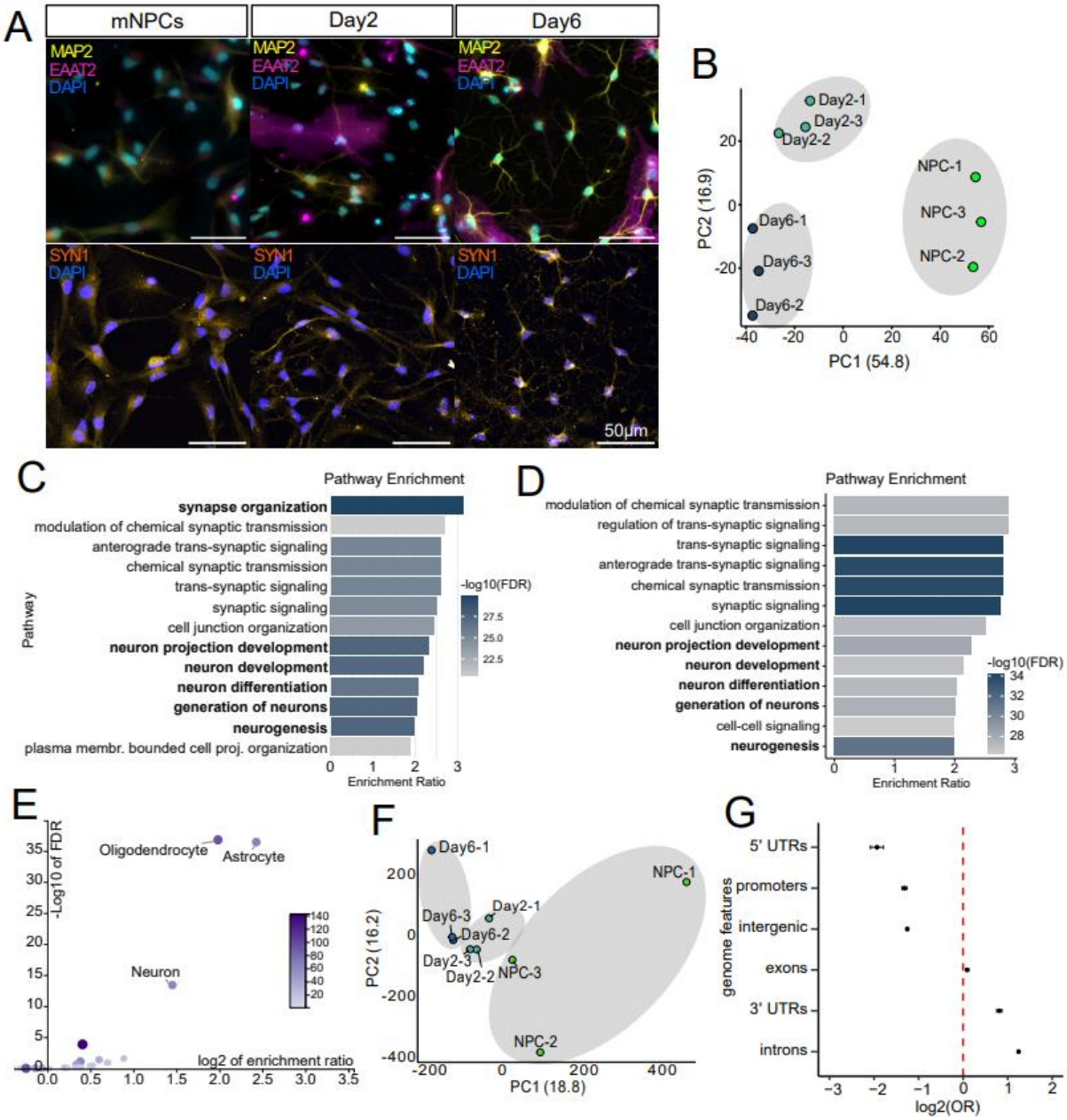


Supplementary Figure 2: *Dnmt1*-KO and *Kmt2a*-KO DMRs. (A) Scatter plot showing transcription factor's motif enrichment in *Dnmt1*-KO associated DMRs. (B,D) Examples of genome-wide significant DMRs in *Dnmt1*-KO mNPCs (blue) compared to controls (orange), located near the neuronal-related gene *Ppfibp1* (B), and a known DNMT1 target gene *Rora* (D) (Tao et al., 2022). (C) Dot plot showing 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. (G,H) Smoothed methylation plots showing two *Dnmt1*-KO DMRs located in an intergenic region near *Gsx1* gene.

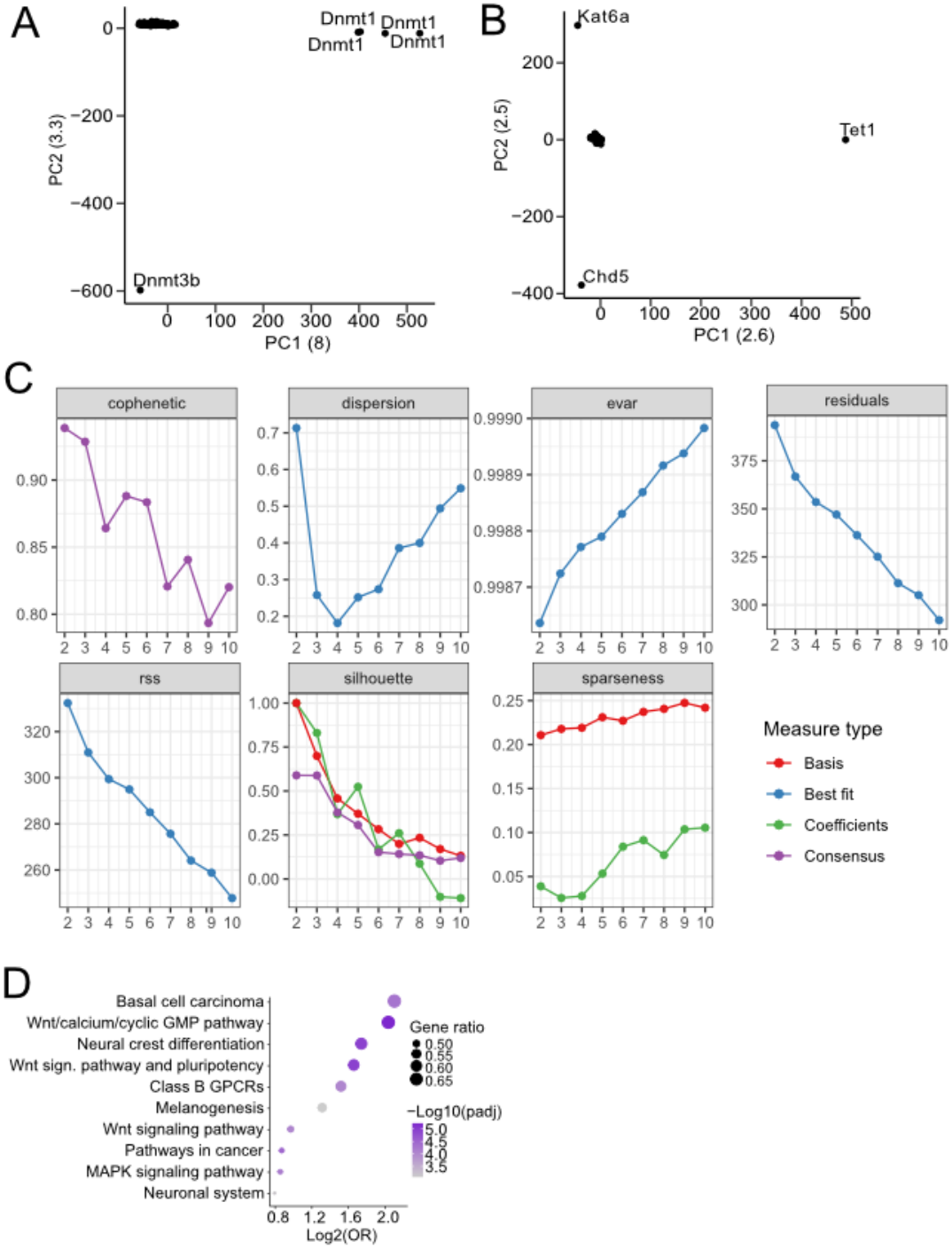


Supplementary Figure 3: RNA abnormalities in *Kmt2a*-KO and *Dnmt1*-KO DEGs. (A-D) Gene ontology (GO) term analysis (Biological Process) for upregulated genes in (A) *Dnmt1*-KOs and (B) *Kmt2a*-KOs, and downregulated genes in (C) *Dnmt1*-KOs and (D) *Kmt2a*-KOs. (E) Representative immunofluorescence images of control, *Kmt2a*-KO and *Dnmt1*-KO mNPCs stained for DAPI (blue) and EAAT2 (magenta). Scale bars: 100 µm. (F)

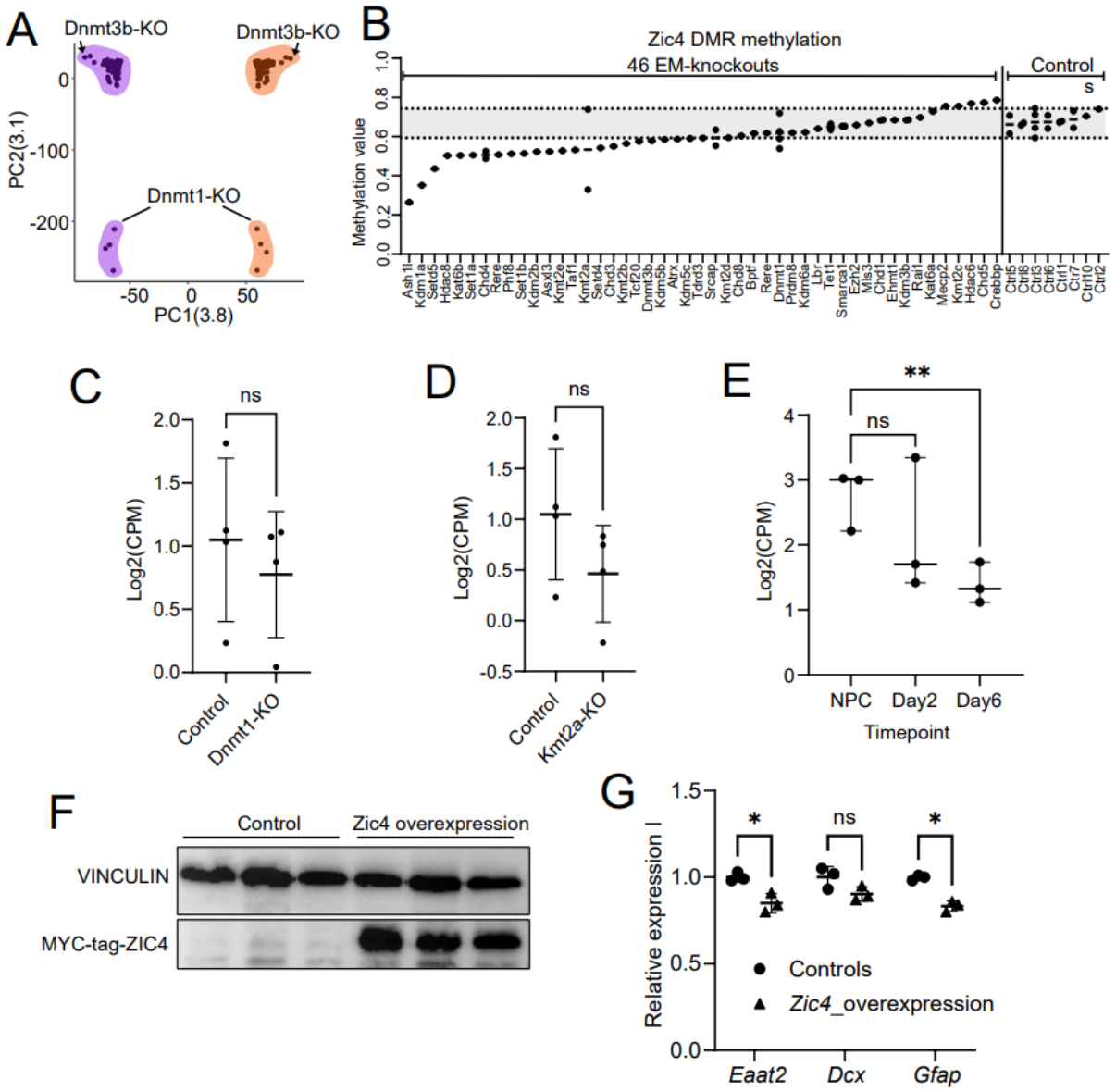
Quantification of EAAT2 mean fluorescence intensity in control, *Kmt2a*-KO and *Dnmt1*-KO mNPCs ($n = 3$, $p < 0.05$, a One-way Anova with Dunnett's correction was used.). (G) WB analysis of PCNA expressions in *Kmt2a*-KO (ns, $p = 0.0695$) and *Dnmt1*-KO (ns, $p = 0.9071$) compared to control mNPC (One-way ANOVA with Dunnett´s correction was used). (H) Representative immunofluorescence images of B6J-Kmt2a$^{+/+}$ and B6J-Kmt2a$^{+/LSL}$ P0 mNPCs, stained for DAPI (blue) and TUBB3 (yellow). (I) Quantification of % of TUBB3 positive cells in B6J-Kmt2a$^{+/+}$ and B6J-Kmt2a$^{+/LSL}$ P0 mNPCs. A student's t-test was used. *$p<0.05$, **$p<0.01$, ns=not significant.



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-D) Normalized counts showing Zic4 expression in *Kmt2a*-KO (not significant, p = 0.1846, C) and *Dnmt1*-KO (not significant, p = 0.5, D), compared to controls. (E) Normalized counts showing Zic4 expression in neurodevelopment, where it is significantly downregulated at Day6 compared with the NPC state (p = 0.00197). (F) WB analysis of Myc-ZIC4 protein after overexpression. (G) 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

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. (2023). Neuron-specific chromatin disruption at CpG islands and aging-related regions in Kabuki syndrome mice. *bioRxiv*. https://doi.org/10.1101/2023.08.01.551456

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

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

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev*, *25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

Do, C., Lang, C. F., Lin, J., Darbary, H., Krupska, I., Gaba, A., Petukhova, L., Vonsattel, J. P., Gallagher, M. P., Goland, R. S., Clynes, R. A., Dwork, A., Kral, J. G., Monk, C., Christiano, A. M., & Tycko, B. (2016). Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am J Hum Genet*, *98*(5), 934–955. https://doi.org/10.1016/j.ajhg.2016.03.027

1049 Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005).
1050 BioMart and Bioconductor: a powerful link between biological databases and microarray data
1051 analysis. *Bioinformatics*, *21*(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525
1052 Elizarraras, J. M., Liao, Y., Shi, Z., Zhu, Q., Pico, A. R., & Zhang, B. (2024). WebGestalt 2024: faster
1053 gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Res*,
1054 *52*(W1), W415–W421. https://doi.org/10.1093/nar/gkae456
1055 Fahrner, J. A., Lin, W. Y., Riddle, R. C., Boukas, L., DeLeon, V. B., Chopra, S., Lad, S. E.,
1056 Luperchio, T. R., Hansen, K. D., & Bjornsson, H. T. (2019). Precocious chondrocyte
1057 differentiation disrupts skeletal growth in Kabuki syndrome mice. *JCI Insight*, *4*(20).
1058 https://doi.org/10.1172/jci.insight.129380
1059 Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., & Liu, H. K. (2013).
1060 The chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC
1061 transcription factors. *Cell Stem Cell*, *13*(1), 62–72. https://doi.org/10.1016/j.stem.2013.05.002
1062 Foroutan, A., Haghshenas, S., Bhai, P., Levy, M. A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi,
1063 A., Pedace, L., Miele, E., Genevieve, D., Heide, S., Alders, M., Zampino, G., Merla, G.,
1064 Fradin, M., Bieth, E., Bonneau, D., Dieterich, K.,…Lebre, A. S. (2022). Clinical Utility of a
1065 Unique Genome-Wide DNA Methylation Signature for KMT2A-Related Syndrome. *Int J Mol*
1066 *Sci*, *23*(3). https://doi.org/10.3390/ijms23031815
1067 Friedrich, G., & Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to
1068 identify and mutate developmental genes in mice. *Genes Dev*, *5*(9), 1513–1523.
1069 https://doi.org/10.1101/gad.5.9.1513
1070 Fu, M. P., Merrill, S. M., Sharma, M., Gibson, W. T., Turvey, S. E., & Kobor, M. S. (2023). Rare
1071 diseases of epigenetic origin: Challenges and opportunities. *Front Genet*, *14*, 1113086.
1072 https://doi.org/10.3389/fgene.2023.1113086
1073 Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C.,
1074 Speed, T. P., Blewitt, M. E., & Ritchie, M. E. (2019). Using long-read sequencing to detect
1075 imprinted DNA methylation. *Nucleic Acids Res*, *47*(8), e46.
1076 https://doi.org/10.1093/nar/gkz107
1077 Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W. B., & Millen, K. J. (2004).
1078 Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker
1079 malformation. *Nat Genet*, *36*(10), 1053–1055. https://doi.org/10.1038/ng1420
1080 Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite
1081 sequencing reads to differentially methylated regions. *Genome Biol*, *13*(10), R83.
1082 https://doi.org/10.1186/gb-2012-13-10-r83
1083 Heberle, E., & Bardet, A. F. (2019). Sensitivity of transcription factors to DNA methylation. *Essays*
1084 *Biochem*, *63*(6), 727–741. https://doi.org/10.1042/EBC20190033
1085 Irion, S., Luche, H., Gadue, P., Fehling, H. J., Kennedy, M., & Keller, G. (2007). Identification and
1086 targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*, *25*(12), 1477–
1087 1482. https://doi.org/10.1038/nbt1362
1088 Isles, A. R. (2022). The contribution of imprinted genes to neurodevelopmental and neuropsychiatric
1089 disorders. *Transl Psychiatry*, *12*(1), 210. https://doi.org/10.1038/s41398-022-01972-4
1090 Jonsson, M. E., Ludvik Brattas, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., Verschuere, S.,
1091 Madsen, S., Hansson, J., Larsson, J., Mansson, R., Meissner, A., & Jakobsson, J. (2019).
1092 Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human
1093 neural progenitors. *Nat Commun*, *10*(1), 3182. https://doi.org/10.1038/s41467-019-11150-8
1094 Kerimoglu, C., Sakib, M. S., Jain, G., Benito, E., Burkhardt, S., Capece, V., Kaurani, L., Halder, R.,
1095 Agis-Balboa, R. C., Stilling, R., Urbanke, H., Kranz, A., Stewart, A. F., & Fischer, A. (2017).
1096 KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions.
1097 *Cell Rep*, *20*(3), 538–548. https://doi.org/10.1016/j.celrep.2017.06.072
1098 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment
1099 and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, *37*(8), 907–915.
1100 https://doi.org/10.1038/s41587-019-0201-4
1101 Kim, S., Kim, H., & Um, J. W. (2018). Synapse development organized by neuronal activity-regulated
1102 immediate-early genes. *Exp Mol Med*, *50*(4), 1–7. https://doi.org/10.1038/s12276-018-0025-1

1103 Kimura, H., & Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for
1104 maintenance DNA methyltransferase, Dnmt1. *J Biol Chem*, *278*(7), 4806–4812.
1105 https://doi.org/10.1074/jbc.M209923200
1106 Korthauer, K., Chakraborty, S., Benjamini, Y., & Irizarry, R. A. (2019). Detection and accurate false
1107 discovery rate control of differentially methylated regions from whole genome bisulfite
1108 sequencing. *Biostatistics*, *20*(3), 367–383. https://doi.org/10.1093/biostatistics/kxy007
1109 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S.,
1110 Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D.,
1111 Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment
1112 analysis web server 2016 update. *Nucleic Acids Res*, *44*(W1), W90–97.
1113 https://doi.org/10.1093/nar/gkw377
1114 Kutner, R. H., Zhang, X. Y., & Reiser, J. (2009). Production, concentration and titration of
1115 pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc*, *4*(4), 495–505.
1116 https://doi.org/10.1038/nprot.2009.22
1117 Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., &
1118 Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput*
1119 *Biol*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118
1120 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–
1121 3100. https://doi.org/10.1093/bioinformatics/bty191
1122 Liang, H., Xiao, G., Yin, H., Hippenmeyer, S., Horowitz, J. M., & Ghashghaei, H. T. (2013). Neural
1123 development is dependent on the function of specificity protein 2 in cell cycle progression.
1124 *Development*, *140*(3), 552–561. https://doi.org/10.1242/dev.085621
1125 Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong,
1126 J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M.,
1127 Feng, X., Fischer, C., Fulton, R. S.,…Paten, B. (2023). A draft human pangenome reference.
1128 *Nature*, *617*(7960), 312–324. https://doi.org/10.1038/s41586-023-05896-x
1129 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for
1130 assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.
1131 https://doi.org/10.1093/bioinformatics/btt656
1132 Liska, O., Bohar, B., Hidas, A., Korcsmaros, T., Papp, B., Fazekas, D., & Ari, E. (2022). TFLink: an
1133 integrated gateway to access transcription factor-target gene interactions for multiple species.
1134 *Database (Oxford)*, *2022*. https://doi.org/10.1093/database/baac083
1135 Lo, H. L., & Yee, J. K. (2007). Production of vesicular stomatitis virus G glycoprotein (VSV-G)
1136 pseudotyped retroviral vectors. *Curr Protoc Hum Genet*, *Chapter 12*, Unit 12 17.
1137 https://doi.org/10.1002/0471142905.hg1207s52
1138 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for
1139 RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-
1140 0550-8
1141 Luperchio, T. R., Boukas, L., Zhang, L., Pilarowski, G., Jiang, J., Kalinousky, A., Hansen, K. D., &
1142 Bjornsson, H. T. (2021). Leveraging the Mendelian disorders of the epigenetic machinery to
1143 systematically map functional epigenetic variation. *Elife*, *10*.
1144 https://doi.org/10.7554/eLife.65884
1145 Mahmood, T., & Yang, P. C. (2012). Western blot: technique, theory, and trouble shooting. *N Am J*
1146 *Med Sci*, *4*(9), 429–434. https://doi.org/10.4103/1947-2714.100998
1147 Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018).
1148 MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol*, *14*(1),
1149 e1005944. https://doi.org/10.1371/journal.pcbi.1005944
1150 Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of
1151 occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770.
1152 https://doi.org/10.1093/bioinformatics/btr011
1153 Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V.
1154 S., Schramm, M., Becher, I., Steidel, M., Noh, K. M., Bergamini, G., Beck, M., Bantscheff,
1155 M., & Savitski, M. M. (2018). Systematic analysis of protein turnover in primary cells. *Nat*
1156 *Commun*, *9*(1), 689. https://doi.org/10.1038/s41467-018-03106-1

Noguchi, H., Kimura, A., Murao, N., Namihira, M., & Nakashima, K. (2016). Prenatal deletion of DNA methyltransferase 1 in neural stem cells impairs neurogenesis and causes anxiety-like behavior in adulthood. *Neurogenesis (Austin)*, *3*(1), e1232679. https://doi.org/10.1080/23262133.2016.1232679

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*, *30*(1), 187–200. https://doi.org/10.1002/pro.3978

Park, D. H., Hong, S. J., Salinas, R. D., Liu, S. J., Sun, S. W., Sgualdino, J., Testa, G., Matzuk, M. M., Iwamori, N., & Lim, D. A. (2014). Activation of neuronal gene expression by the JMJD3 demethylase is required for postnatal and adult brain neurogenesis. *Cell Rep*, *8*(5), 1290–1299. https://doi.org/10.1016/j.celrep.2014.07.060

Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G. W., & Schonhuth, A. (2015). WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J Comput Biol*, *22*(6), 498–509. https://doi.org/10.1089/cmb.2014.0157

Pei, Z., Wang, B., Chen, G., Nagao, M., Nakafuku, M., & Campbell, K. (2011). Homeobox genes Gsx1 and Gsx2 differentially regulate telencephalic progenitor maturation. *Proc Natl Acad Sci U S A*, *108*(4), 1675–1680. https://doi.org/10.1073/pnas.1008824108

Persengiev, S. P., Kondova, II, & Kilpatrick, D. L. (1999). E2F4 actively promotes the initiation and maintenance of nerve growth factor-induced cell differentiation. *Mol Cell Biol*, *19*(9), 6048–6056. https://doi.org/10.1128/MCB.19.9.6048

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*, *34*(7), 695–697. https://doi.org/10.1038/nbt.3583

Plaisier, C. L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C. M., Ding, Y., Reiss, D. J., Paddison, P. J., & Baliga, N. S. (2016). Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Syst*, *3*(2), 172–186. https://doi.org/10.1016/j.cels.2016.06.006

Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M., Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen, T. S., Weissman, J. S., & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol*, *38*(8), 954–961. https://doi.org/10.1038/s41587-020-0470-y

Reynisdottir, T., Anderson, K. J., Brinn, A., Franklin, K., Ouyang, J., Snorradottir, A. O., Lutz, C. M., Zuberi, A. R., DeLeon, V. B., & Bjornsson, H. T. (2024). In-utero rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome. *bioRxiv*, 2024.2007.2019.604339. https://doi.org/10.1101/2024.07.19.604339

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rothbart, S. B., & Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta*, *1839*(8), 627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001

Rybachuk, O., Kopach, O., Pivneva, T., & Kyryk, V. (2019). Isolation of Neural Stem Cells from the embryonic mouse hippocampus for in vitro growth or engraftment into a host tissue. *Bio Protoc*, *9*(4), e3165. https://doi.org/10.21769/BioProtoc.3165

Schoof, M., Launspach, M., Holdhof, D., Nguyen, L., Engel, V., Filser, S., Peters, F., Immenschuh, J., Hellwig, M., Niesen, J., Mall, V., Ertl-Wagner, B., Hagel, C., Spohn, M., Lutz, B., Sedlacik, J., Indenbirken, D., Merk, D. J., & Schuller, U. (2019). The transcriptional coactivator and histone acetyltransferase CBP regulates neural precursor cell development and migration. *Acta Neuropathol Commun*, *7*(1), 199. https://doi.org/10.1186/s40478-019-0849-5

Sheu, J. R., Hsieh, C. Y., Jayakumar, T., Lin, G. Y., Lee, H. N., Huang, S. W., & Yang, C. H. (2019). HDAC6 dysfunction contributes to impaired maturation of adult neurogenesis in vivo: vital role on functional recovery after ischemic stroke. *J Biomed Sci*, *26*(1), 27. https://doi.org/10.1186/s12929-019-0521-1

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*, *14*(4), 407–410. https://doi.org/10.1038/nmeth.4184

Syal, C., Seegobin, M., Sarma, S. N., Gouveia, A., Hsu, K., Niibori, Y., He, L., Wondisford, F. E., Frankland, P. W., & Wang, J. (2018). Ectopic expression of aPKC-mediated phosphorylation in p300 modulates hippocampal neurogenesis, CREB binding and fear memory differently with age. *Sci Rep*, *8*(1), 13489. https://doi.org/10.1038/s41598-018-31657-2

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 a negative regulator of osteogenesis. *Biol Open*, *11*(3). https://doi.org/10.1242/bio.058534

Tycko, B. (2010). Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet*, *19*(R2), R210–220. https://doi.org/10.1093/hmg/ddq376

Veyrac, A., Besnard, A., Caboche, J., Davis, S., & Laroche, S. (2014). The transcription factor Zif268/Egr1, brain plasticity, and memory. *Prog Mol Biol Transl Sci*, *122*, 89–129. https://doi.org/10.1016/B978-0-12-420170-5.00004-0

Wei, Z., Zheng, S., Spangler, S. A., Yu, C., Hoogenraad, C. C., & Zhang, M. (2011). Liprin-mediated large signaling complex organization revealed by the liprin-alpha/CASK and liprin-alpha/liprin-beta complex structures. *Mol Cell*, *43*(4), 586–598. https://doi.org/10.1016/j.molcel.2011.07.021

Wickham, H. (2016). ggplot2. In: Springer International Publishing:.

Wulfridge, P., Langmead, B., Feinberg, A. P., & Hansen, K. D. (2019). Analyzing whole genome bisulfite sequencing data from highly divergent genotypes. *Nucleic Acids Res*, *47*(19), e117. https://doi.org/10.1093/nar/gkz674

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., & Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*, *148*(4), 816–831. https://doi.org/10.1016/j.cell.2011.12.035

You, L., Yan, K., Zou, J., Zhao, H., Bertos, N. R., Park, M., Wang, E., & Yang, X. J. (2015). The lysine acetyltransferase activator Brpf1 governs dentate gyrus development through neural stem cells and progenitors. *PLoS Genet*, *11*(3), e1005034. https://doi.org/10.1371/journal.pgen.1005034

Zhang, J., Ji, F., Liu, Y., Lei, X., Li, H., Ji, G., Yuan, Z., & Jiao, J. (2014). Ezh2 regulates adult hippocampal neurogenesis and memory. *J Neurosci*, *34*(15), 5184–5199. https://doi.org/10.1523/JNEUROSCI.4129-13.2014

Zhang, W., Macias-Velasco, J. F., Zhuo, X., Belter, E. A., Jr., Tomlinson, C., Garza, J., Tekkey, N., Li, D., & Wang, T. (2025). methylGrapher: genome-graph-based processing of DNA methylation data from whole genome bisulfite sequencing. *Nucleic Acids Res*, *53*(3). https://doi.org/10.1093/nar/gkaf028

Zhou, Q., Obana, E. A., Radomski, K. L., Sukumar, G., Wynder, C., Dalgard, C. L., & Doughty, M. L. (2016). Inhibition of the histone demethylase Kdm5b promotes neurogenesis and derepresses Reln (reelin) in neural stem cells from the adult subventricular zone of mice. *Mol Biol Cell*, *27*(4), 627–639. https://doi.org/10.1091/mbc.E15-07-0513

protocol from the MD Anderson Cancer center at University of Texas, available at https://www.mdanderson.org/documents/core-facilities/Functional%20Genomics%20Core/Homemade%204fold%20lentivirus%20concentrator.pdf).

Figure 1

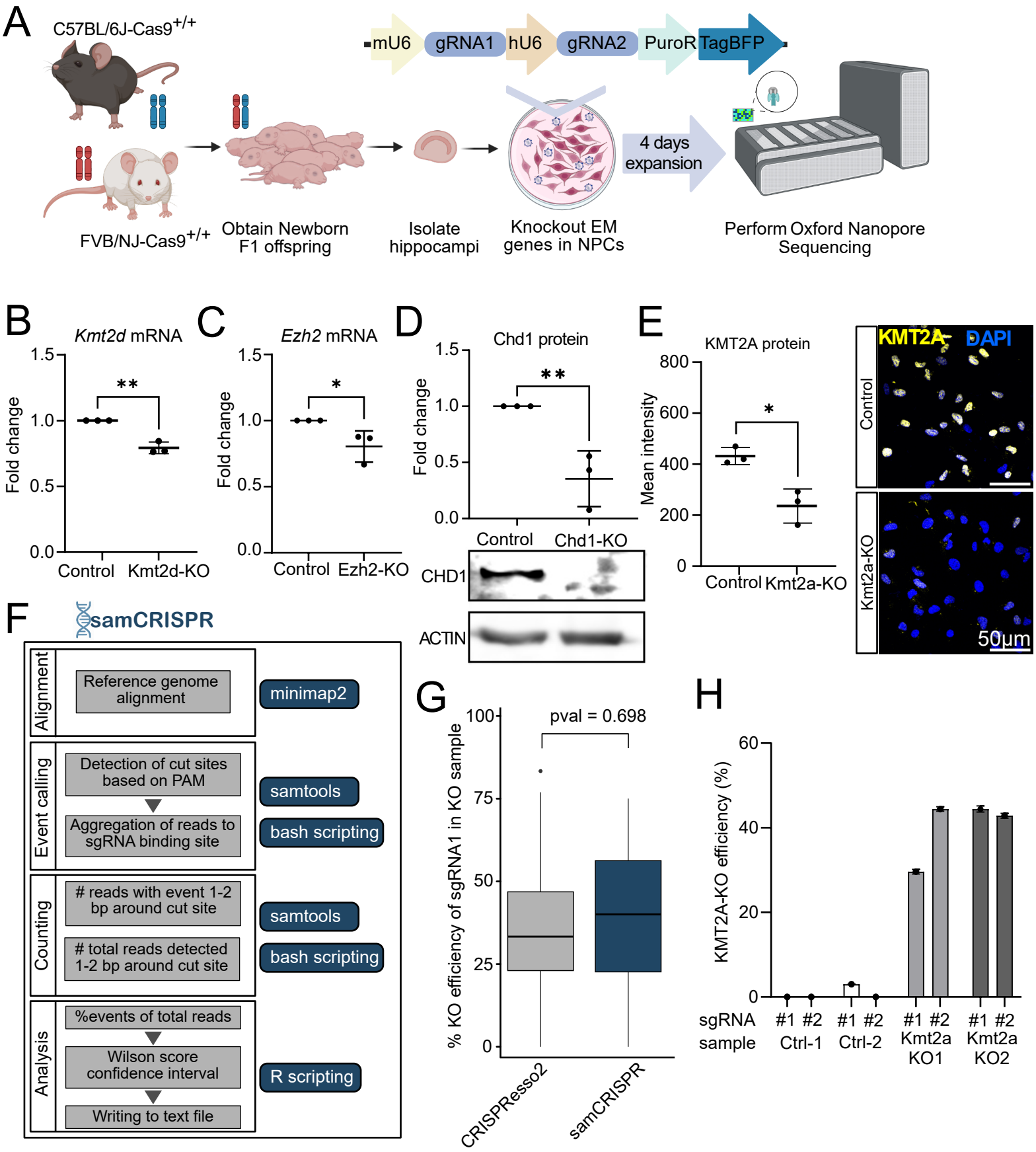

A
C57BL/6J-Cas9+/+
FVB/NJ-Cas9+/+
Obtain Newborn F1 offspring
Isolate hippocampi
mU6
gRNA1
hU6
gRNA2
PuroR
TagBFP
Knockout EM genes in NPCs
4 days expansion
Perform Oxford Nanopore Sequencing
B
Kmt2d mRNA
Fold change
Control
Kmt2d-KO
C
Ezh2 mRNA
Control
Ezh2-KO
D
Chd1 protein
Control
Chd1-KO
CHD1
ACTIN
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

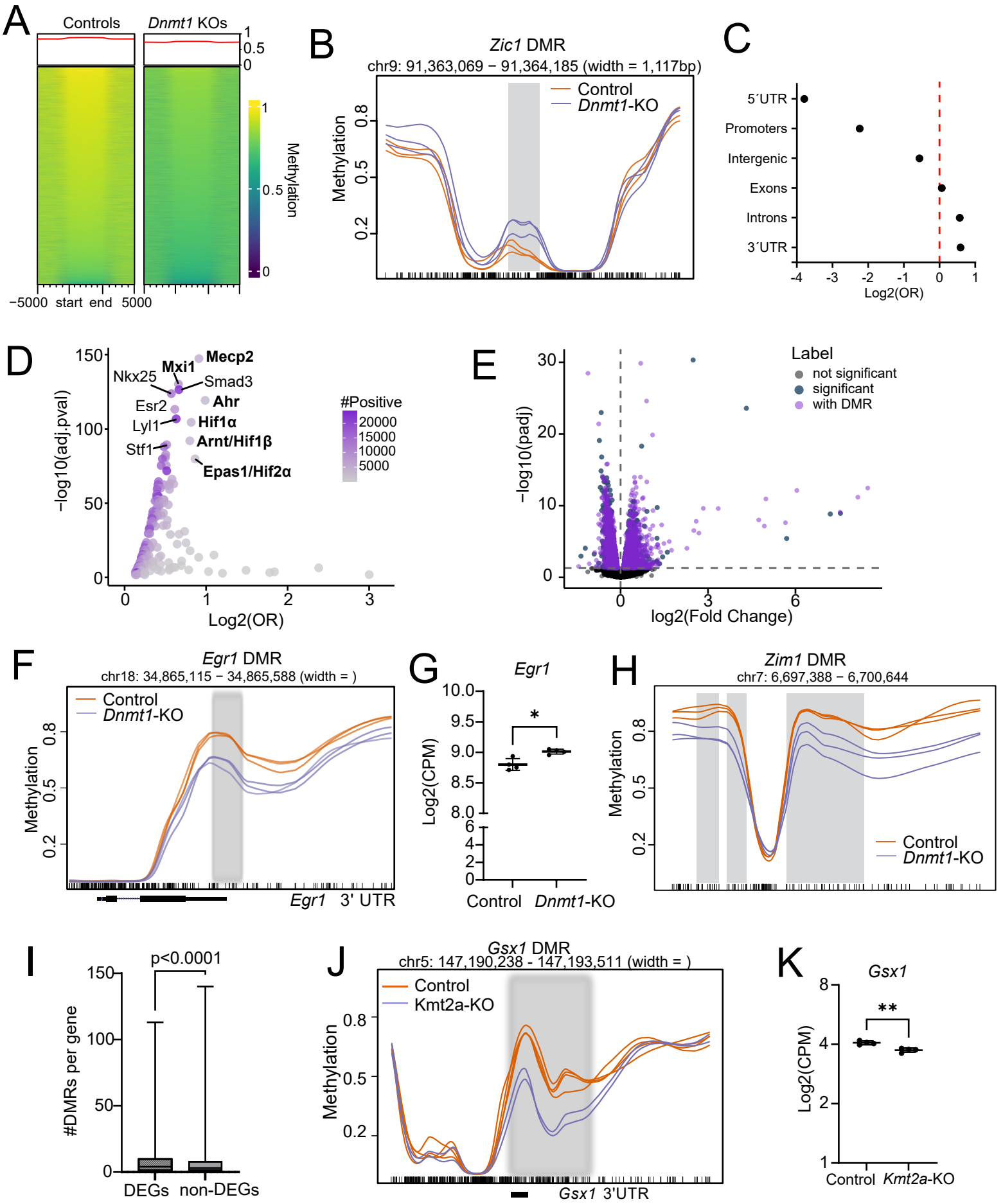

A
Controls
Dnmt1 KOs
Methylation
−5000 start end 5000
B
Zic1 DMR
chr9: 91,363,069 − 91,364,185 (width = 1,117bp)
Control
Dnmt1-KO
Methylation
C
5´UTR
Promoters
Intergenic
Exons
Introns
3´UTR
Log2(OR)
D
Mecp2
Mxi1
Nkx25
Smad3
Ahr
Esr2
Hif1α
Lyl1
Arnt/Hif1β
Stf1
Epas1/Hif2α
#Positive
−log10(adj.pval)
Log2(OR)
E
Label
not significant
significant
with DMR
−log10(padj)
log2(Fold Change)
F
Egr1 DMR
chr18: 34,865,115 − 34,865,588 (width =)
Control
Dnmt1-KO
Methylation
Egr1 3' UTR
G
Egr1
Log2(CPM)
Control
Dnmt1-KO
H
Zim1 DMR
chr7: 6,697,388 − 6,700,644
Methylation
Control
Dnmt1-KO
I
p<0.0001
#DMRs per gene
DEGs
non-DEGs
J
Gsx1 DMR
chr5: 147,190,238 - 147,193,511 (width =)
Control
Kmt2a-KO
Methylation
Gsx1 3'UTR
K
Gsx1
Log2(CPM)
Control
Kmt2a-KO

Figure 3

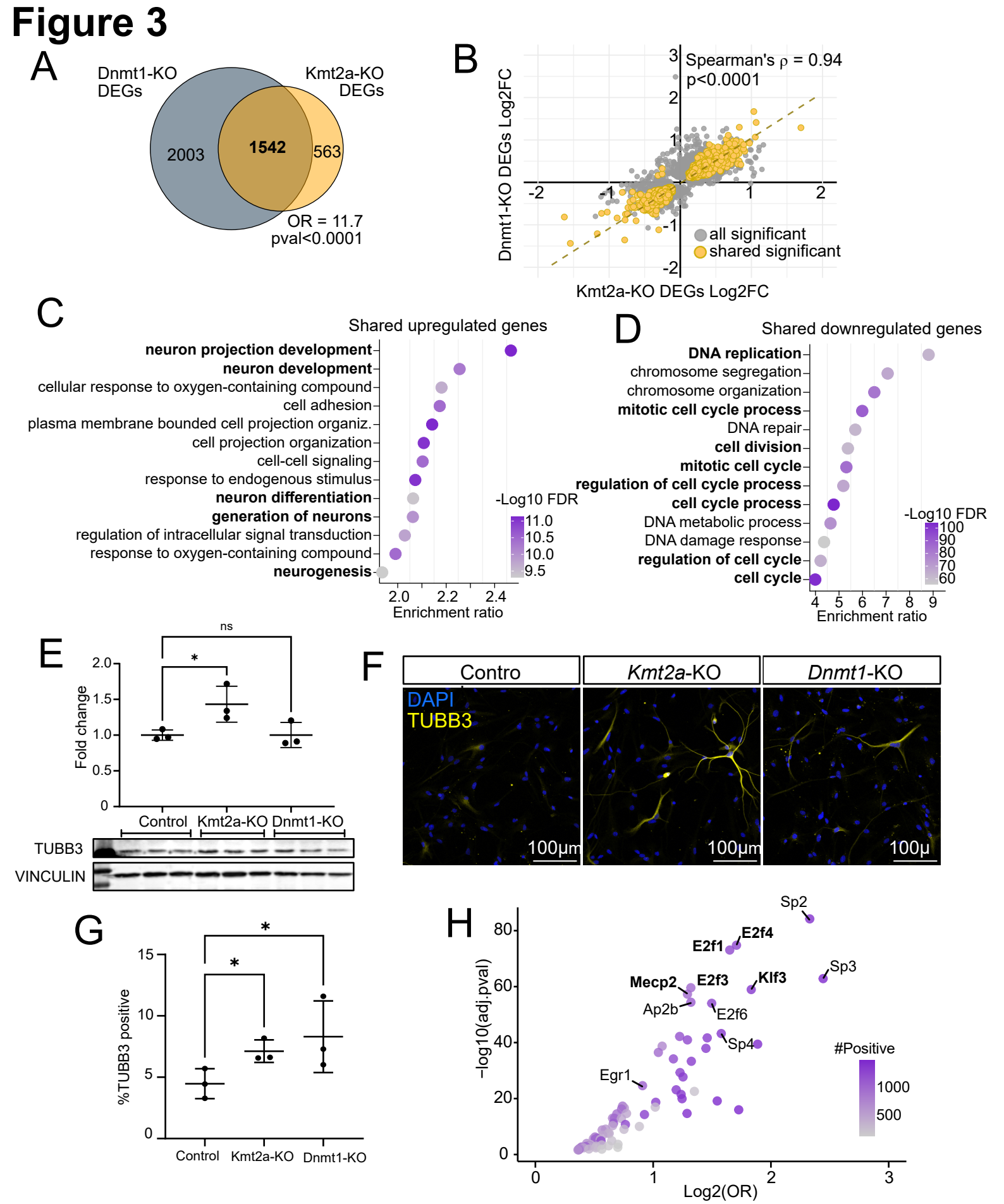

A
Dnmt1-KO DEGs
Kmt2a-KO DEGs
2003
1542
563
OR = 11.7
pval<0.0001
B
Spearman's ρ = 0.94
p<0.0001
Dnmt1-KO DEGs Log2FC
Kmt2a-KO DEGs Log2FC
all significant
shared significant
C
Shared upregulated genes
neuron projection development
neuron development
cellular response to oxygen-containing compound
cell 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
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
Enrichment ratio
E
ns
Fold change
Control
Kmt2a-KO
Dnmt1-KO
TUBB3
VINCULIN
F
Contro
Kmt2a-KO
Dnmt1-KO
DAPI
TUBB3
100µm
G
%TUBB3 positive
Control
Kmt2a-KO
Dnmt1-KO
H
Sp2
E2f4
E2f1
Sp3
Mecp2
E2f3
Klf3
Ap2b
E2f6
Sp4
Egr1
#Positive
−log10(adj.pval)
Log2(OR)

Figure 4

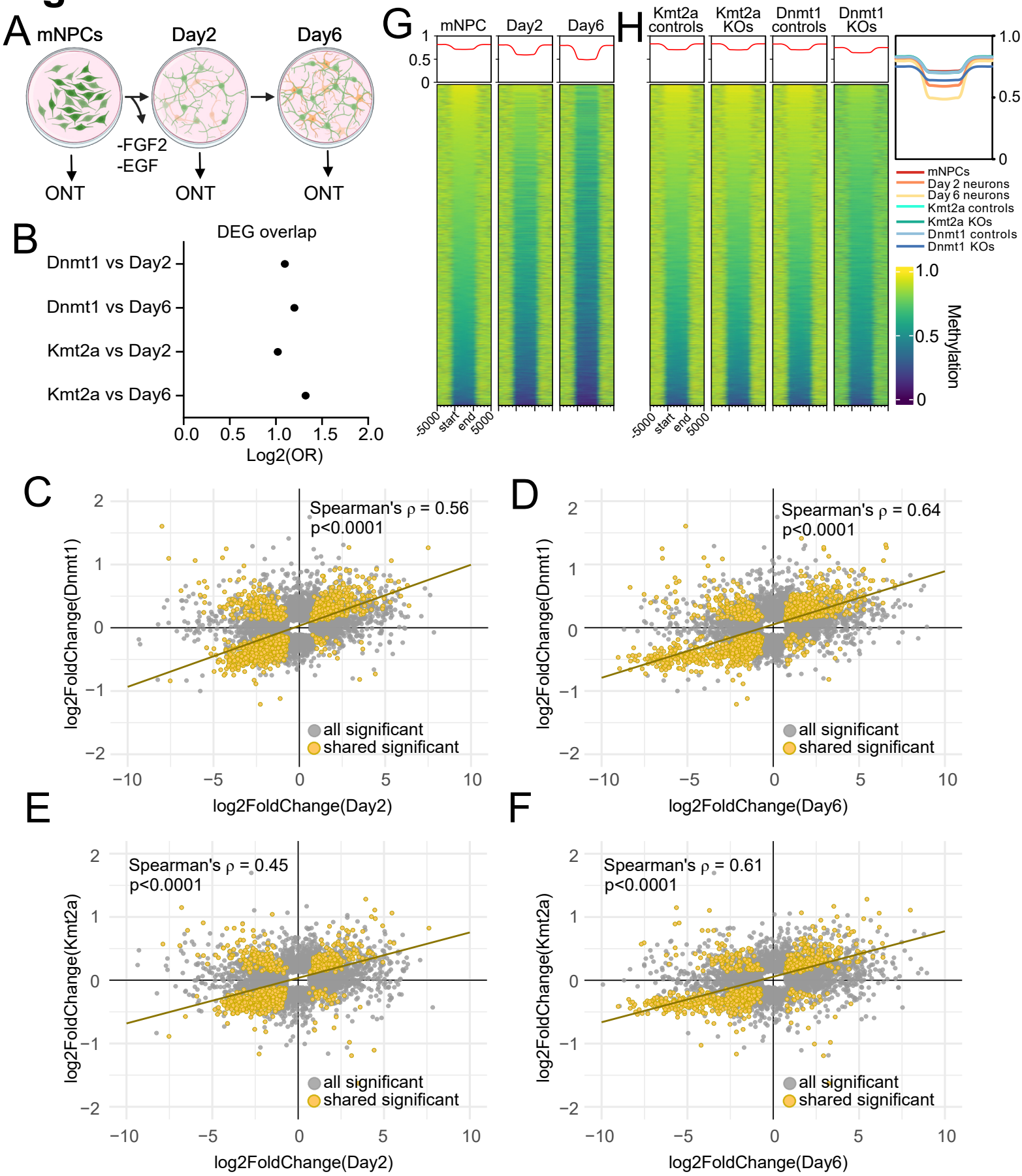

A
mNPCs
Day2
Day6
-FGF2
-EGF
ONT
ONT
ONT
B
DEG overlap
Dnmt1 vs Day2
Dnmt1 vs Day6
Kmt2a vs Day2
Kmt2a vs Day6
0.0
0.5
1.0
1.5
2.0
Log2(OR)
G
mNPC
Day2
Day6
1
0.5
0
-5000
start
end
5000
H
Kmt2a controls
Kmt2a KOs
Dnmt1 controls
Dnmt1 KOs
1.0
0.5
0
mNPCs
Day 2 neurons
Day 6 neurons
Kmt2a controls
Kmt2a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

Figure 5

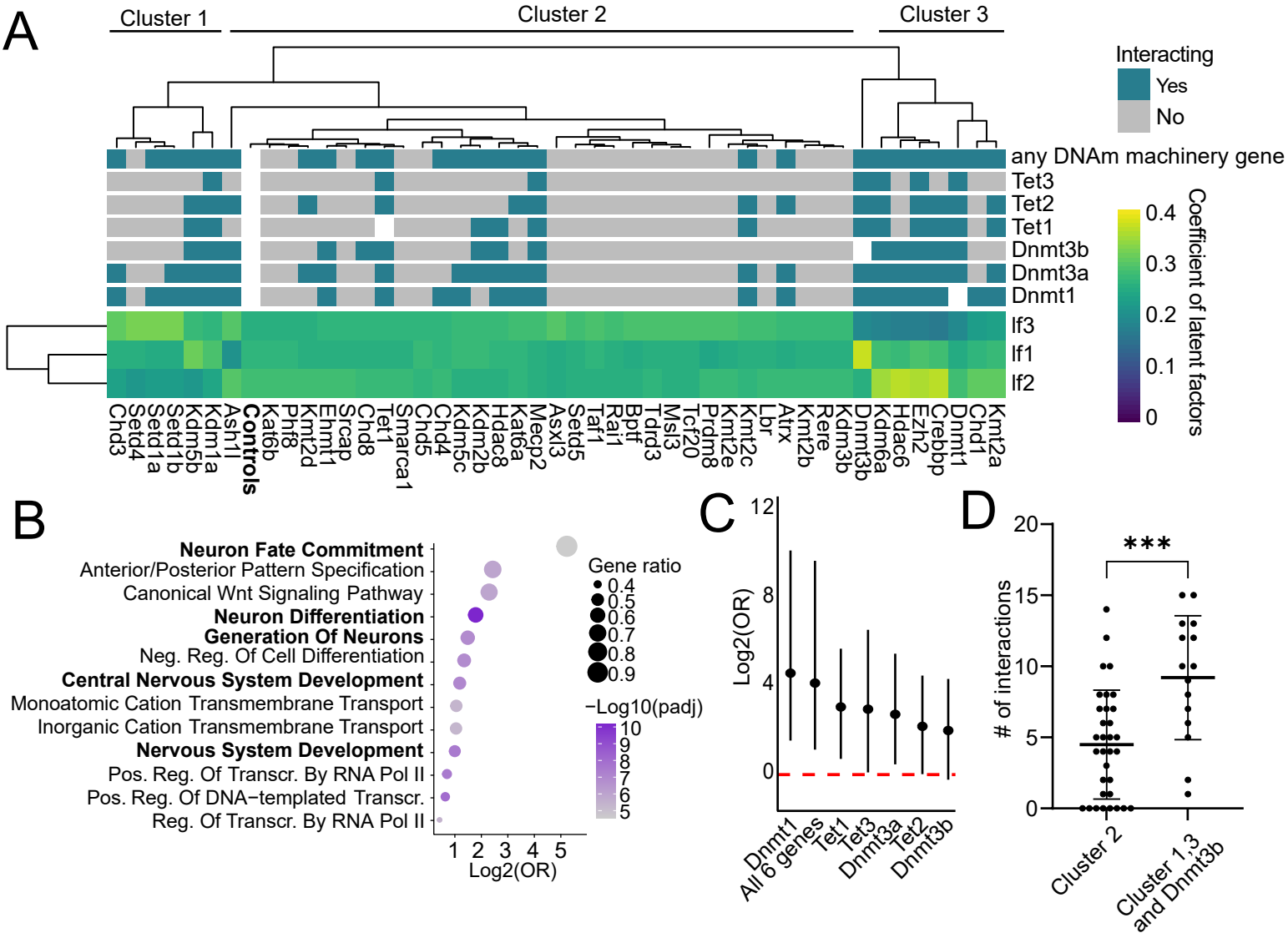
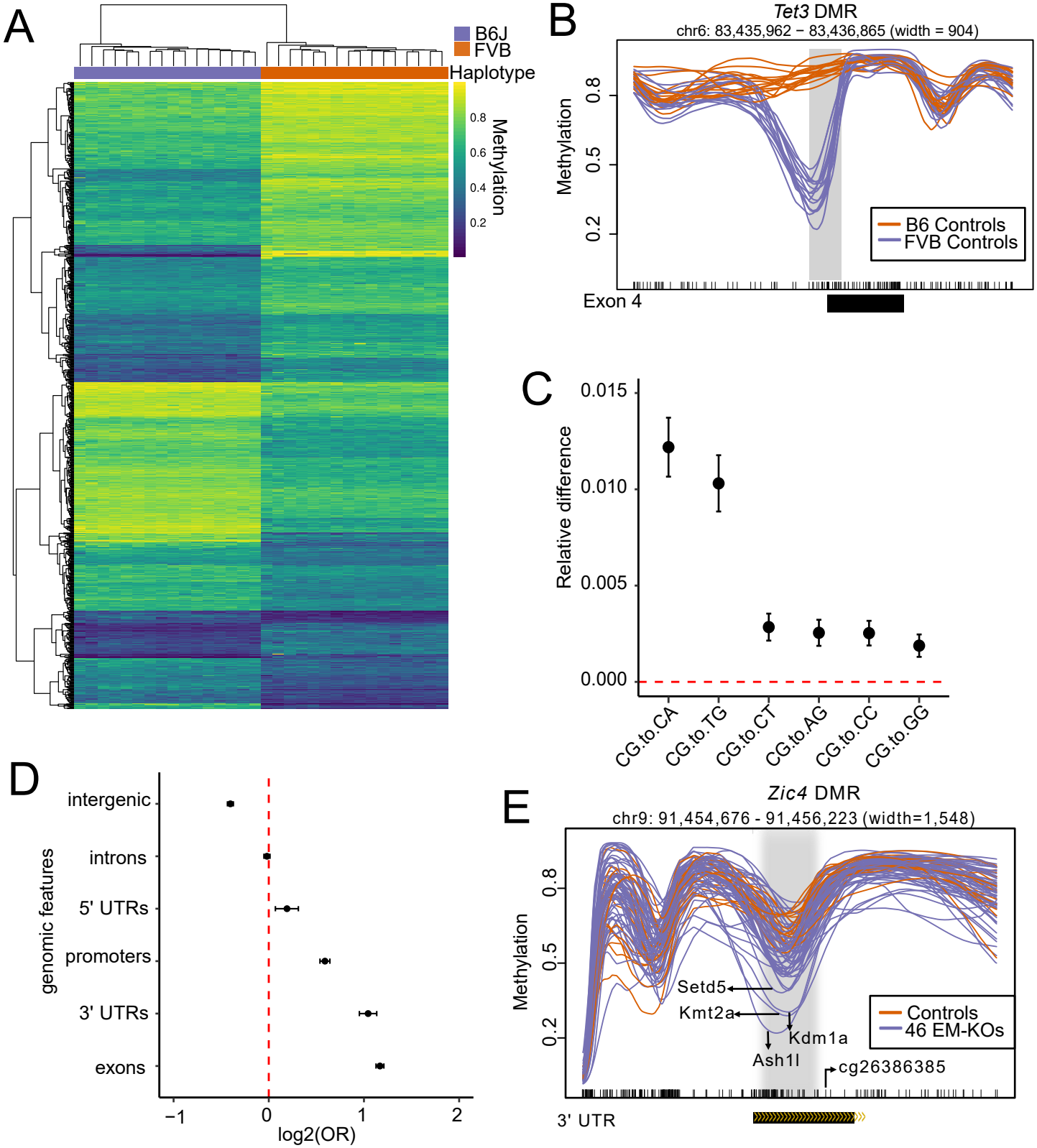

A
Cluster 1
Cluster 2
Cluster 3
Interacting
Yes
No
any DNAm machinery gene
Tet3
Tet2
Tet1
Dnmt3b
Dnmt3a
Dnmt1
If3
If1
If2
Coefficient of latent factors
0.4
0.3
0.2
0.1
0
Chd3
Setd4
Setd1a
Setd1b
Kdm5b
Kdm1a
Ash1l
Controls
Kat6b
Phf8
Kmt2d
Ehmt1
Srcap
Chd8
Tet1
Smarca1
Chd5
Chd4
Kdm5c
Kdm2b
Hdac8
Kat6a
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: 83,435,962 − 83,436,865 (width = 904)
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
genomic features
intergenic
introns
5' UTRs
promoters
3' UTRs
exons
−1
0
1
2
log2(OR)
E
Zic4 DMR
chr9: 91,454,676 - 91,456,223 (width=1,548)
Methylation
0.8
0.5
0.2
Setd5
Kmt2a
Kdm1a
Ash1l
cg26386385
Controls
46 EM-KOs
3' UTR