Shared molecular consequences of epigenetic machinery disruption in murine neuronal progenitors

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

[1]Louma G. Laboratory of Epigenetic Research, Faculty of Medicine, University of Iceland, Reykjavik, Iceland; [2]California Institute of Technology, Pasadena, California, USA; [3]Department of Genetics and Molecular Medicine, Landspitali University Hospital, Reykjavik, Iceland; [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA; [5]Department of Biomedical Engineering, 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 D. Hansen, khansen@jhsph.edu

Abstract

The Mendelian disorders of the epigenetic machinery (MDEMs) are an emerging cause of intellectual disability and growth abnormalities, which commonly disrupt hippocampal function. To investigate consequences of epigenetic machinery (EM) disruption during neurodevelopment, we systematically knocked out (KO) EM factors in neuronal progenitors isolated from the murine hippocampus and established a neurodifferentiation model to interrogate their functions. We then profiled gene expression and DNA methylation (DNAm) in the EM-KOs using RNA sequencing and Nanopore long-read DNA sequencing. While *Dnmt1*-KO induced extensive DNAm alterations, *Kmt2a*-KO had little effect on methylation. Nevertheless, the disruption of *Kmt2a* and *Dnmt1* produced strikingly convergent transcriptional changes. Loss of either EM factor led to premature neuronal differentiation, partially explaining this convergence, and MYC appeared as a shared regulatory node linked to downregulation of cell-cycle programs in these cells. Extending our methylation analysis to 46 EM genes, we found that loss of DNA methyltransferases induced the strongest DNAm changes, whereas other EM-KOs had subtle or negligible effects. However, clustering of EM-KOs based on promoter DNAm levels revealed three distinct EM subgroups, of which two were enriched for interactions with the DNAm machinery. Allele-specific analysis of DNAm further identified a single differentially methylated region shared across the 46 EM-KOs, localized to the FVB allele over the *Zic4* 3'UTR. Furthermore, *Zic4* overexpression appears to maintain the neuronal progenitor state, suggesting functional relevance of this locus. Taken together, our results reveal both gene-specific and convergent effects across diverse EM-KOs and provide new insight into the molecular etiology of the MDEMs.

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

Introduction

In the past 15 years clinicians have witnessed the rapid discovery of a group of at least 90 distinct disorders, caused mainly by heterozygous variants in the components that directly maintain the epigenome. 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 this multitude of names, all parties agree that the disruption of these genes is likely to lead to secondary epigenetic dysregulation. Although these genes have been associated with broad phenotypic presentations (Bjornsson, 2015), they also exhibit unifying disease phenotypes such as intellectual disability and growth abnormalities (Bjornsson, 2015) which are present in the vast majority of patients. We have previously hypothesized that these shared phenotypic abnormalities may be caused by shared epigenetic disruption. In support of this hypothesis, we have identified common signatures of epigenetic dysregulation in three such MDEMs (Kabuki syndrome 1&2 and Rubinstein-Taybi syndromes): identified in blood cells for all three disorders (Luperchio et al., 2021) and in neurons for the first two (Boukas et al., 2024).

Another defining feature of these disorders is that many of them (42) exhibit distinct blood cell DNA methylation (DNAm) episignatures with diagnostic utility (Aref-Eshghi et al., 2020). Remarkably, although blood may not always be the most disease-relevant tissue, many of these disorders have shared epigenetic abnormalities (Aref-Eshghi et al., 2018). Here, we have investigated whether such 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. To do so, 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).

Recent advances in Oxford Nanopore Technologies (ONT) long-read direct DNA sequencing allow for the analysis of DNAm genome-wide and further enables haplotype phasing for the identification of allele-specific effects on DNAm (Gigante et al., 2019). Here, we profiled DNAm abnormalities seen in 46 EM knockouts (KOs) in a single disease-relevant cell-type context, murine neuronal progenitor cells (mNPCs). Our work identifies several transcription factors (TFs) which appear to mediate aspects of the observed chromatin abnormalities or show dysregulation secondarily to the epigenetic abnormality. Our approach,

which enables haplotype phasing of the DNAm data, also highlights the considerable effect that strain-specific genetic variation has on DNAm levels. Together, our results lay the groundwork for understanding how the epigenetic machinery (EM) helps orchestrate hippocampal neurogenesis and how disruption of this process through EM loss uniformly leads to learning and neurological defects.

Results

Systematic KO of EM factors in a single neuronal system

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

To confirm the efficiency of the KO in our system, we performed qPCR, western blot (WB), and immunofluorescence (IF) assays on individual EM-KO mNPCs. As expected, four days after transduction, we routinely saw a decrease in the relevant mRNA and protein levels for individual EM-KOs. As an example, we observed significant ($p < 0.05$) downregulation of *Kmt2d* and *Ezh2* at the RNA level (**Figure 1B-C**) as well as a significant ($p < 0.05$) decrease in protein levels of CHD1, KMT2A, and DNMT1 (**Figure 1D-E, Supplementary Figure 1B-C**) in each of their respective KOs.

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

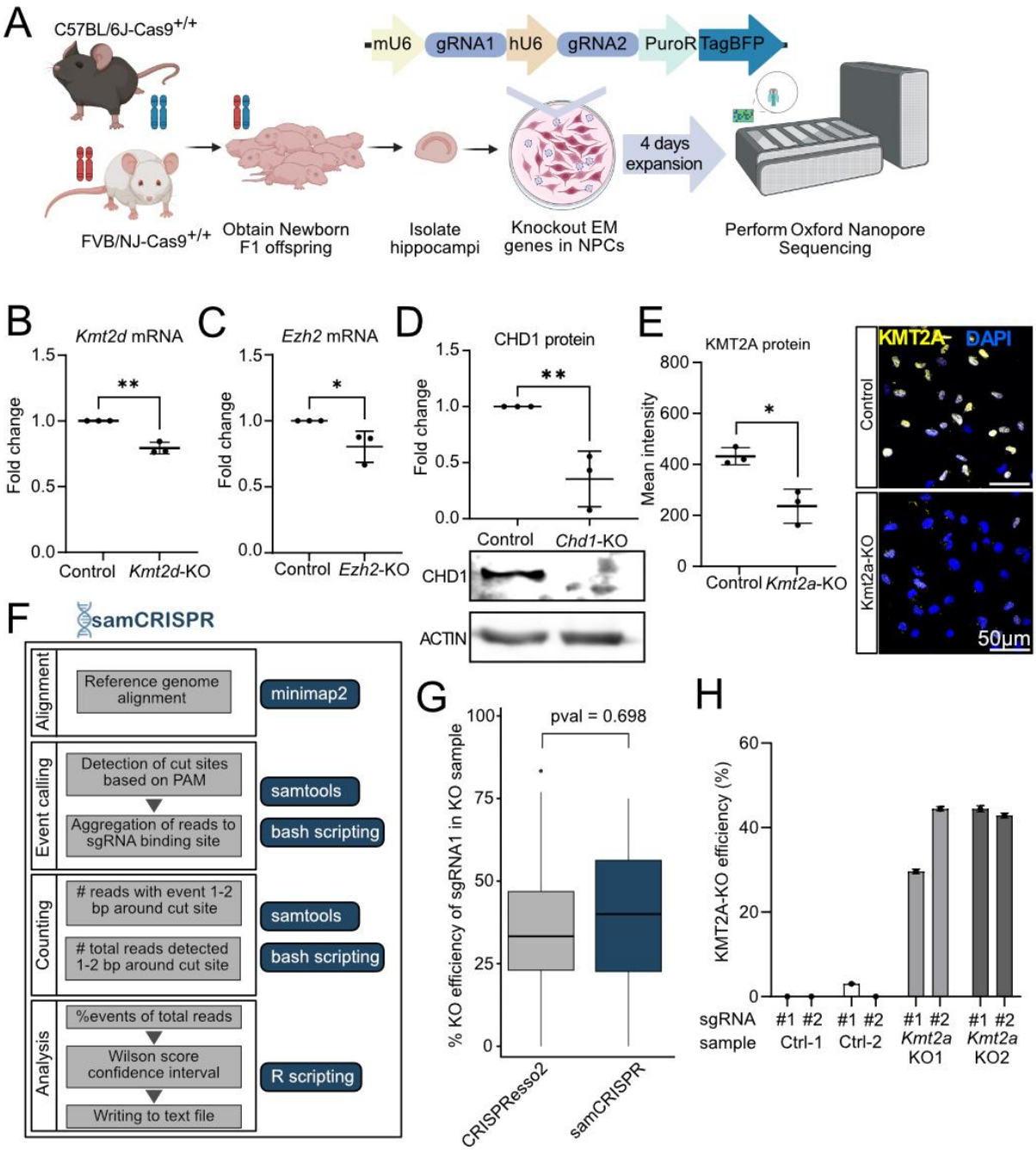


Figure 1: Effective CRISPR-Cas9 targeting in a disease relevant cellular system. (A) Schematic workflow of EM-KO in mNPCs using lentiviral vectors followed by evaluation of DNAm with ONT sequencing. (B) Gene expression changes of *Kmt2d* (p=0.0013) and (C) *Ezh2* (p=0.0229) in mNPCs after knocking out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

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

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

143 Having verified the KO efficiency of our system, we investigated if the disruption of
144 individual EM genes in our mNPCs would result in DNAm abnormalities. As *Dnmt1* encodes
145 the key maintenance DNA methyltransferase, we used it as a positive control for assessing the
146 ONT sequencing approach. We performed ONT sequencing (Jain et al., 2016) on *Dnmt1*-KO
147 and scrambled controls in biological triplicates. Differential methylation analysis revealed a
148 global demethylation pattern in *Dnmt1*-KO cells, including 141,721 hypomethylated
149 differentially methylated regions (DMRs) (qval < 0.05, **Figure 2A, Supplementary Table 2**)
150 and a sole hypermethylated region in *Zic1* gene body (**Figure 2B**). This is consistent with
151 prior studies showing that loss of DNMT1 in human NPCs causes global demethylation
152 (Jonsson et al., 2019). *Dnmt1*-KO DMRs were enriched in 3´UTRs, introns, and exons but
153 depleted in promoters, intergenic regions, and 5´UTRs (**Figure 2C, Supplementary Table 3**).

154 Given the extensive DNAm changes in *Dnmt1*-KO, we explored whether *Dnmt1*-KO
155 DMRs show enrichment of TF binding motifs using an Analysis of Motif Enrichment (AME;
156 MEME-suite, mouse HOCOMOCO v11 core motifs). This analysis identified 170 known TF
157 binding motifs that were significantly enriched (Fisher's exact test, adjusted pval < 0.05,
158 **Figure 2D and Supplementary Table 4**), including a motif for MECP2, which has been
159 shown to recruit DNMT1 to hemi or fully methylated CpGs (Kimura & Shiota, 2003).
160 However, we also found other noteworthy TF motifs, some with CpG sites in their binding
161 motifs, such as HIF1α, AHR and ARNT, and others lacking CpGs in their motifs, including
162 SMAD3, ESR2, NKX2-5 and MXI1 (**Figure 2D, Supplementary Table 4**). Several hypoxia-
163 related TF motifs (HIF1α, EPAS1 and ARNT) emerged. These are of particular interest given
164 our recent observation of dysregulated hypoxia signaling in Kabuki syndrome, another
165 MDEM, caused by mutations in *KMT2D* (Carosso et al., 2019). We next asked if the
166 enrichment of TF motifs in the *Dnmt1*-KO associated DMRs could reflect known DNMT1
167 interaction partners. We compared these motifs with known DNMT1 interactors from
168 BioGRID (Oughtred et al., 2021) and found that 13 of the 24 interactors represented in the
169 HOCOMOCO database exhibited motif enrichment in *Dnmt1*-KO DMRs (**Supplementary
170 Figure 2A**).

171 To investigate whether methylation changes in *Dnmt1*-KOs coincide with changes in
172 transcription, we performed RNA-seq on *Dnmt1*-KO cells. We identified 3,454 differentially
173 expressed genes (DEGs) compared with controls (**Supplementary Table 5**). We used
174 HOMER to annotate the *Dnmt1*-KO DMRs and found that 2,713 of the DEGs (78.5%) are
175 associated with 26,909 of the DMRs (19%; **Figure 2E, purple**). Among these were *Ppfibp1*
176 (**Supplementary Figure 2B**), a gene encoding a factor involved in synaptic protein-protein
177 interactions (PPIs)(Wei et al., 2011), the immediate early genes *Jun*, *Fos* and *Egr1* (**Figure**
178 **2F-G, Supplementary Figure 2C)** (Kim et al., 2018; Veyrac et al., 2014), as well as *Rora*
179 and *Itga8* (**Supplementary Figure 2C-D**), previously identified DNMT1 target genes (Tao et
180 al., 2022). We also found that mouse imprinted genes were significantly overrepresented in
181 the *Dnmt1*-KO associated DEGs (Fisher's exact test, OR = 1.83, p = 0.0118), including *Zim1*
182 (**Figure 2H**)(Kim et al., 1999).

183 We next examined the impact of knocking out KMT2A, a histone methyltransferase
184 which is expected to have a more indirect influence on DNAm compared to DNMT1.
185 Nevertheless, pathogenic variants in *KMT2A* are the cause of Wiedemann-Steiner syndrome
186 (OMIM, 605130) which has a distinct DNAm signature in peripheral blood (Foroutan et al.,
187 2022). To investigate whether *Kmt2a* loss also perturbs DNAm in neuronal progenitors, we
188 performed ONT sequencing on *Kmt2a*-KO NPCs in duplicates, which revealed only three
189 genome-wide significant DMRs compared to controls. One of these DMRs was located in the
190 3'UTR of the *Gsx1* gene (meanDiff = 0.001121656, qval = 0.042, **Figure 2I**). We also
191 performed RNA-seq after *Kmt2a* KO and found 2,105 DEGs compared to controls
192 (**Supplementary Table 6**). Among the previously identified DMRs, only one is associated
193 with a gene exhibiting significant changes in gene expression, *Gsx1* (**Figure 2J**). GSX1 is a
194 TF expressed early in NPCs and is known to promote maturation in specific neuronal
195 subtypes (Coltogirone et al., 2023; Pei et al., 2011). Notably, a prior study reported reduced
196 H3K4me1 near this DMR in *Kmt2a* knockdown neurons derived from a mouse model
197 (Kerimoglu et al., 2017) (**Supplementary Figure 2E**), supporting that *Gsx1* may be a direct
198 target of KMT2A in mNPCs. We then queried GSX1 targets genes in our *Kmt2a*-KO DEGs.
199 Using the TFBS database, which predicts TF-target gene interactions from DNA motifs and
200 ENCODE DNase I footprints (Plaisier et al., 2016), we identified 133 GSX1 target genes
201 (19.2% of 693 expressed targets) that overlapped with *Kmt2a*-KO DEGs. However, this
202 overlap was not significantly enriched (Fisher's exact test, OR = 1.18 and p = 0.055). To
203 further explore whether *Gsx1* regulation might also be influenced by other EM genes, we

204 examined the *Dnmt1*-KO. While we identified two intergenic DMRs near (~30K and ~16.5K
205 upstream) the *Gsx1* locus in *Dnmt1*-KO (**Supplementary Figure 2F-G**), *Gsx1* did not exhibit
206 corresponding transcriptional changes in *Dnmt1*-KO. Interestingly, its downstream targets
207 were significantly enriched amongst *Dnmt1*-KO DEGs (Fisher's exact test, OR = 1.16, p =
208 0.044). Finally, in contrast to the *Dnmt1*-KO DEGs, we found no enrichment of imprinting
209 genes in *Kmt2a*-KO DEGs (Fisher's exact test, OR = 1.23, p = 0.295). Thus, although
210 imprinting is regulated through chromatin state, the effect is likely EM-gene specific.

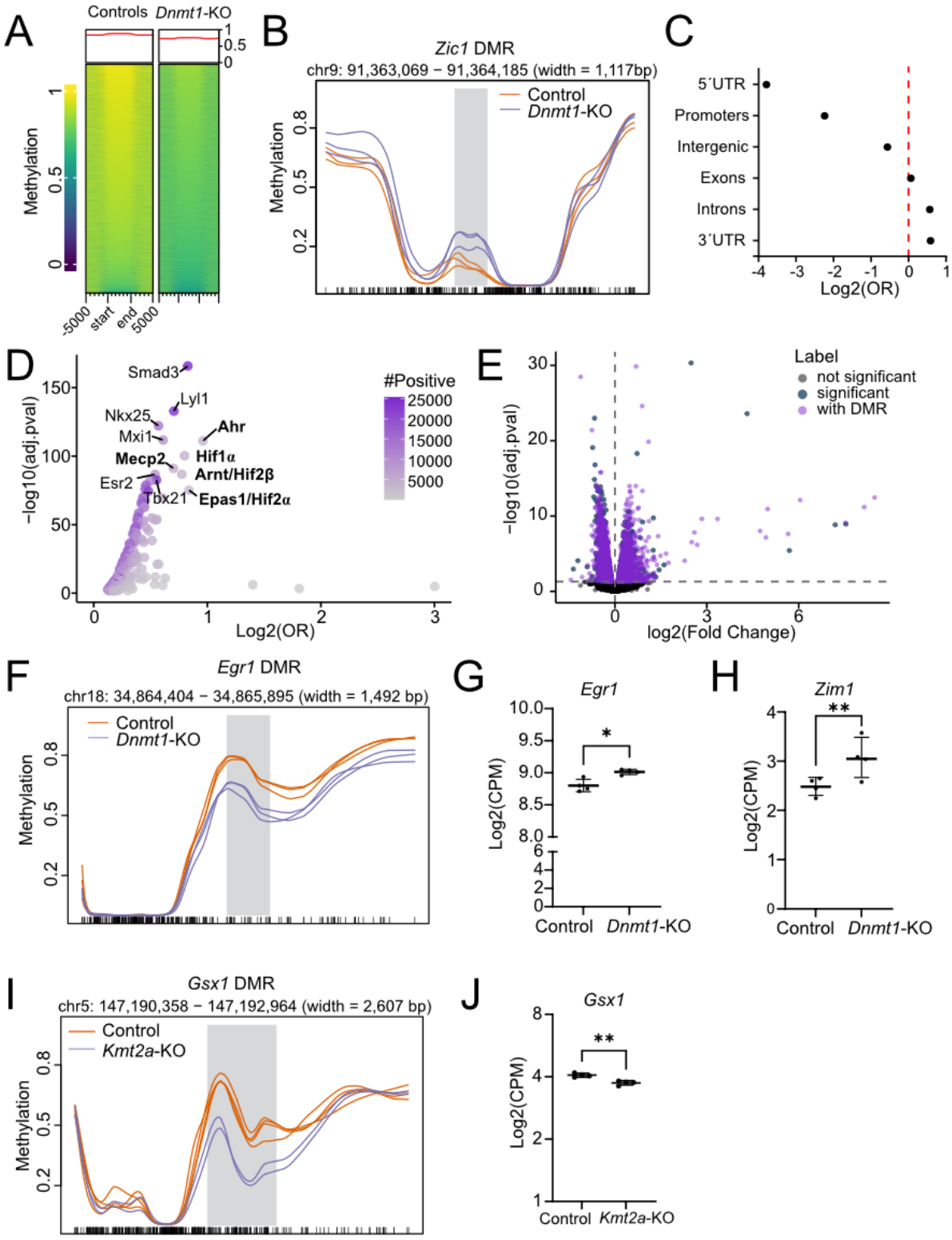


Figure 2: Extent and impact of DNA methylation changes in *Dnmt1*-KO and *Kmt2a*-KO mNPCs. (A) Enrichment heatmaps of smoothed CpG methylation across 141,722 *Dnmt1*-associated DMRs. Each horizontal line represents the average smoothed methylation profile across CpG sites within a DMR and ±5kb flanking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

218 regions of DMR are ±5kb. On the x-axis, each tick mark represents a CpG in the mouse
219 genome. (C) The enrichment analysis of the DMR distribution in *Dnmt1*-KOs, using Fisher's
220 exact test. All categories are significantly enriched or depleted ($p < 0.05$). The exact value is
221 listed in **Supplementary Table 3**. (D) Scatter plot showing enriched TF motifs in *Dnmt1*-KO
222 DMRs (padj < 0.05). (E) A volcano plot showing the distribution of *Dnmt1*-KO DEGs, with
223 significant genes (padj < 0.05) shown in dark blue, and DEGs with DMRs associated shown
224 in purple. (F) *Egr1* DMR (qval < 0.05, gray vertical bar) in *Dnmt1*-KOs (purple) compared to
225 control NPCs (orange). Flanking regions of DMR are ±5kb. On the x-axis, each tick mark
226 represents a CpG in the mouse genome. (G) RNA-seq results showing normalized *Egr1*
227 counts in *Dnmt1*-KOs compared to control NPCs. Adjusted pval < 0.05 (H) RNA-seq results
228 showing normalized *Zim1* counts in *Dnmt1*-KOs compared to control NPCs. Adjusted pval <
229 0.05 (I) *Gsx1* DMR (qval < 0.05, gray vertical bar) found in *Kmt2a*-KO NPCs (orange)
230 compared to control NPCs (purple). Flanking regions of DMR are ±5kb. On the x-axis, each
231 tick mark represents a CpG in the mouse genome. (J) RNA-seq results showing normalized
232 *Gsx1* counts in *Kmt2a*-KOs and controls. *p < 0.05, **p < 0.01, ns = not significant.

233

234 **Convergence of gene expression changes after *Dnmt1* and *Kmt2a* disruption**

235 Given the phenotypic overlap among multiple MDEMs, we next investigated the
236 overlap in molecular consequences between *Kmt2a*-KOs and *Dnmt1*-KOs. When we
237 compared the DEGs from both *Dnmt1*-KO and *Kmt2a*-KO cells, we found a substantial
238 overlap of 1,543 DEGs (44.7% of *Dnmt1*-KO DEGs and 73.3% of *Kmt2a*-KO DEGs),
239 including 679 upregulated and 861 downregulated genes (Fisher's exact test, OR=11.6, $p <$
240 0.0001, **Figure 3A, Supplementary Table 7**). Moreover, gene expression changes were
241 highly correlated between the two datasets (Spearman's $\rho = 0.94$, $p<0.0001$, **Figure 3B**), and
242 almost all shared DEGs were concordant in regard to the direction of change. In both datasets,
243 the upregulated genes were enriched in neuronal differentiation and maturation GO terms
244 (**Figure 3C**), while the downregulated genes were enriched in GO terms linked to cell cycle
245 and chromosome organization (**Figure 3D**). We also performed a gene set enrichment
246 analysis (GSEA) of both datasets, which revealed significant enrichment of downregulated
247 genes in Hallmark terms such as MYC and E2F targets, and G2M checkpoint pathways (FDR
248 < 0.05, **Figure 3E, Supplementary Figure 3A-C**), consistent with reduced proliferative
249 activity. While *Myc* itself, a master regulator of cell-cycle progression (Bretones et al., 2015),
250 was not differentially expressed in either dataset (**Supplementary Figure 3D-E**), many
251 canonical MYC downstream effectors, including E2F genes, Cyclins, *Cdk1*, *Cdk2*,
252 *Cdkn2b/p15*, and *Cdkn1a/p21*, followed the expected direction of change (**Figure 3F,**
253 **Supplementary Figure 3F**)(Bretones et al., 2015). MYC regulators such as *Usp28* and *Skp2*
254 were also downregulated (**Figure 3F, Supplementary Figure 3F**). Given that both KMT2A

255 and DNMT1 have reported physical interactions with MYC on BioGRID (Kalkat et al.,
256 2018), these patterns suggest that disruption of either factor may drive cells towards cell-cycle
257 exit and premature differentiation.

258 We also explored the expression of other EM factors and found that 27 EM genes
259 were differentially expressed in both datasets (**Supplementary Figure 3G**), including *Ezh2*,
260 which has previously been identified as a barrier to neuronal differentiation (Ciceri et al.,
261 2024). Interestingly, both *Dnmt1* and a key interactor *Urhf1* (Liu et al., 2013) are significantly
262 downregulated in both datasets. This suggests that core DNAm machinery is disrupted in both
263 KOs although our data conclusively shows only a minimal effect on DNAm in the *Kmt2a*-KO
264 and therefore establishes that the primary cause of the shared phenotypes is not due to
265 changes in DNAm itself. Finally, *Trim28* and *Setdb1*, co-repressors of transposable elements
266 (Fasching et al., 2015; Kato et al., 2018), were also found significantly downregulated in both
267 datasets.

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

280 To determine whether *Kmt2a*-KO and *Dnmt1*-KO mNPCs undergo premature
281 differentiation, we explored markers for neuronal and glial differentiation in both KOs and
282 scrambled control mNPCs, using WB and IF staining. WB analysis revealed a significant
283 increase in the neuronal marker TUBB3 in *Kmt2a*-KO mNPCs, but not the *Dnmt1*-KO
284 mNPCs, compared to controls (**Supplementary Figure 4A**). No notable differences in the
285 proliferation marker PCNA were detected in either KO model (**Supplementary Figure 4B**).
286 Interestingly, IF staining demonstrated a greater number of TUBB3-positive cells in both
287 *Kmt2a*-KO and *Dnmt1*-KO cultures, compared to scrambled controls, along with distinct

288 morphological changes indicative of differentiation, such as smaller and elongated cell-bodies
289 (**Figure 3H-I**). Increased expression of the astrocyte marker EAAT2 was also observed in
290 *Dnmt1*-KO cultures, whereas no significant change was detected in *Kmt2a*-KO cells
291 (**Supplementary Figure 4C-D**). Prior work has shown prematurely poised neuronal
292 enhancers in primary mNPCs from a *Kmt2a*-haploinsufficient model, a phenotype which
293 normalized after genetic rescue of *Kmt2a* (Reynisdottir et al., 2025). We found that NPCs
294 from this model exhibit a similar premature neuronal differentiation phenotype as in our full
295 KO (**Supplementary Figure 4E-F**), supporting that this is a feature of *Kmt2a* disruption.
296 These findings support that loss of both factors contributes to premature differentiation of
297 mNPCs but that they differ in impact. Specifically, loss of *Kmt2a* leads to a robust and
298 uniform neuronal differentiation phenotype, whereas *Dnmt1* deficiency promotes both
299 neuronal and glial differentiation, but potentially in a more variable manner.

300

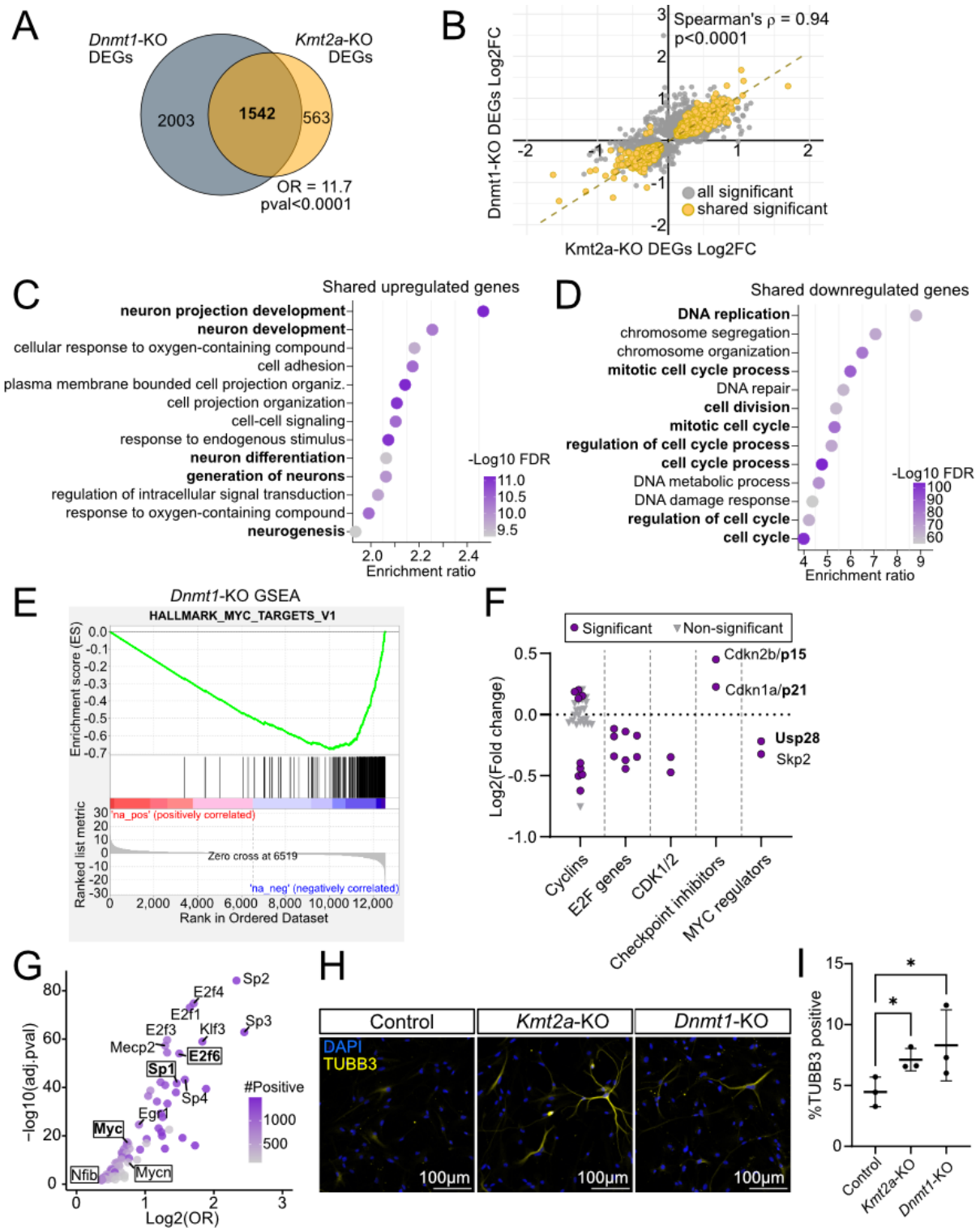


Figure 3: Premature differentiation in *Kmt2a*-KO and *Dnmt1*-KO. (A) A Venn diagram showing extensive overlap of DEGs in *Dnmt1*-KO and *Kmt2a*-KO cells compared to control NPCs, with results from Fisher's exact test (OR = 11.7, $p < 0.001$). (B) Scatter plot showing Log2FC of significant DEGs from both *Dnmt1*-KOs and *Kmt2a*-KOs in gray, and the shared significant DEGs in yellow. Spearman's correlation was calculated for the shared DEGs (R = 0.94, p-val < 0.0001). (C) GO term analysis (Biological Process) of overlapping upregulated

genes, showing enrichment ratio of each GO-term on the x-axis and color depicting the FDR. (D) GO term analysis (Biological Process) of overlapping downregulated genes showing enrichment ratio for each GO-term on the x-axis and color depicting the FDR. (E) GSEA plots showing the MYC_targets_V1 Hallmark category for *Dnmt1*-KO genes. (F) RNA-seq results showing gene expression changes in *Dnmt1*-KO NPCs compared with controls. Y-axis shows Log2fold changes of genes belonging to the categories on the x-axis. Purple dots show significant genes; gray triangles show non-significant genes. (G) Scatter plot of TFs ranked by the number of promoters containing their binding motifs. The odds ratio indicates motif overrepresentation in promoters of shared DEGs. Boxes label shared BioGRID interactors. (H) IF staining of mNPCs after *Kmt2a*-KO and *Dnmt1*-KO compared with controls (blue = DAPI, yellow = TUBB3, magenta = EAAT2). (I) Quantification of the percentage of TUBB3 positive cells in *Kmt2a*-KO, *Dnmt1*-KO and controls. A one-way ANOVA with an FDR correction was used to calculate significance. *$p < 0.05$.

Evaluation of changes that occur during normal differentiation

Given the premature differentiation observed in EM-KOs, some of the expression changes may reflect differences in cell state composition across genotypes. To explore the extent of this in our EM-KOs, we differentiated wild-type (WT) mNPCs into mixed neuronal cultures (see methods) and demonstrated that these cultures express typical markers for neurons and astrocytes, as well as synaptic markers at Day 6 (**Supplementary Figure 5A**). We then performed RNA-seq on mNPCs, Day 2, and Day 6 differentiating cells (schematic in **Figure 4A**), which yielded 3706 and 3441 DEGs at Day 2 and Day 6, respectively, compared with mNPCs (**Supplementary Table 9-10**), and changes were consistent with developmental time-point (**Supplementary Figure 5B**). Both time-points were enriched in neurogenic and synaptic terms, consistent with ongoing neural differentiation and maturation of the culture (**Supplementary Figure 5C-D**). Additionally, a GSEA of cell-types showed significant enrichment in cell-type markers for the three main brain-cell lineages (neurons, astrocytes and oligodendrocytes, **Supplementary Figure 5E**), as would be expected.

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

344 between KO-induced and differentiation-related transcriptional programs, consistent with a
345 premature differentiation phenotype in the previously described EM-KOs.

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

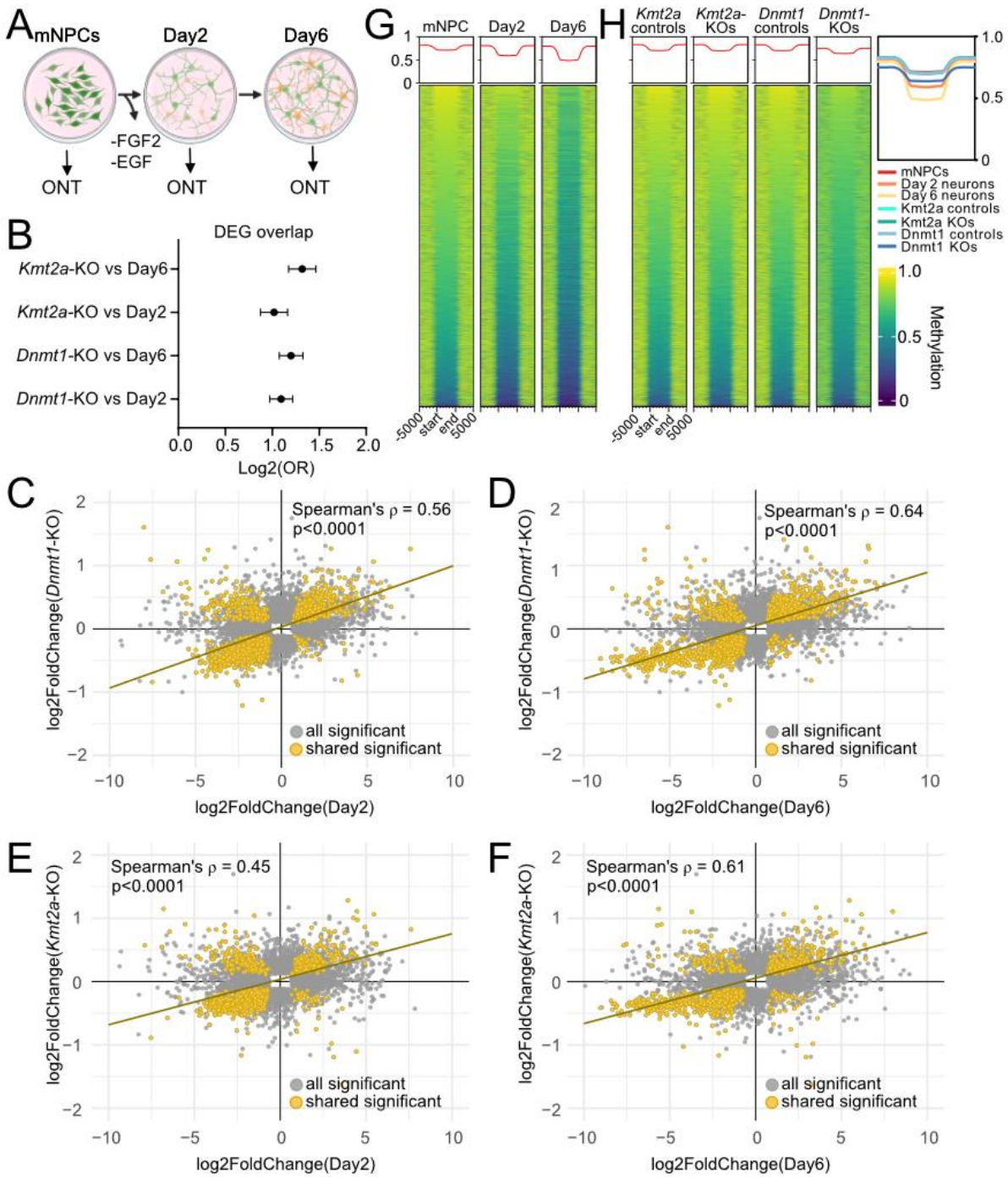


Figure 4: Changes observed in DNAm and gene expression during differentiation of NPCs. (A) A schematic of the neurodevelopmental model. (B) Odds ratio for the overlap between *Kmt2a*-KO and *Dnmt1*-KO DEGs with NPC-Day 2 and NPC-Day 6 DEGs, with 95% confidence intervals shown (Fisher's exact test). (C-F) 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 the corresponding DMR. DMRs were aligned at the DMR start site (start) and end site (end), ±5kb on either side. Yellow shows higher smooth methylation but blue shows lower smooth methylation values.

382

383 **An expanded cohort of EM-KOs shows subtle DNAm changes**

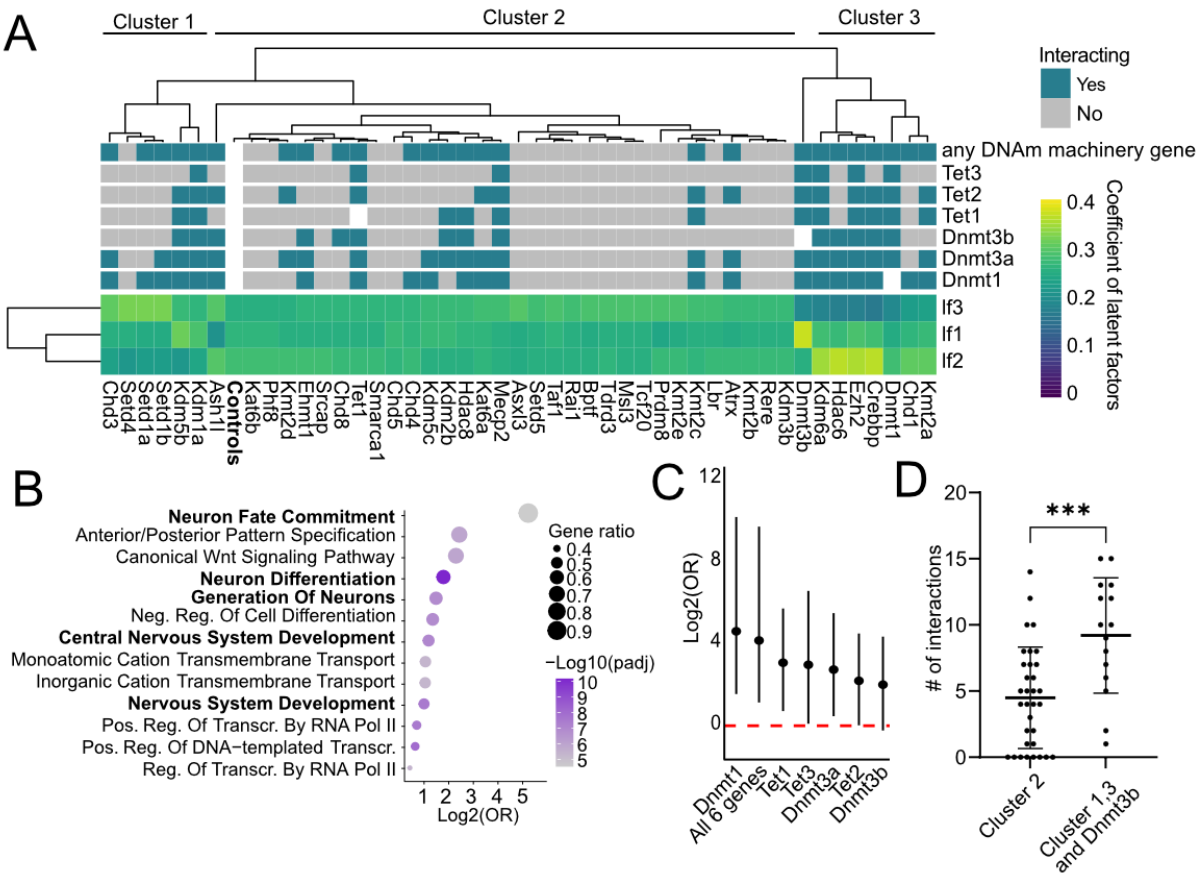
384 We next broadened our analysis to assess DNAm abnormalities across a larger number
385 of EM factors. Using ONT long-read whole genome sequencing, we profiled 44 additional
386 EM-KO cells alongside matched scrambled controls. These data were pooled with previously
387 generated data from the *Dnmt1*-KO and *Kmt2a*-KO, resulting in a total of 46 EM-KOs and 16
388 control samples. To systematically evaluate genome-wide DNAm changes, we partitioned the
389 mouse genome into ~2.7 million 1 kb tiles (each containing > 2 CpGs) and compared
390 methylation levels across groups. A PCA plot of these tiles revealed that two known DNAm
391 machinery components, *Dnmt1* and *Dnmt3b*, showed the most dramatic DNAm abnormalities
392 (**Supplementary Figure 6A**). When these two were excluded from the analysis, *Kat6a*, *Chd5*,
393 and *Tet1* emerged as outliers, clustering separately from the remaining EM-KOs
394 (**Supplementary Figure 6B**). Despite this variation, no shared DMRs were identified across
395 the 46 EM-KOs.

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

397 Given the diverse roles of EM genes in modulating DNAm, we hypothesized that
398 shared DMRs may exist within specific subgroups of EM-KOs with related molecular
399 functions. Therefore, classifying EM genes based on regional DNAm patterns could help
400 uncover convergence. Promoter regions are well-established regulatory elements that are
401 easily assigned to the corresponding genes and involved in transcriptional control (Deaton &
402 Bird, 2011). To investigate whether a subset of EM-KOs exhibit similar DNAm changes at
403 gene promoters, we applied non-negative matrix factorization (NMF) to cluster the 46 EM-
404 KOs based on DNAm patterns at the promoters of 20,482 mouse protein-coding genes
405 (**Supplementary Figure 6C)**. Hierarchical clustering of EM-KOs using the NMF-derived
406 coefficients of latent factors further characterized three latent factors (**Figure 5A**).
407 Specifically, cluster 1 included *Chd3, Setd4, Setd1a, Setd1b, Kdm5b*, and *Kdm1a*, showing the
408 highest coefficient in latent factor 3. In contrast, cluster 2 included the controls and showed
409 low coefficients in all three latent factors. Finally, cluster 3 included the genes *Kmt2a*, *Dnmt1*,
410 *Chd1*, *Kdm6a*, *Hdac6*, *Ezh2*, and *Crebbp* and showed the highest coefficient in latent factor 2
411 (**Figure 5A**). *Dnmt3b*-KO showed the highest coefficient in latent factor 1 and did not cluster
412 with other EM genes (**Figure 5A**). Interestingly, genes with the highest weight in latent factor
413 2 were significantly enriched in GO terms related to generation of neurons and

414 neurodifferentiation (**Figure 5B**) as well as the Wnt signaling pathway (**Supplementary**
415 **Figure 6D**). Conversely, no significant enrichment in GO terms was found for the other latent
416 factor groups. However, separate DNAm analysis comparing each of these clusters to controls
417 did not reveal any genome-wide significant DMRs. In the two clusters that did not group with
418 controls (clusters 1 and 3), nearly all the gene members (excluding SETD4) had STRING
419 defined interactions with the maintenance methyltransferase DNMT1 (**Figure 5A**). Indeed,
420 components of these clusters had a significant enrichment in interactions with all the
421 enzymatic DNAm machinery components (**Figure 5A, C**). This enrichment was strongest for
422 the maintenance methyltransferase (DNMT1) but less so with *de novo* methyltransferases
423 (**Figure 5C**). Furthermore, the same groups (Clusters 1&2 and DNMT3B), had a significantly
424 higher number of STRING interactions with other members of the DNAm machinery (n = 25)
425 (Boukas et al., 2019) compared to what was seen in cluster 2 (**Figure 5D**, one-sided Wilcoxon
426 rank-sum test, W = 349.5, p-value = 0.0014). Additionally, these same groups all involve
427 factors with enzymatic activity, while readers of epigenetic marks are only present in cluster
428 2. These results suggest that enzymatic EM factors, outside the core DNAm machinery, may
429 influence promoter DNAm through interactions with components of the DNAm machinery.

430



431
432 **Figure 5: DNAm analysis of 46 EM-KO genes in NPCs reveals shared promoter effects.**

433 (A) A composite graph showing STRING interactions of EM components with enzymatic
434 DNAm components on top, the heatmap showing coefficient of latent factors for each EM-
435 KO across three latent factors. The color scale represents the coefficient of each latent factor
436 (yellow shows higher but blue a lower latent factor coefficient). (B) Gene ontology (GO) term
437 enrichment (Biological Process) of the genes associated with latent factor 2. Size represents
438 the ratio of genes in the relevant GO term, color indicates the -log10(adjusted p value). (C)
439 Forest plot showing enrichment of enzymatic DNAm machinery-interacting EMs in two
440 relevant clusters compared to rest. Significance testing was done with a Fisher's exact test.
441 Dots represent Fisher's exact test OR ($\log_2$ scale), with lines indicating 95% confidence
442 intervals. The horizontal dotted red line represents statistical significance. (D) Comparison of
443 number of STRING interactions with 25 DNAm machinery components between the clusters.
444 A Wilcoxon rank sum test was used for significance. lf1 = latent factor 1, lf2 = latent factor 2,
445 lf3 = latent factor 3. ***$p < 0.001$, ns = not significant.

446

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

448 We used mNPCs derived from F1 offspring of the two strains (FVB maternal × B6J
449 paternal, **Figure 1A**). As the B6J and FVB mouse strains differ genetically, with a SNV
450 between the two genomes roughly every 630bp, it is possible to phase sequenced reads and
451 perform an allele-specific methylation analysis. This is an average statistic, in practice
452 roughly half of the two genomes can be phased. PCA of the phased data shows the same
453 pattern between KOs and controls as for the unphased data (**Supplementary Figure 7A**) but
454 supports the presence of haplotype-specific methylation between the two strains. Indeed,
455 within the control samples, we identified 1,074 haplotype-specific DMRs (FWER < 0.05)
456 (**Figure 6A, Supplementary Table 13**) (Do et al., 2016; Tycko, 2010), including one located
457 in the exon 4 of *Tet3* (**Figure 6B**). Due to the consistent parental origin of the two genomes in
458 these F1 crosses, imprinted loci will manifest as allele-specific methylation patterns within a
459 single cross direction. Using 23 known ICR regions (Gigante et al., 2019), we have coverage
460 on both haplotypes in 16 of the 23, and all 16 exhibit strong methylation differences between
461 the two alleles (example in **Supplementary Figure 7B**).

462 To determine if the EM-clusters, identified via the NMF analysis (**Figure 5A**) yielded
463 any shared haplotype-specific DMRs, we performed a grouped analysis on the phased data for
464 each of these clusters, compared to the controls. Cluster 1 (*Chd3*, *Setd4*, *Setd1a*, *Setd1b*,
465 *Kdm5b*, and *Kdm1a*) showed a shared hypomethylated DMR localized to the FVB allele 4.6
466 kb upstream of the gene *Ebf1* (FWER = 0.03, **Figure 6C**), which encodes a TF involved in
467 early neurogenesis (Garel et al., 1999; La Manno et al., 2021). In contrast, cluster 2 and 3 did
468 not yield any shared DMRs compared to controls. Finally, we searched for haplotype-specific
469 DMRs associated with the entire group of 46 EM-KOs and identified a single shared

470 hypomethylated DMR localized to the B6J allele (FWER<0.05). This DMR is located in the
471 3' UTR of *Zic4* (**Figure 6D, Supplementary Figure 7C**), a TF that plays a role in cerebellar
472 development (Grinberg et al., 2004). Studies using DNA methylation arrays in humans, have
473 shown that (a) the 3' UTR of this gene is dynamically methylated during differentiation from
474 hESC to neurons (Samara et al., 2022) and (b) methylation of the gene body contributes to an
475 episignature for Sotos syndrome (Choufani et al., 2015). Although *Zic4* was not differentially
476 expressed in *Dnmt1*-KO nor *Kmt2a*-KO mNPCs compared to scrambled controls
477 (**Supplementary Figure 7D-E**), it was significantly downregulated (log2FC = -1.978,
478 adjusted p-val = 0.00197) at Day 6 compared to Day 0 mNPCs in our neurodevelopmental
479 model (**Supplementary Figure 7F**). To evaluate the functional role of ZIC4 in our system,
480 we overexpressed Myc-tagged *Zic4* in mNPCs and compared with the Myc-tag alone. The
481 results showed reduced expression of neuronal maturation markers (**Figure 6E,**
482 **Supplementary Figure 7G**), suggesting that ZIC4 may maintain the NPC state and raising
483 the possibility that it may act as a brake on neuronal maturation.

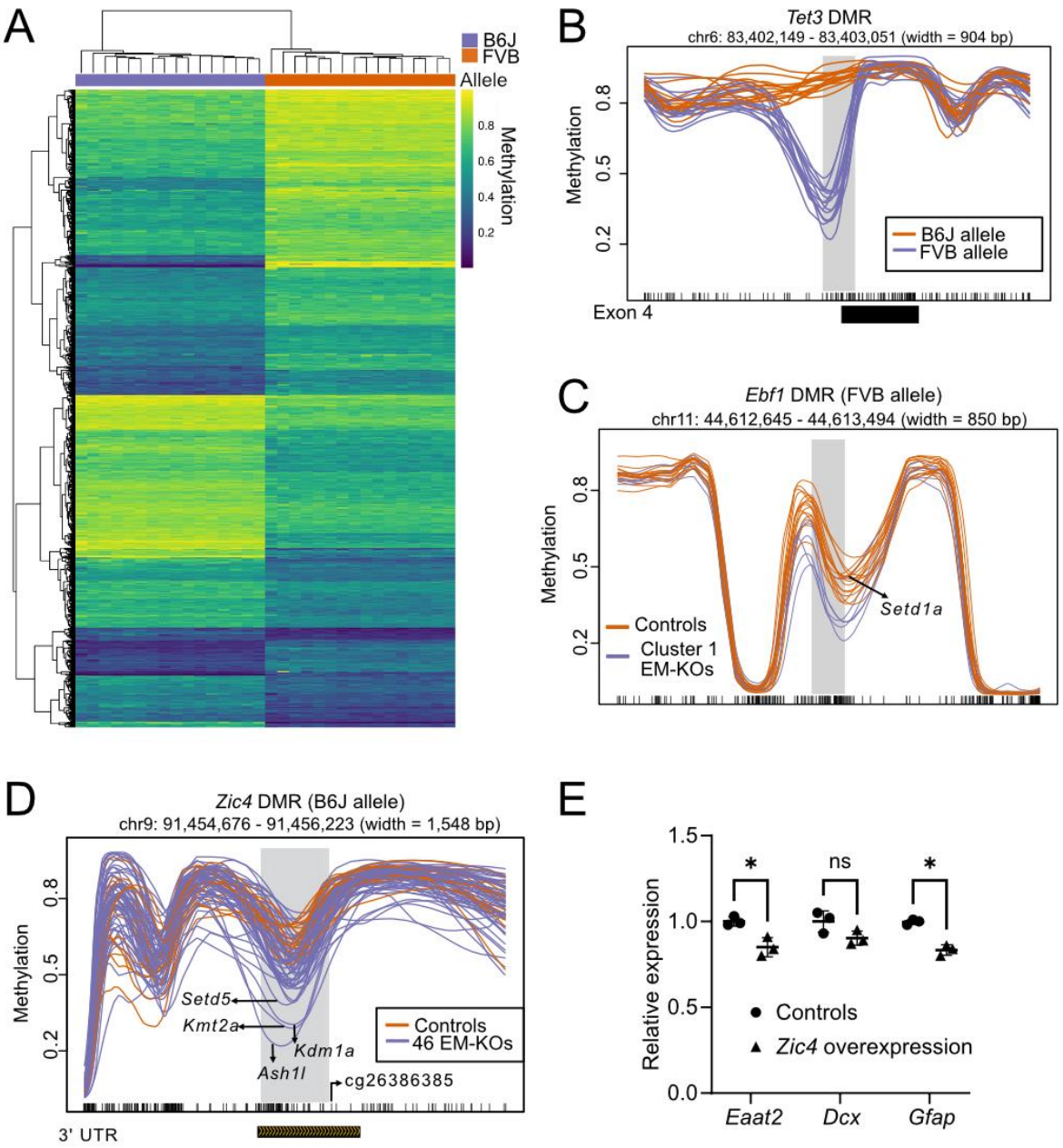


Figure 6: Haplotype-specific changes of DNAm. (A) A heatmap plot showing smoothed methylation of 1074 haplotype-specific DMRs across B6J and FVB haplotypes. Yellow and blue represent higher and lower DNAm, respectively. Each line in y-axis represents a single DMR across haplotyped controls but x-axis represents haplotyped controls. (B) A haplotype-specific DMR in the *Tet3* gene. (C) A line plot showing a DMR in 5' UTR of *Ebf1* between the FVB allele of control mNPC (orange) and the FVB allele of Cluster 1 EM-KOs (purple). The DMR is highlighted in gray. Flanking regions of DMR are ±5kb. On the x-axis, each tick mark represents a CpG in the mouse genome. (D) A line plot showing a DMR in 3' UTR of *Zic4* between the B6J allele of control mNPC (orange) and the B6J allele of 46 EM-KOs (purple). The DMR is highlighted in gray. Flanking regions of DMR are ±5kb. On the x-axis, each tick mark represents a CpG in the mouse genome. (E) WB analysis of Myc-ZIC4 protein after overexpression. (H) qPCR results showing relative mRNA levels of *Eaat2, Dcx* and

498 *Gfap* after ZIC4 overexpression, compared with Myc-only controls; significance calculated
499 using a t-test.

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 transcriptome profiling, we characterized the functional consequences of 46 EM gene KOs. Disruption of *Dnmt1* leads, as expected, to extensive changes in DNAm. A PCA of genome-wide DNAm suggests that the second largest effect on DNAm is from *Dnmt3b* and, interestingly, *Kat6a*, although further investigation is required to assess these changes genome-wide. Since we aimed to identify DNAm changes that were likely to be causative rather than a consequence of the disease phenotype, we isolated and profiled the methylation levels relatively soon after KO (4 days). However, across the 46 EM-KOs, we observe little-to-no shared changes in DNAm and little evidence for KO-specific changes with the exception of a few KOs such as *Dnmt1* and *Dnmt3b*. This does not exclude the possibility that DNAm changes could occur at a later time point. Our data suggests that multiple TFs such as GSX1, MYC, SP and the E2F family TFs may mediate or interpret EM-related epigenetic disruption. For instance, *Gsx1* was differentially expressed in *Kmt2a*-deficient cells, and many of its target genes were dysregulated in both *Kmt2a*-KOs and *Dnmt1*-KOs. Similarly, MYC, SP and E2F family TF binding motifs were enriched in promoters of shared DEGs, and GSEA showed MYC targets significantly negatively enriched. Interestingly, MYC, SP1 and E2F6 specifically are known binding partners of both KMT2A and DNMT1. 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 via interactions with the TFs themselves. Similarly, our clustering of promoter DNAm suggests that, at least in some EM contexts, physical interactions with the DNAm machinery plays a role. Importantly, some TFs such as EGR1 are known to recruit TET1, further linking transcriptional networks to active demethylation pathways (Sun et al., 2019).

Interestingly, we identified a hypomethylated DMR in the 3' UTR of *Zic4* which was shared across all 46 EM-KOs in one of the two haplotypes. This gene is implicated in both murine and human neurodevelopment and DNAm 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 this gene is not differentially expressed in our EM-KOs.

533 Nevertheless, *Zic4* is downregulated during differentiation and overexpression of *Zic4*
534 reduced expression of markers for neuronal and glial differentiation (**Supplementary Figure**
535 **7F, Figure6E**) suggesting that it may play a role in maintaining the NPC state. It is not clear
536 whether this gene is important in explaining the phenotypic overlap in the MDEM group, but
537 it could nevertheless have potential utility as a biomarker.

538 We have established a premature differentiation phenotype arising because of either
539 *Dnmt1*-KO or *Kmt2a*-KO. This is something we and others have observed as a consequence
540 of disrupting several other EM factors. For example, in KMT2A-deficient mice we previously
541 found H3K4me1 enrichment at neuronal enhancers in undifferentiated NPCs (Reynisdottir et
542 al., 2025), suggesting that these enhancers show premature poising. Additionally, prior work
543 on Kabuki syndrome, another MDEM caused by mutations in *KMT2D*, the same phenotype
544 has been reported in two disease relevant contexts (Carosso et al., 2019; Fahrner et al., 2019).
545 Moreover, studies of Kabuki syndrome neurons from a mouse model have shown an
546 overrepresentation of aging-related loci (Boukas et al., 2024), further suggesting that
547 premature differentiation may represent a unifying phenotype across a subset of MDEMs.
548 Interestingly, in our NMF analysis, KMT2A and DNMT1 clustered with EZH2, KDM6A and
549 CREBBP, all of which have been shown to regulate neuronal stem cell maintenance and
550 differentiation (Lei & Jiao, 2018; Pereira et al., 2010; Wang et al., 2010). Genes associated
551 with the latent factor of this group were enriched for functions relating to neuronal
552 differentiation (**Figure 5B**). This could suggest that even though the DNAm changes are very
553 small, the disruption of these EM factors affects a core NPC gene network.

554 Despite the modest DNAm changes, we observed an impressive convergence of
555 transcriptional dysregulation between *Dnmt1*- and *Kmt2a*-KOs, along with the premature
556 differentiation phenotype. Comparing the transcriptional changes in these two EM-KOs to
557 transcriptional changes during normal neuronal differentiation, we conclude that part of the
558 overlap is secondary to the premature differentiation. Nonetheless, we note that the overlap
559 between *Dnmt1*-KO and *Kmt2a*-KO cells appears stronger (larger OR) than either of the
560 overlaps with normal differentiation. Combined with the lack of methylation changes in
561 *Kmt2a*-KOs, we have shown that the expression changes are not a consequence of DNA
562 methylation changes and are more likely to drive this particular phenotype.

563 Our data also points to a coordinated disruption of MYC-dependent regulatory
564 programs after *Kmt2a*- and *Dnmt1*-KO. Although *Myc* transcript levels were unchanged, both
565 KMT2A and DNMT1 are known MYC interactors (Kalkat et al., 2018), and we observed

566 consistent downregulation of MYC target genes in the expected direction (**Figure 3E-F**). This
567 pattern strongly suggests reduced MYC activity. As MYC is a master cell-cycle regulator
568 (Dong et al., 2014; García-Gutiérrez et al., 2019), either the loss of interaction with DNMT1
569 or KMT2A or the downregulation of its deubiquitinator USP28 (Popov et al., 2007), might be
570 a critical factor contributing to the premature cell-cycle exit and neuronal differentiation
571 phenotypes we observe. Further work is required to understand how the shared expression
572 changes arise, and especially to identify which epigenetic changes, if any, are causal in this
573 context. Furthermore, differentiation state should be carefully considered when analyzing
574 gene expression, since premature or altered differentiation can be a major source of the
575 observed changes and thus a confounder of such studies.

576 Beyond providing mechanistic insight, this work establishes a framework for using
577 long-read sequencing, haplotype phasing, and transcriptomic profiling in disease-relevant
578 cells to understand the role of epigenetic regulators in health and disease.

579 **Materials and methods**

580 **Animals**

581 B6J.129(Cg)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J(Strain#026179) (B6J) and
582 FVB.129(B6)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J(Strain#026558) (FVB) mice,
583 each carrying a heterozygous Cas9 transgene, were obtained from Jackson laboratory. Mice
584 were bred within their respective background strains (B6J and FVB) to generate homozygous
585 Cas9-expressing offspring. DNA was isolated from ear clips, and Polymerase Chain Reaction
586 (PCR) was used to confirm the Cas9 cassette using primers design provided by the Jackson
587 laboratory (**Supplementary Table 14**). All mice were kept on a 12 h dark-light cycle at room
588 temperature (21-23°C) with relative humidity around 40% and maintained with ad libitum
589 diet and water. Mouse importation and experimental protocols were in accordance with
590 Icelandic Food and Veterinary Authority (license no. 2208602) and approved by the National
591 Expert Advisory Board on Animal Welfare.

592 **Dissections and primary NPCs isolation**

593 The NPCs were isolated from an F1 hybrids of homozygous Cas9 mice (B6J and FVB). The
594 isolation procedure was based on published protocols (Bernas et al., 2017; Rybachuk et al.,
595 2019) with minor modifications, see Supplementary Materials for details. NPCs were grown
596 in neuronal growth medium (NGM), which included Neurobasal medium supplemented with

597 1% B27 supplement (ThermoFisher, 17504044), 1% Glutamax (ThermoFisher, 35050061),
598 0.25% Penicillin-streptomycin (ThermoFisher, 15140122) 20 ng/ml EGF (PeproTech, AF-
599 100-15), 20 ng/ml FGF2 (PeproTech, 100-18B) and 2 µg/ml Heparin (MP Biomedicals,
600 0210193125) on PDL/Laminin coated 6-well plates, and frozen down in NGM + 10% DMSO
601 (Santa Cruz, sc-358801).

602 **Neuronal progenitor differentiation**

603 NPCs (passage 3-4) were differentiated into mixed neuronal cultures by collecting them as
604 previously described, counting and seeding 150.000 cells/well on a PDL/Laminin coated 24-
605 well plate in 0.5 ml N2-Neuronal Induction Medium (N2NIM) consisting of Neurobasal Plus
606 Medium (ThermoFisher, A3582901) with 1x B-27 Plus Supplement (ThermoFisher,
607 A3582801), 1x CultureOne Supplement (ThermoFisher, A3320201), 1x N2 Supplement
608 (ThermoFisher, 17502048), 0.25x Penicillin-streptomycin and 0.25x GlutaMax. This was
609 supplemented with 20ng/ml FGF2 and 10ng/ml BDNF (Peprotech, 450-02). On day 2 and 4,
610 the medium was changed to N2NIM supplemented with 10ng/ml FGF2, 10ng/ml BDNF and
611 10ng/ml GDNF (ORF Genetics, IK0900). At Day2 and Day6 of differentiation, neuronal
612 cultures were collected and DNA and RNA isolated as described below.

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

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

625 **Cell culture**

626 For virus production, HEK293T cells were cultured in culturing media, DMEM/F12
627 supplemented with GlutaMAX™ (ThermoFisher, 31331093) and 10% Fetal Bovine Serum

628 (FBS) (ThermoFisher, 10270106). All cells were incubated at 37°C with 5% CO_2. HEK293T
629 cells were frozen down in culture media supplemented with 10% DMSO.

630 **Virus production and transduction into NPCs**

631 HEK293T cells in a T75 flask were transfected to produce lentivirus. The transfection mixture
632 contained 2.29 µg lentiviral CRISPR plasmid, 2.08 µg psPAX2 (Addgene, 12260), 0.63 µg
633 pMD2.G (Addgene, 12259), combined with 15 µl Lipofectamine 3000 and 10 µl P3000
634 (ThermoFisher, L3000015), and brought to a final volume of 250 µl with Opti-MEM™
635 (ThermoFisher, 31985047). The transfected HEK293T cells were cultured for 72 hours, and
636 viral particles collected from the medium after 48 and 72 hours, passing the supernatant
637 through a 0.45 µm syringe filter (Santa Cruz, sc-358814) to remove cell debris. After the
638 second collection, 4× Lentivirus concentrator solution was prepared following the protocol
639 from the MD Anderson Cancer Center Functional Genomics Core, and was used to precipitate
640 viral particles (Kutner et al., 2009; Lo & Yee, 2007). Fifty µl viral aliquots were stored in -
641 80°C until use. Transduction efficiency for each virus was tested in NPCs at several
642 concentrations, using serial dilutions of the virus together with 8 µg/ml polybrene (Santa
643 Cruz, sc-134220). Transduction efficiency was determined using fluorescence-activated cell
644 sorting (FACS) of BFP fluorescent cells in NPC cells, 48 hours after transduction, and an
645 efficiency of ~30% was used to conduct all experiments.

646 **EM-KO in NPCs and validation**

647 For EM-KO experiments, the NPCs were seeded on a 6-well plate 24 h prior to transduction.
648 Cells were transduced with lentivirus in the presence of 8 µg/ml polybrene. Each lentivirus
649 targeted a single EM gene and non-targeting controls were used as negative controls for viral
650 infection and culturing effects. After 24 h, cells were split and selected by 0.5 µg/ml
651 puromycin (ThermoFisher, A1113802) for 2 days. Cells surviving puromycin treatment were
652 cultured for another day without puromycin to allow them to recover, then harvested for
653 experiments using Accutase as described above. Timing was selected based on the fact that
654 average protein half-life in mouse primary neurons is ~86 hours (Mathieson et al., 2018) and
655 for EM proteins it is considered even less (~40 hours), thus most of the EM proteins would be
656 expected to be degraded at the time of our evaluation. The KO efficiency was determined by
657 PCR, WB and IF staining.

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

659 Cultured cells were harvested and lysed in TRIzol™ Reagent (ThermoFisher, 15596018).
660 Total RNA was isolated using Direct-zol™ RNA Microprep Kit (Zymo Research, R2062)
661 based on the manufacturer's protocol. The concentration and quality of RNA was evaluated
662 with Nanodrop (ThermoFisher). One µg total RNA was used for cDNA synthesis with High-
663 Capacity cDNA Reverse Transcription kit (ThermoFisher, 4368814). RT-qPCR reactions
664 were performed using Luna® Universal qPCR Master Mix (NEB, M3003) and a CFX384
665 Touch Real-Time PCR Detection System (BIO-RAD), according to manufacturer's protocol.
666 Any primers used are provided in **Supplementary Table 14.**

667 **Protein isolation and WB**

668 NPCs were collected using Accutase and counted with a Countess II FL Automated Cell
669 Counter (ThermoFisher). Cells were lysed at 15000-20000 cells/µl in 1x sample buffer
670 (60mM Tris HCl pH 6.8, 2% SDS, 10% Glycerol, 0.01% Bromophenol blue, 1.25% Beta-me)
671 supplemented with 1:100 Halt™ Protease Inhibitor Cocktail (ThermoFisher, 78437), 1:100
672 phosphatase inhibitor Cocktail (Cell signaling, 5870) and 1:100 PMSF Protease Inhibitor
673 (ThermoFisher, 36978). The lysate was heated at 95°C for 5 min and 1:100 Benzonase
674 nuclease (Sigma-Aldrich, sc-202391) added to degrade genomic DNA. The lysate was
675 aliquoted and stored in -80°C. Western blotting was performed as has previously been
676 described (Mahmood & Yang, 2012), where 15-25 µg of protein was loaded per sample,
677 blotted onto PVDF membrane (ThermoFisher, 88520) and stained overnight using the
678 following primary antibodies: rabbit anti-CHD1, 1:1000 (Cell signaling, 4351), mouse anti-
679 ACTIN, 1:1000 (Abcam, ab8224), rabbit anti-EAAT2, 1:1200 (Abcam, ab41621), mouse
680 anti-beta III Tubulin, 1:800 (Abcam, ab78078), rabbit anti-DNMT1, 1:1000 (Proteintech,
681 24206-1-AP), mouse anti-VINCULIN, 1:1000 (Sigma, V9131), rabbit anti-Myc-Tag, 1:1000
682 (Cell signaling, 2272). Mouse anti-PCNA, 1:7000 (Proteintech, 60097-1-Ig) was incubated at
683 RT for 1.5 h. Blots were stained with secondary antibodies: Donkey anti-Mouse IgG, 1:15000
684 (LICOR, 926-68072), Donkey anti-Rabbit IgG, 1:15000 (LICOR, 926-32213) for 1h at RT.
685 After washing, blots were visualized with Odyssey® CLx Imager and protein quantification
686 was evaluated by Image J (Schindelin et al., 2012).

687 **Immunofluorescence**

688 NPCs were cultured on PDL/Laminin-coated 8-well culture slide (Falcon, 354118) for 24-48
689 hours, fixed with 4% paraformaldehyde (PFA), blocked with 5% NGS in 0.1% PBS-T and
690 stained using primary antibodies against rabbit anti-KMT2A, 1:200 (Cell signaling, 14197),

691 rabbit anti-EAAT2 1:400 (Abcam, ab41621) and mouse beta III Tubulin 1:200 (Abcam,
692 ab78078), overnight. The following day, cells were stained using secondary antibodies: Alexa
693 goat anti-rabbit 647, 1:1000 (ThermoFisher, A-21244) and Alexa goat anti-mouse 555,
694 1:1000 (ThermoFisher, A-32727). Cells were mounted with Fluoromount-G™ Mounting
695 Medium, with DAPI (ThermoFisher, 00-4959-52). Images were captured with Olympus
696 Confocal microscopy (Olympus Fluoroview, FV4000), processed with Fiji and quantified
697 with CellProfiler. Statistical analysis was done using GraphPad Prism v10.4.

698 **Overexpression**

699 NPCs were seeded at a density of 250,000-300,000 cells per well in PDL and Laminin coated
700 6-well plates, 24 hours prior to transfection. The following day, cells were co-transfected with
701 2.5 µg of either the Myc-tagged *Zic4* overexpression plasmid (Origene, MR216949) or the
702 corresponding Myc-alone plasmid backbone (control), along with 7.5 µl of Lipofectamine
703 3000, following the manufacturer's protocol. Cells were harvested 24 hours post-transfection
704 for downstream analysis.

705 **DNA sample isolation and preparation**

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

718 **ONT-DNA sequencing library construction**

719 Library preparation for ONT-DNA sequencing used the ONT Ligation Sequencing kit V14
720 (ONT, SQK-LSK114), using 1 µg sheared genomic HMW DNA in 47 µl nuclease-free water
721 (ThermoFisher, Am9937) as input, followed by library preparing and loading R10.4.1 flow

722 cell (FLO-PRO114M, ONT) on a PromethION 24 sequencer following the manufacturer's
723 instructions, see supplementary methods for details. Nanopore sequencing was performed
724 using the following settings: a Run limit of 96 h, a Minimum read length of 200 bp, and
725 Super-Accurate Basecalling Model with 5mC and 5hmC modified base context. To increase
726 the output, flow cells were washed once after 22-24 h of sequencing using Flow Cell Wash kit
727 XL (ONT, EXP-WAS004-XL), followed by 1-1.5 h waiting time. Then priming and loading
728 library procedures were the same as described previously. Around 70 Gigabases estimated
729 bases with an average read length of 20 kb were achieved per sample.

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

731 We developed a computational tool, *samCRISPR*, to estimate CRISPR/Cas9 KO efficiency
732 using ONT long-read whole-genome sequencing (WGS) data. The name "samCRISPR"
733 references both SAMtools, from which the mpileup format is used in KO efficiency
734 calculations, and CRISPR-Cas9 genome editing technology. SAMtools is used to sort, index,
735 and filter aligned reads, and to generate mpileup output at the target site, which provides base-
736 level resolution of read alignments. This output is parsed by our tool, *samCRISPR*, to identify
737 substitutions, insertions, and deletions based on mpileup letter codes, allowing precise
738 quantification of editing events and KO efficiency. The tool identifies sequencing reads that
739 carry CRISPR-induced events within designated target regions using the mpileup output from
740 SAMtools (Danecek et al., 2021) (**Figure 1F**). To eliminate the chance of counting
741 sequencing errors in the vicinity of target regions as actual CRISPR events, the counting of
742 CRISPR events was restricted to 1bp upstream and 1bp downstream of Cas9 cut sites.
743 CRISPR events were defined as mutations occurring within ±1 base pairs of the predicted cut
744 site, located 3 bp upstream of the PAM sequence. This narrow window was chosen to
745 minimize false positives caused by sequencing or alignment errors near Cas9 cleavage sites.
746 KO efficiency was calculated as the ratio of reads with CRISPR events in the target region to
747 the total number of reads mapped to that region. To validate samCRISPR results, KO
748 efficiencies for selected EM-KOs were manually inspected using the Integrative Genomics
749 Viewer (IGV) (Robinson et al., 2011). *samCRISPR* accepts a BAM file, a reference FASTA,
750 or BED-formatted sgRNA coordinates as input, and supports batch analysis of multiple BAM
751 files and sgRNAs. It requires a Linux environment with SAMtools for mpileup generation and
752 R computing environment for calculating Wilson score confidence intervals (Agresti & Coull,
753 1998).

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

755 We utilized a custom graph genome-based approach, termed PLASMA (Pipeline for Long-
756 read Allele-Specific Methylation Analysis), to integrate both FVB and B6 genomic sequences
757 into a unified non-linear reference coordinate suitable for allele-specific methylation analysis
758 with minimal introduction of reference bias (Davidovich, et al., in preparation). A genome
759 graph, a structure composed of nodes and edges, captures alternative alleles, structural
760 variants, and indels across multiple genomes (Liao et al., 2023; Zhang et al., 2025). Unlike
761 linear references, genome graphs support the simultaneous representation of multiple
762 haplotypes, making them particularly suited for admixed or hybrid genomes.

763 The whole-genome alignment between the FVB and B6J reference genomes was performed
764 using MUMmer4 v4.0 (Marcais et al., 2018) to identify heterozygous variants. Next, Jellyfish
765 (Marcais & Kingsford, 2011) was used to enumerate all 45-mers in each genome, which were
766 subsequently classified into conserved nodes and variant edges. Edges shorter than 3 kb were
767 aligned between genomes to derive CIGAR strings, enabling base-level coordinate
768 conversion. From these, a pseudo-hybrid intermediate genome was constructed by selecting
769 the longer edge between B6 and FVB, thus forming a unified coordinate system
770 encompassing diversity of both strains.

771 Guppy v6.5.7 in MinKNOW software v23.04.5 was used to call canonical and modified bases
772 (5mC, 5hmC) from raw current signal files fast5/pod5. Reads were then aligned to both
773 references using Minimap2 v2.26 (Li, 2018), retaining only high-confidence alignments
774 (MAPQ > 20). Strain-specific SNPs and alignment scores were used to assign reads to B6 or
775 FVB haplotypes using WhatsHap v2.8 (Patterson et al., 2015). Methylation of each CpG was
776 calculated using modkit v0.1.2 (https://github.com/nanoporetech/modkit). Importantly, all
777 DNAm data were projected onto the pseudo-hybrid coordinate system, minimizing reference
778 bias during haplotype-specific methylation analysis and allowing for a systematic comparison
779 of methylation data from distinct reference coordinate systems.

780 **Identification of differentially methylated regions**

781 Methylation levels of each CpG for each sample were extracted from bedMethyl files
782 generated by modkit, which contain per-CpG entries detailing genomic coordinates and read
783 counts. Each file included approximately 21 million CpG sites. These datasets were imported
784 into R to create BSseq objects (Hansen et al., 2012) for downstream DMR analysis. To ensure
785 data reliability, unphased datasets retained only CpGs with ≥3× coverage in at least 80% of
786 samples. For phased data, where coverage was inherently lower, CpGs with ≥1× coverage

787 across all samples were included. DMRs were identified in phased data using the bsseq 1.39.4
788 (Hansen et al., 2012) and unphased data using dmrseq 1.22.1 (Korthauer et al., 2019) R
789 packages. Dmrseq controls the false discovery rate (FDR) whereas bsseq controls the family-
790 wise error rate (FWER). For the phased data smoothing was performed separately for each
791 allele. Strain-specific CpGs were included during smoothing but were removed prior to DMR
792 identification, following previous work (Wulfridge et al., 2019). This strategy enhances
793 statistical power while reducing spurious DMR detection due to strain-specific sequence
794 differences. DMRs were filtered using the following criteria phased analysis: ≥ 3 CpGs per
795 region, > 5% mean methylation difference, and inverse CpG density ≤ 300 (defined as the
796 number of CpGs divided by the region's genomic width), unphased analysis ≥ 5 CpGs, and
797 inverse CpG density ≤ 120.

798 **Annotation of differentially methylated regions**

799 We used annotatePeaks.pl program to annotate DMRs by using UCSC mm10 mouse
800 reference fasta and UCSC mm10 refGene gtf files. That program calls a separate program,
801 assignGenomeAnnotation, to efficiently assign DMRs to possible annotations genome wide.
802 By default, in the case that some annotations overlap, HOMER assigns a priority based on the
803 following: TSS (by default defined from -1kb to +100bp), TTS (by default defined from -100
804 bp to +1kb), CDS Exons, 5' UTR Exons, 3' UTR Exons, introns, and intergenic. Intergenic
805 regions were considered either upstream (TSS) or downstream (TTS) of genes.
806
807 To annotate DMRs in phased data, a custom pipeline was employed: The UCSC mm10 GTF
808 file was converted to BED format using a Bash script. A Python script mapped coordinates
809 between the mm10 genome and the pseudo-hybrid intermediate genome. This allowed
810 accurate functional annotation of DMRs in a unified coordinate space representing both B6
811 and FVB haplotypes.

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

813 Canonical transcripts of protein-coding genes in *Mus musculus* were downloaded from
814 Ensembl using the biomaRt package v2.58.2 (Durinck et al., 2005) and annotated using
815 GenomicFeatures package v1.54.4 (Lawrence et al., 2013) in conjunction with the UCSC
816 genome annotation TxDb.Mmusculus.UCSC.mm10.knownGene v3.10.0. A subset of these
817 genomic features was then selected to include only the canonical protein-coding transcripts.
818 For CpG enrichment, we used the R computing environment to construct a 2×2 contingency

819 table with the number of CpGs for each genomic feature and DMRs. A Fisher's exact test was
820 used to assess the significance of enrichment.

821 **DNA methylation overlap across individual datasets**

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

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

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

837 **RNA isolation and sequencing**

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

850 **RNA sequencing analysis**

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

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

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

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

878 We used NMF on smoothed DNAm levels of promoters of protein-coding genes (Cemgil,
879 2009). CpGs with more than 3x coverage across at least more than 80% of samples were kept
880 in our analysis. Any promoters with no retained CpGs were discarded. The number of latent
881 factors (parameter k) was chosen to be three based on diagnostic plots of NMF and PPI-

882 interactions of EM components. The brunet model in the R package nmf v0.28 was employed
883 in decomposition.

884 **TF motif enrichment analysis**

885 For identification of TF motif enrichment in DMRs and promoter regions, the Analysis for
886 Motif Enrichment (AME) tool within MEME-suite v5.5.8 (https://meme-
887 suite.org/meme/tools/ame) was used (Bailey et al., 2015). The top 30k most differentially
888 methylated regions from the *Dnmt1*-KO were tested against a background that included 30k
889 random regions from the mm10 genome that were matched in size distribution and CpG
890 density and did not include any of the *Dnmt1*-KO DMRs (**Supplementary Table 15**). For
891 promoter regions of shared DEGs we obtained mouse protein-coding promoters for canonical
892 transcripts as described above, filtered them for the shared DEGs and used all other protein-
893 coding genes as a background. In both cases, we tested against the mouse HOCOMOCO v11
894 core database and estimated enrichment using a Fisher's exact test. Odds ratios were
895 calculated using the following formula:

896 $$OR = \frac{\%TP \times (100 - \%FP)}{\%FP \times (100 - \%TP)}$$

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

899 **Inspection of protein-protein interactions using STRING and BioGRID**

900 STRING (Bernas et al., 2017) interactions consisted of three categories: known interactions
901 derived from curated databases and experimental data; predicted interactions based on gene
902 neighborhood, gene fusions, and co-occurrence; and other interactions inferred from text
903 mining, co-expression, and protein homology. We used STRING interactions (confidence
904 score cutoff of > 0.400) between members of the DNAm machinery (Boukas et al., 2019) and
905 the 46 EM-KOs to determine whether the shared impact of EM-KOs on promoter DNAm in
906 **Figure 5A** is associated with these interactions. To identify if TFs with motif enrichments in
907 the *Dnmt1*-KO DMRs or the shared DEGs between *Kmt2a*-KOs and *Dnmt1*-KOs were also
908 known interactors of these proteins, we downloaded all known direct human protein-protein
909 interactions from BioGRID (Oughtred et al., 2021). We filtered the list for known interactors
910 of KMT2A and DNMT1 and compared them with the lists of TFs in R.

911 **Enrichment analysis and visualization**

912 To test whether the number of overlapping genes between different RNA sequencing datasets
913 was higher than expected, Fisher's exact tests were performed using the function fisher.test()
914 in R. Shared detected genes were considered background. To test if TF targets or imprinted
915 genes were enriched in DEGs lists, Fisher's exact tests were performed using the function
916 fisher.test() in R, using detected genes as background. TF target genes were obtained from
917 https://tfbsdb.systemsbiology.net/ (Plaisier et al., 2016) and the TFlink (https://tflink.net/)
918 (Liska et al., 2022) databases. A list of mouse imprinted genes was obtained from the
919 Imprinted Gene Database (https://www.geneimprint.com/site/genes-by-
920 species.Mus+musculus), filtering away genes with "unknown" or "not imprinted" status. To
921 test whether known BioGRID interactors for KMT2A and DNMT1 overlapped significantly,
922 we filtered the human BioGRID list for interactors with either KMT2A or DNMT1. Fisher's
923 exact test was performed using the fisher.test() function in R, using all interacting partners in
924 the human BioGRID dataset, that are expressed in the two RNA-seq datasets, as universe.

925 **Gene ontology and gene set enrichment analysis**

926 To estimate enrichment in certain gene ontology (GO) terms in RNA sequencing datasets, the
927 Webgestalt 2024 (Elizarraras et al., 2024) webtool was used, using mouse protein-coding
928 genes as background. The GO term "Biological Process" was evaluated and FDR< 0.05 used
929 to estimate significance. The top 10-15 terms were reported for each analysis. For enrichment
930 of certain GO terms and KEGGs pathways in genes associated with latent factors in the NMF
931 analysis, the EnrichR (Kuleshov et al., 2016) webtool was used. For GSEA of RNA-seq data
932 we ranked the gene lists from *Kmt2a*-KO and *Dnmt1*-KO NPCs using the signed log10
933 adjusted p-value from the DESeq2 analysis, where the sign was determined by the direction of
934 change. Preranked lists were loaded to the GSEA software (v.4.3.2, BROAD Institute) and
935 enrichment evaluated against the Molecular Signature Database (MSigDB) 2024 Hallmark
936 mouse gene set. Significance was assessed using 1,000 gene set permutations, and results with
937 FDR q < 0.05 were considered significant.

938 $$OR = \frac{\%TP \times (100 - \%FP)}{\%FP \times (100 - \%TP)}$$

939 **Data visualization**

940 For scatter plots and volcano plots, the R package ggplot2 was used (Wickham, 2016). For
941 Venn diagrams, the R package VennDiagram was used (Chen & Boutros, 2011). The
942 heatmaps were generated through the *pheatmap()* function of the R package pheatmap. For

individual gene expression values from RNA sequencing data, and for quantification results from qPCR, WB and IF experiments, Graphpad Prism (v10.4) was used. Schematic figures were generated using BioRender, and figure assembled into panels using Affinity Designer.

Data availability

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

Code availability

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

Acknowledgements: We would like to thank Elvar Örn Jónsson and Þorgils Árni Hjálmarsson for their dedication and valuable efforts in troubleshooting data storage and analysis issues within the high-performance computing (HPC) cluster at the University of Iceland. 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; J.O., K.M., K.W. S.P., and T.P. performed experiments. K.O., K.M., K.D.H., J.O., S.P, and A.D. analyzed the data. H.T.B., K.D.H., K.M., J.O. and K.O. wrote manuscript. K.A. K.W., S.P., T.P., and A.D. edited manuscript.

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

973 **Tables**

974 **List of supplementary tables**

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

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

979 **Supplementary Table 3.** Genomic region enrichment analysis of the *Dnmt1*-KO DMRs.

980 **Supplementary Table 4.** TF motif enrichment in the top 30K *Dnmt1*-KO DMRs compared to
981 a length and CpG density matched background list.

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

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

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

985 **Supplementary Table 8.** TF motif enrichment of promoters of shared DEGs between *Dnmt1*-
986 KO and *Kmt2a*-KO.

987 **Supplementary Table 9.** Differential expression analysis results after 2 days of
988 differentiation.

989 **Supplementary Table 10.** Differential expression analysis results after 6 days of
990 differentiation.

991 **Supplementary Table 11.** Genome-wide DMRs between time points of neurodevelopmental
992 model.

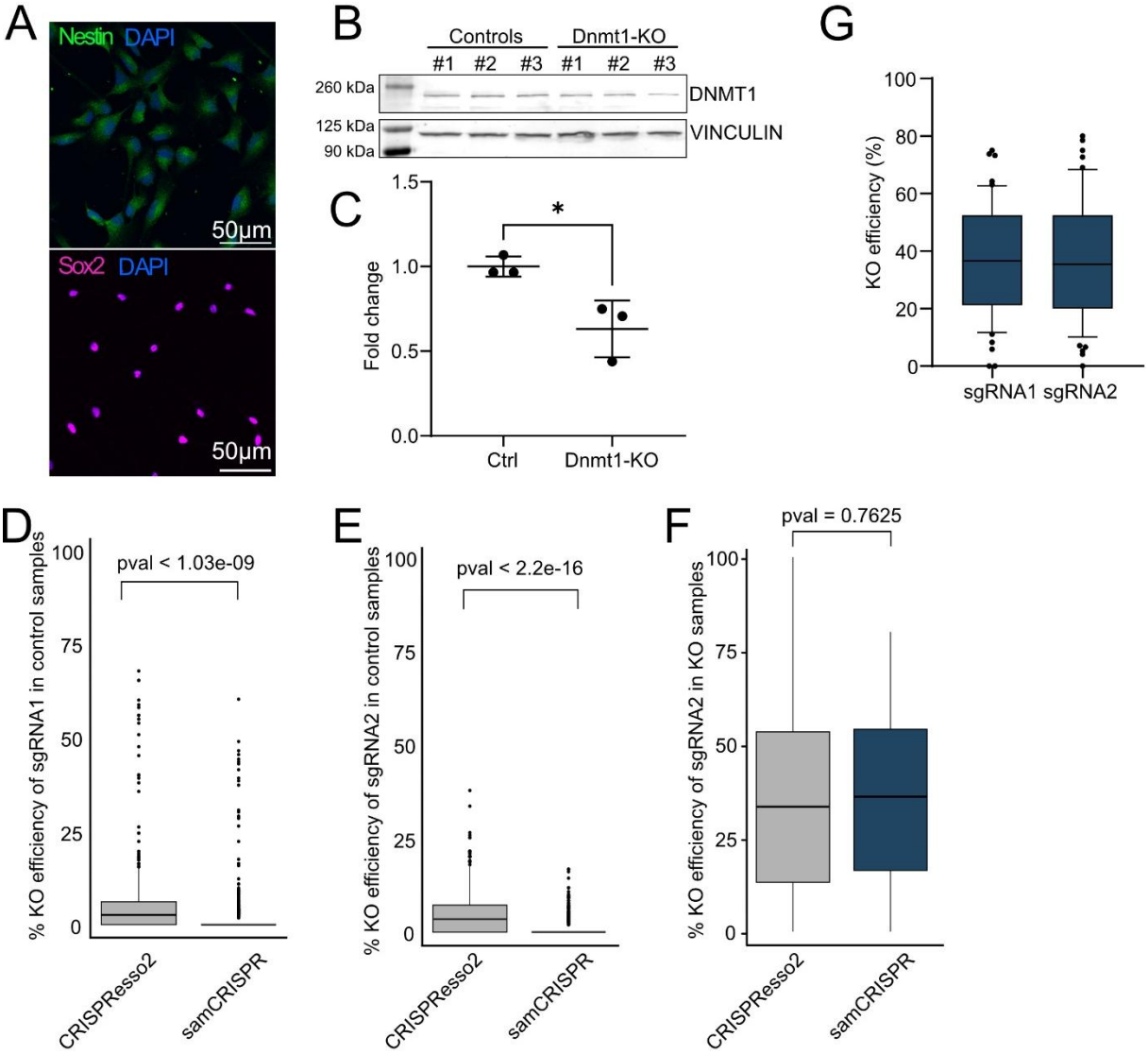
993 **Supplementary Table 12.** Overlap between *Dnmt1*-KO DMRs and Day6 DMRs.

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

995 **Supplementary Table 14**. Primers used during the current study.

996 **Supplementary Table 15.** A list of 30k DMRs that were used as background for motif
997 enrichment analysis.

998 **Supplemental figures**

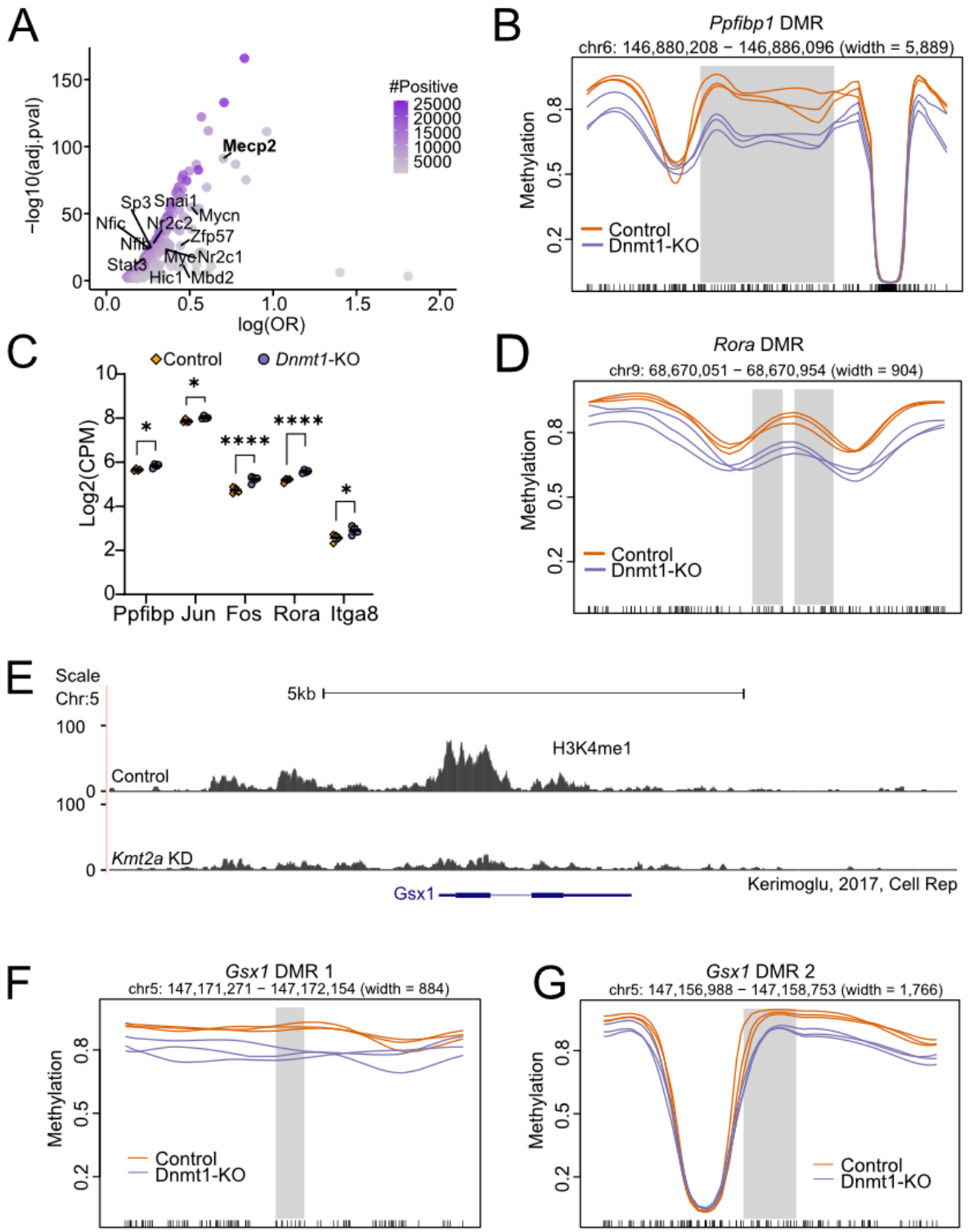


999

1000 **Supplementary Figure 1: CRISPR-Cas9 KO system evaluation**. (A) IF staining of isolated
1001 mouse NPCs, staining for Nestin (green, upper) and Sox2 (magenta, lower) and DAPI (blue).
1002 (B-C) WB and quantification showing *Dnmt1*-KO and scrambled controls in triplicates.
1003 Results in (C) show DNMT1 (upper in B) normalized to VINCULIN (lower in B). A t-test
1004 was used for significance. (D) KO efficiency of sgRNA 1 in control samples across methods
1005 (two-tailed t-test). (E) KO efficiency of sgRNA 2 in control samples across methods (two
1006 tailed t-test). (F) KO efficiency of sgRNA 2 in EM-KOs across methods (two-tailed t-test).
1007 (G) KO efficiency of sgRNA 1 and 2 in EM-KOs using *samCRISPR*.

1008

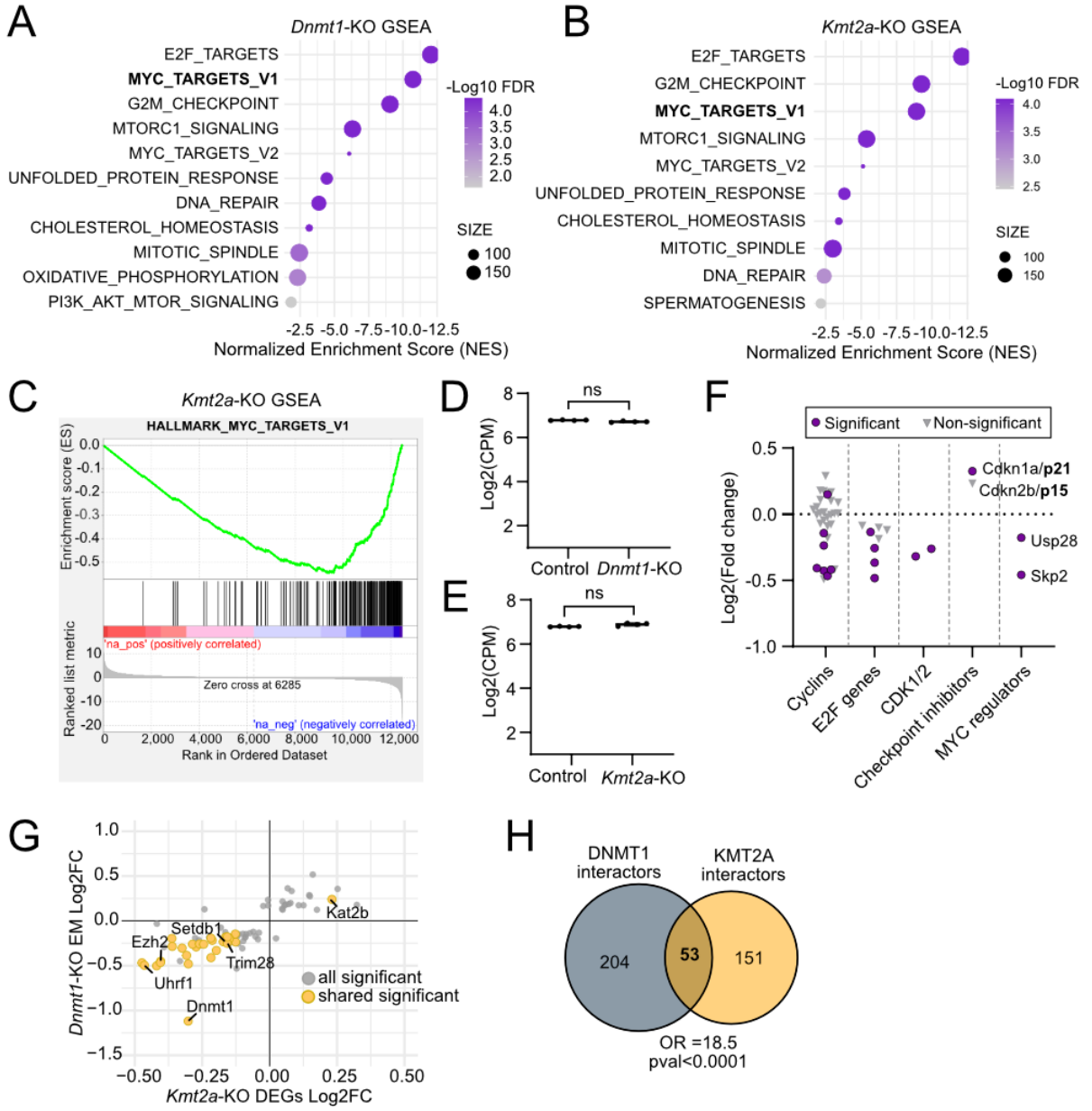
1009



1010

1011 **Supplementary Figure 2: *Dnmt1*-KO and *Kmt2a*-KO DMRs.** (A) A scatter plot showing
1012 TFs motif enrichment in *Dnmt1*-KO associated DMRs. Known binding partners of DNMT1
1013 from BioGRID are labelled. (B,D) Examples of genome-wide significant DMRs in *Dnmt1*-
1014 KO mNPCs (blue) compared to controls (orange), located near the neuronal-related gene
1015 *Ppfibp1* (B), and a known DNMT1 target gene *Rora* (D) (Tao et al., 2022). (C) Dot plot
1016 showing normalized expression 5 examples of *Dnmt1*-KO DMR-associated DEGs. (E)
1017 Publicly available ChIP-seq data (Kerimoglu et al., 2017) showing a loss of H3K4me1 peaks
1018 at the *Gsx1* exon in *Kmt2a* knockdown cells compared to control. (F,G) Smoothed

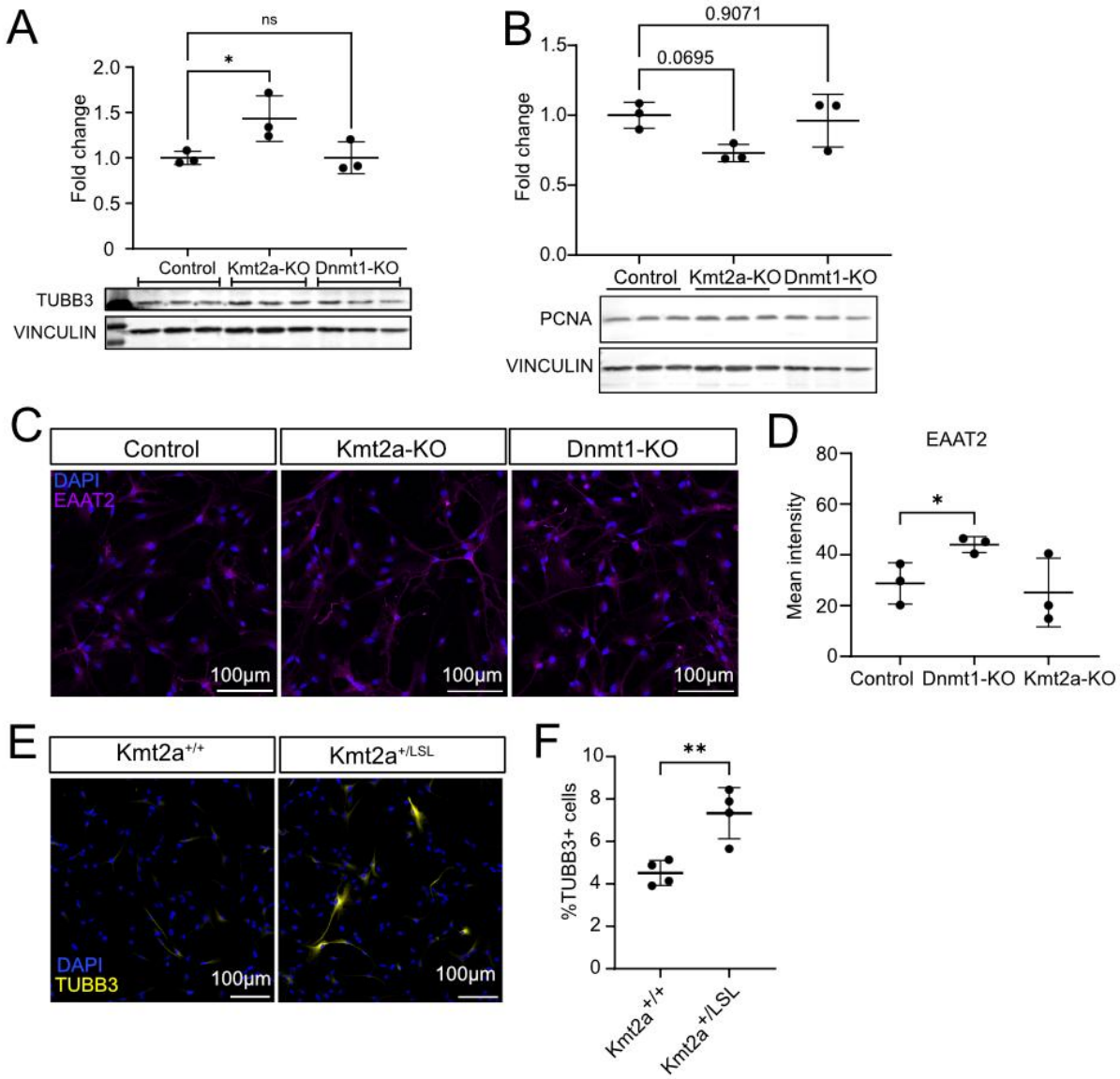
methylation plots showing two *Dnmt1*-KO DMRs located in an intergenic region near *Gsx1* gene.



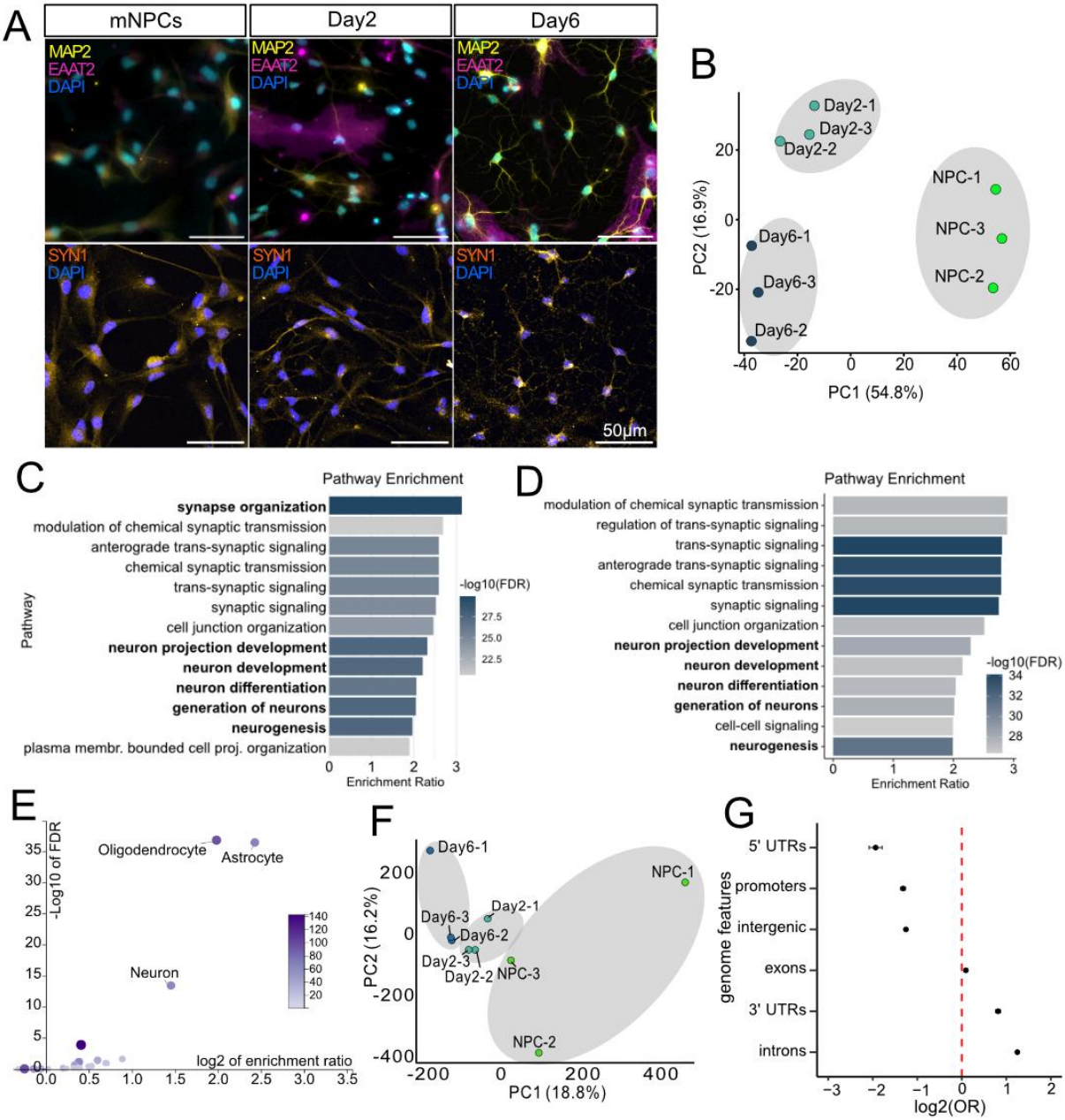
Supplementary Figure 3: RNA abnormalities in *Kmt2a*-KO and *Dnmt1*-KO DEGs.

(A-B) GSEA analysis of (A) *Dnmt1*-KO and (B) *Kmt2a*-KO genes, tested against the MSigDB HALLMARK gene set. The x-axis shows the normalized enrichment score; color denotes FDR adjusted q-values and size indicates the number of genes tested in each category. (C) GSEA plots showing the MYC_targets_V1 Hallmark category for *Kmt2a*-KO genes. (D-E) Log2 normalized *Myc* expression values from the (D) *Dnmt1*-KO and (E) *Kmt2a*-KO RNA-seq data. (F) RNA-seq results showing gene expression changes of MYC targets in *Kmt2a*-KO NPCs compared with controls. Y-axis shows Log2fold changes of genes belonging to the categories on the X-axis. Purple dots show significant genes; gray triangles show non-significant genes. (G) Scatter plot showing log2fold changes of differentially expressed EM genes in *Dnmt1*-KO on y-axis and *Kmt2a*-KO on x-axis. Grey dots show significant EM genes in either one of the datasets, yellow dots show EM genes significant in

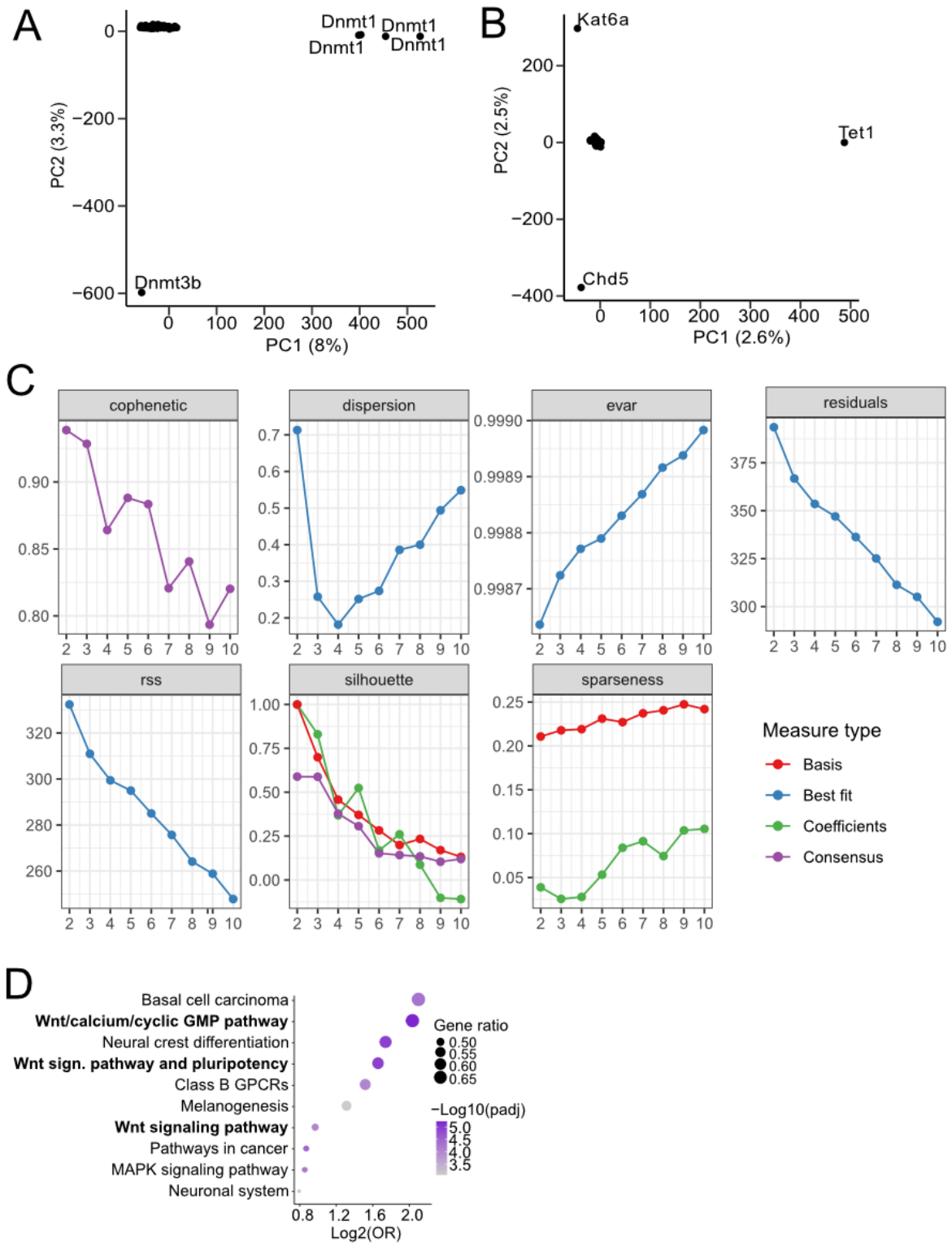
both datasets. (H) Venn diagram showing the overlap of expressed BioGRID interactors between KMT2A and DNMT1, Fisher's exact test OR=18.5, $p < 0.0001$.



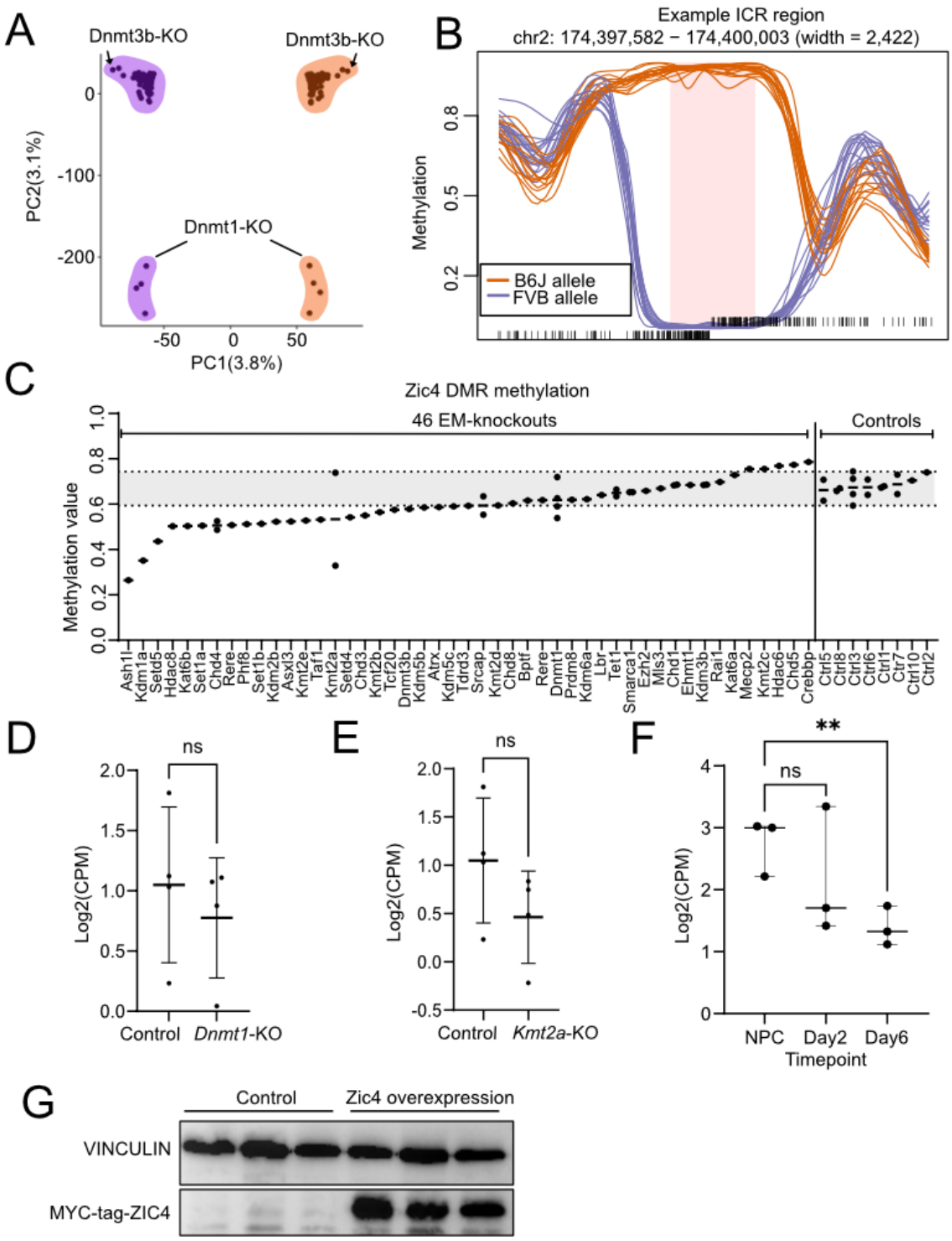
Supplementary Figure 4: Premature differentiation after *Dnmt1*- and *Kmt2a*-KO
(A) WB analysis of TUBB3 protein after knocking out *Kmt2a* (padj=0.0474) and *Dnmt1* (padj>0.05, ns) with CRISPR/Cas9. A one-way ANOVA with Dunnett´s multiple comparisons test was used to calculate significance. (B) WB analysis of PCNA expressions in *Kmt2a*-KO (ns, $p = 0.0695$) and *Dnmt1*-KO (ns, $p = 0.9071$) compared to control mNPC. A one-way ANOVA with Dunnett´s correction was used. (C) Representative IF images of control, *Kmt2a*-KO and *Dnmt1*-KO mNPCs stained for DAPI (blue) and EAAT2 (magenta). Scale bars: 100 µm. (D) 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.(E) Representative IF images of B6J-*Kmt2a*$^{+/+}$ and B6J-*Kmt2a*$^{+/LSL}$ P0 mNPCs, stained for DAPI (blue) and TUBB3 (yellow). (F) 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 5: DNA methylation changes in neurodevelopment. (A) IF 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 6. 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 7. A potential marker gene for EM disruption. (A) A PCA plot showing *Dnmt1*-KO and *Dnmt3b*-KO are two outliers (purple is the B6J haplotype, orange is the FVB haplotype). (B) Smoothed methylation plots showing a haplotype-specific ICR located on chromosome 2. (C) A 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 the minimum and maximum values of the controls. (D-E) Normalized counts showing *Zic4* expression in *Kmt2a*-KO (not significant, p = 0.1846, D) and *Dnmt1*-KO (not significant, p = 0.5, E), compared to controls. (F) Normalized counts showing *Zic4* expression in neurodevelopment, where it is significantly downregulated at Day6 compared with the NPC

1084 state ($p = 0.00197$). (G) A western-blot showing overexpression of Myc-tagged Zic4 and
1085 Vinculin as a loading control.

References

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

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

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

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

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

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

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

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

Bretones, G., Delgado, M. D., & León, J. (2015). Myc and cell cycle control. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1849*(5), 506-516. https://doi.org/https://doi.org/10.1016/j.bbagrm.2014.03.013

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

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

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

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

Ciceri, G., Baggiolini, A., Cho, H. S., Kshirsagar, M., Benito-Kwiecinski, S., Walsh, R. M., Aromolaran, K. A., Gonzalez-Hernandez, A. J., Munguba, H., Koo, S. Y., Xu, N., Sevilla, K. J., Goldstein, P. A., Levitz, J., Leslie, C. S., Koche, R. P., & Studer, L. (2024). An epigenetic barrier sets the timing of human neuronal maturation. *Nature*, *626*(8000), 881-890. https://doi.org/10.1038/s41586-023-06984-8

1136 Ciptasari, U., & van Bokhoven, H. (2020). The phenomenal epigenome in neurodevelopmental
1137 disorders. *Hum Mol Genet*, *29*(R1), R42-R50. https://doi.org/10.1093/hmg/ddaa175
1138 Coltogirone, R. A., Sherfinski, E. I., Dobler, Z. A., Peterson, S. N., Andlinger, A. R., Fadel, L. C., Patrick,
1139 R. L., & Bergeron, S. A. (2023). Gsx2, but not Gsx1, is necessary for early forebrain patterning
1140 and long-term survival in zebrafish. *Dev Dyn*, *252*(3), 377-399.
1141 https://doi.org/10.1002/dvdy.542
1142 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T.,
1143 McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools.
1144 *Gigascience*, *10*(2). https://doi.org/10.1093/gigascience/giab008
1145 Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev*, *25*(10),
1146 1010-1022. https://doi.org/10.1101/gad.2037511
1147 Do, C., Lang, C. F., Lin, J., Darbary, H., Krupska, I., Gaba, A., Petukhova, L., Vonsattel, J. P., Gallagher,
1148 M. P., Goland, R. S., Clynes, R. A., Dwork, A., Kral, J. G., Monk, C., Christiano, A. M., & Tycko,
1149 B. (2016). Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA
1150 Methylation. *Am J Hum Genet*, *98*(5), 934-955. https://doi.org/10.1016/j.ajhg.2016.03.027
1151 Dong, P., Maddali, M. V., Srimani, J. K., Thélot, F., Nevins, J. R., Mathey-Prevot, B., & You, L. (2014).
1152 Division of labour between Myc and G1 cyclins in cell cycle commitment and pace control.
1153 *Nature Communications*, *5*(1), 4750. https://doi.org/10.1038/ncomms5750
1154 Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart
1155 and Bioconductor: a powerful link between biological databases and microarray data
1156 analysis. *Bioinformatics*, *21*(16), 3439-3440. https://doi.org/10.1093/bioinformatics/bti525
1157 Elizarraras, J. M., Liao, Y., Shi, Z., Zhu, Q., Pico, A. R., & Zhang, B. (2024). WebGestalt 2024: faster
1158 gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Res*,
1159 *52*(W1), W415-W421. https://doi.org/10.1093/nar/gkae456
1160 Fahrner, J. A., Lin, W. Y., Riddle, R. C., Boukas, L., DeLeon, V. B., Chopra, S., Lad, S. E., Luperchio, T. R.,
1161 Hansen, K. D., & Bjornsson, H. T. (2019). Precocious chondrocyte differentiation disrupts
1162 skeletal growth in Kabuki syndrome mice. *JCI Insight*, *4*(20).
1163 https://doi.org/10.1172/jci.insight.129380
1164 Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jönsson, M. E., Männe, C., Turelli, P., Jern, P.,
1165 Cammas, F., Trono, D., & Jakobsson, J. (2015). TRIM28 represses transcription of endogenous
1166 retroviruses in neural progenitor cells. *Cell Rep*, *10*(1), 20-28.
1167 https://doi.org/10.1016/j.celrep.2014.12.004
1168 Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., & Liu, H. K. (2013). The
1169 chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription
1170 factors. *Cell Stem Cell*, *13*(1), 62-72. https://doi.org/10.1016/j.stem.2013.05.002
1171 Foroutan, A., Haghshenas, S., Bhai, P., Levy, M. A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi, A.,
1172 Pedace, L., Miele, E., Genevieve, D., Heide, S., Alders, M., Zampino, G., Merla, G., Fradin, M.,
1173 Bieth, E., Bonneau, D., Dieterich, K.,…Lebre, A. S. (2022). Clinical Utility of a Unique Genome-
1174 Wide DNA Methylation Signature for KMT2A-Related Syndrome. *Int J Mol Sci*, *23*(3).
1175 https://doi.org/10.3390/ijms23031815
1176 Friedrich, G., & Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to
1177 identify and mutate developmental genes in mice. *Genes Dev*, *5*(9), 1513-1523.
1178 https://doi.org/10.1101/gad.5.9.1513
1179 Fu, M. P., Merrill, S. M., Sharma, M., Gibson, W. T., Turvey, S. E., & Kobor, M. S. (2023). Rare diseases
1180 of epigenetic origin: Challenges and opportunities. *Front Genet*, *14*, 1113086.
1181 https://doi.org/10.3389/fgene.2023.1113086
1182 García-Gutiérrez, L., Bretones, G., Molina, E., Arechaga, I., Symonds, C., Acosta, J. C., Blanco, R.,
1183 Fernández, A., Alonso, L., Sicinski, P., Barbacid, M., Santamaría, D., & León, J. (2019). Myc
1184 stimulates cell cycle progression through the activation of Cdk1 and phosphorylation of p27.
1185 *Scientific Reports*, *9*(1), 18693. https://doi.org/10.1038/s41598-019-54917-1

1186 Garel, S., Marin, F., Grosschedl, R., & Charnay, P. (1999). Ebf1 controls early cell differentiation in the
1187 embryonic striatum. *Development*, *126*(23), 5285-5294.
1188 https://doi.org/10.1242/dev.126.23.5285
1189 Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed,
1190 T. P., Blewitt, M. E., & Ritchie, M. E. (2019). Using long-read sequencing to detect imprinted
1191 DNA methylation. *Nucleic Acids Research*, *47*(8), e46. https://doi.org/10.1093/nar/gkz107
1192 Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W. B., & Millen, K. J. (2004). Heterozygous
1193 deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat*
1194 *Genet*, *36*(10), 1053-1055. https://doi.org/10.1038/ng1420
1195 Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite
1196 sequencing reads to differentially methylated regions. *Genome Biol*, *13*(10), R83.
1197 https://doi.org/10.1186/gb-2012-13-10-r83
1198 Irion, S., Luche, H., Gadue, P., Fehling, H. J., Kennedy, M., & Keller, G. (2007). Identification and
1199 targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*, *25*(12), 1477-
1200 1482. https://doi.org/10.1038/nbt1362
1201 Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of
1202 nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 239.
1203 https://doi.org/10.1186/s13059-016-1103-0
1204 Jonsson, M. E., Ludvik Brattas, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., Verschuere, S.,
1205 Madsen, S., Hansson, J., Larsson, J., Mansson, R., Meissner, A., & Jakobsson, J. (2019).
1206 Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human
1207 neural progenitors. *Nat Commun*, *10*(1), 3182. https://doi.org/10.1038/s41467-019-11150-8
1208 Kalkat, M., Resetca, D., Lourenco, C., Chan, P.-K., Wei, Y., Shiah, Y.-J., Vitkin, N., Tong, Y.,
1209 Sunnerhagen, M., Done, S. J., Boutros, P. C., Raught, B., & Penn, L. Z. (2018). MYC Protein
1210 Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive
1211 Tumorigenesis. *Molecular Cell*, *72*(5), 836-848.e837.
1212 https://doi.org/https://doi.org/10.1016/j.molcel.2018.09.031
1213 Kato, M., Takemoto, K., & Shinkai, Y. (2018). A somatic role for the histone methyltransferase Setdb1
1214 in endogenous retrovirus silencing. *Nature Communications*, *9*(1), 1683.
1215 https://doi.org/10.1038/s41467-018-04132-9
1216 Kerimoglu, C., Sakib, M. S., Jain, G., Benito, E., Burkhardt, S., Capece, V., Kaurani, L., Halder, R., Agis-
1217 Balboa, R. C., Stilling, R., Urbanke, H., Kranz, A., Stewart, A. F., & Fischer, A. (2017). KMT2A
1218 and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Rep*,
1219 *20*(3), 538-548. https://doi.org/10.1016/j.celrep.2017.06.072
1220 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment
1221 and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, *37*(8), 907-915.
1222 https://doi.org/10.1038/s41587-019-0201-4
1223 Kim, J., Lu, X., & Stubbs, L. (1999). Zim1, a Maternally Expressed Mouse Kruppel-Type Zinc-Finger
1224 Gene Located in Proximal Chromosome 7. *Human Molecular Genetics*, *8*(5), 847-854.
1225 https://doi.org/10.1093/hmg/8.5.847
1226 Kim, S., Kim, H., & Um, J. W. (2018). Synapse development organized by neuronal activity-regulated
1227 immediate-early genes. *Exp Mol Med*, *50*(4), 1-7. https://doi.org/10.1038/s12276-018-0025-
1228 1
1229 Kimura, H., & Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for
1230 maintenance DNA methyltransferase, Dnmt1. *J Biol Chem*, *278*(7), 4806-4812.
1231 https://doi.org/10.1074/jbc.M209923200
1232 Korthauer, K., Chakraborty, S., Benjamini, Y., & Irizarry, R. A. (2019). Detection and accurate false
1233 discovery rate control of differentially methylated regions from whole genome bisulfite
1234 sequencing. *Biostatistics*, *20*(3), 367-383. https://doi.org/10.1093/biostatistics/kxy007
1235 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S.,
1236 Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen,
1237 G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web

1238 server 2016 update. *Nucleic Acids Res*, *44*(W1), W90-97.
1239 https://doi.org/10.1093/nar/gkw377
1240 Kutner, R. H., Zhang, X. Y., & Reiser, J. (2009). Production, concentration and titration of pseudotyped
1241 HIV-1-based lentiviral vectors. *Nat Protoc*, *4*(4), 495-505.
1242 https://doi.org/10.1038/nprot.2009.22
1243 La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth,
1244 C., Khven, I., Lederer, A. R., Dratva, L. M., Johnsson, A., Nilsson, M., Lonnerberg, P., &
1245 Linnarsson, S. (2021). Molecular architecture of the developing mouse brain. *Nature*,
1246 *596*(7870), 92-96. https://doi.org/10.1038/s41586-021-03775-x
1247 Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019).
1248 CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids*
1249 *Res*, *47*(W1), W171-W174. https://doi.org/10.1093/nar/gkz365
1250 Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey,
1251 V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol*, *9*(8),
1252 e1003118. https://doi.org/10.1371/journal.pcbi.1003118
1253 Lei, X., & Jiao, J. (2018). UTX Affects Neural Stem Cell Proliferation and Differentiation through PTEN
1254 Signaling. *Stem Cell Reports*, *10*(4), 1193-1207. https://doi.org/10.1016/j.stemcr.2018.02.008
1255 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-
1256 3100. https://doi.org/10.1093/bioinformatics/bty191
1257 Liang, H., Xiao, G., Yin, H., Hippenmeyer, S., Horowitz, J. M., & Ghashghaei, H. T. (2013). Neural
1258 development is dependent on the function of specificity protein 2 in cell cycle progression.
1259 *Development*, *140*(3), 552-561. https://doi.org/10.1242/dev.085621
1260 Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J.,
1261 Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X.,
1262 Fischer, C., Fulton, R. S.,…Paten, B. (2023). A draft human pangenome reference. *Nature*,
1263 *617*(7960), 312-324. https://doi.org/10.1038/s41586-023-05896-x
1264 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for
1265 assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930.
1266 https://doi.org/10.1093/bioinformatics/btt656
1267 Liska, O., Bohar, B., Hidas, A., Korcsmaros, T., Papp, B., Fazekas, D., & Ari, E. (2022). TFLink: an
1268 integrated gateway to access transcription factor-target gene interactions for multiple
1269 species. *Database (Oxford)*, *2022*. https://doi.org/10.1093/database/baac083
1270 Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., & Wong, J. (2013). UHRF1 targets DNMT1
1271 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated
1272 H3K9. *Nature Communications*, *4*(1), 1563. https://doi.org/10.1038/ncomms2562
1273 Lo, H. L., & Yee, J. K. (2007). Production of vesicular stomatitis virus G glycoprotein (VSV-G)
1274 pseudotyped retroviral vectors. *Curr Protoc Hum Genet*, *Chapter 12*, Unit 12 17.
1275 https://doi.org/10.1002/0471142905.hg1207s52
1276 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for
1277 RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-
1278 0550-8
1279 Luperchio, T. R., Boukas, L., Zhang, L., Pilarowski, G., Jiang, J., Kalinousky, A., Hansen, K. D., &
1280 Bjornsson, H. T. (2021). Leveraging the Mendelian disorders of the epigenetic machinery to
1281 systematically map functional epigenetic variation. *Elife*, *10*.
1282 https://doi.org/10.7554/eLife.65884
1283 Mahmood, T., & Yang, P. C. (2012). Western blot: technique, theory, and trouble shooting. *N Am J*
1284 *Med Sci*, *4*(9), 429-434. https://doi.org/10.4103/1947-2714.100998
1285 Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4:
1286 A fast and versatile genome alignment system. *PLoS Comput Biol*, *14*(1), e1005944.
1287 https://doi.org/10.1371/journal.pcbi.1005944

1288 Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of
1289 occurrences of k-mers. *Bioinformatics*, *27*(6), 764-770.
1290 https://doi.org/10.1093/bioinformatics/btr011
1291 Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V. S.,
1292 Schramm, M., Becher, I., Steidel, M., Noh, K. M., Bergamini, G., Beck, M., Bantscheff, M., &
1293 Savitski, M. M. (2018). Systematic analysis of protein turnover in primary cells. *Nat Commun*,
1294 *9*(1), 689. https://doi.org/10.1038/s41467-018-03106-1
1295 Noguchi, H., Kimura, A., Murao, N., Namihira, M., & Nakashima, K. (2016). Prenatal deletion of DNA
1296 methyltransferase 1 in neural stem cells impairs neurogenesis and causes anxiety-like
1297 behavior in adulthood. *Neurogenesis (Austin)*, *3*(1), e1232679.
1298 https://doi.org/10.1080/23262133.2016.1232679
1299 Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas,
1300 N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers,
1301 M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein,
1302 genetic, and chemical interactions. *Protein Sci*, *30*(1), 187-200.
1303 https://doi.org/10.1002/pro.3978
1304 Park, D. H., Hong, S. J., Salinas, R. D., Liu, S. J., Sun, S. W., Sgualdino, J., Testa, G., Matzuk, M. M.,
1305 Iwamori, N., & Lim, D. A. (2014). Activation of neuronal gene expression by the JMJD3
1306 demethylase is required for postnatal and adult brain neurogenesis. *Cell Rep*, *8*(5), 1290-
1307 1299. https://doi.org/10.1016/j.celrep.2014.07.060
1308 Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G. W., & Schonhuth, A. (2015).
1309 WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J*
1310 *Comput Biol*, *22*(6), 498-509. https://doi.org/10.1089/cmb.2014.0157
1311 Pei, Z., Wang, B., Chen, G., Nagao, M., Nakafuku, M., & Campbell, K. (2011). Homeobox genes Gsx1
1312 and Gsx2 differentially regulate telencephalic progenitor maturation. *Proc Natl Acad Sci U S*
1313 *A*, *108*(4), 1675-1680. https://doi.org/10.1073/pnas.1008824108
1314 Pereira, J. D., Sansom, S. N., Smith, J., Dobenecker, M. W., Tarakhovsky, A., & Livesey, F. J. (2010).
1315 Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal
1316 and differentiation in the cerebral cortex. *Proc Natl Acad Sci U S A*, *107*(36), 15957-15962.
1317 https://doi.org/10.1073/pnas.1002530107
1318 Persengiev, S. P., Kondova, II, & Kilpatrick, D. L. (1999). E2F4 actively promotes the initiation and
1319 maintenance of nerve growth factor-induced cell differentiation. *Mol Cell Biol*, *19*(9), 6048-
1320 6056. https://doi.org/10.1128/MCB.19.9.6048
1321 Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016).
1322 Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*, *34*(7), 695-
1323 697. https://doi.org/10.1038/nbt.3583
1324 Plaisier, C. L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C. M., Ding, Y., Reiss, D. J.,
1325 Paddison, P. J., & Baliga, N. S. (2016). Causal Mechanistic Regulatory Network for
1326 Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Syst*, *3*(2), 172-186.
1327 https://doi.org/10.1016/j.cels.2016.06.006
1328 Popov, N., Wanzel, M., Madiredjo, M., Zhang, D., Beijersbergen, R., Bernards, R., Moll, R., Elledge, S.
1329 J., & Eilers, M. (2007). The ubiquitin-specific protease USP28 is required for MYC stability.
1330 *Nature Cell Biology*, *9*(7), 765-774. https://doi.org/10.1038/ncb1601
1331 Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M.,
1332 Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen,
1333 T. S., Weissman, J. S., & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by
1334 direct guide RNA capture and targeted sequencing. *Nat Biotechnol*, *38*(8), 954-961.
1335 https://doi.org/10.1038/s41587-020-0470-y
1336 Reynisdottir, T., Anderson, K. J., Möller, K., Pétursson, S., Brinn, A., Franklin, K. P., Ouyang, J.,
1337 Snorradottir, A. O., Lutz, C. M., Zuberi, A. R., DeLeon, V. B., & Bjornsson, H. T. (2025). In utero
1338 rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome. *JCI*
1339 *Insight*, *10*(20). https://doi.org/10.1172/jci.insight.187039

1340 Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P.
1341 (2011). Integrative genomics viewer. *Nat Biotechnol*, *29*(1), 24-26.
1342 https://doi.org/10.1038/nbt.1754
1343 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for
1344 differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-
1345 140. https://doi.org/10.1093/bioinformatics/btp616
1346 Rybachuk, O., Kopach, O., Pivneva, T., & Kyryk, V. (2019). Isolation of Neural Stem Cells from the
1347 embryonic mouse hippocampus for in vitro growth or engraftment into a host tissue. *Bio*
1348 *Protoc*, *9*(4), e3165. https://doi.org/10.21769/BioProtoc.3165
1349 Samara, A., Spildrejorde, M., Sharma, A., Falck, M., Leithaug, M., Modafferi, S., Bjornstad, P. M.,
1350 Acharya, G., Gervin, K., Lyle, R., & Eskeland, R. (2022). A multi-omics approach to visualize
1351 early neuronal differentiation from hESCs in 4D. *iScience*, *25*(11), 105279.
1352 https://doi.org/10.1016/j.isci.2022.105279
1353 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden,
1354 C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak,
1355 P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature*
1356 *Methods*, *9*(7), 676-682. https://doi.org/10.1038/nmeth.2019
1357 Schoof, M., Launspach, M., Holdhof, D., Nguyen, L., Engel, V., Filser, S., Peters, F., Immenschuh, J.,
1358 Hellwig, M., Niesen, J., Mall, V., Ertl-Wagner, B., Hagel, C., Spohn, M., Lutz, B., Sedlacik, J.,
1359 Indenbirken, D., Merk, D. J., & Schuller, U. (2019). The transcriptional coactivator and histone
1360 acetyltransferase CBP regulates neural precursor cell development and migration. *Acta*
1361 *Neuropathol Commun*, *7*(1), 199. https://doi.org/10.1186/s40478-019-0849-5
1362 Sheu, J. R., Hsieh, C. Y., Jayakumar, T., Lin, G. Y., Lee, H. N., Huang, S. W., & Yang, C. H. (2019). HDAC6
1363 dysfunction contributes to impaired maturation of adult neurogenesis in vivo: vital role on
1364 functional recovery after ischemic stroke. *J Biomed Sci*, *26*(1), 27.
1365 https://doi.org/10.1186/s12929-019-0521-1
1366 Sun, Z., Xu, X., He, J., Murray, A., Sun, M. A., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., Li, L., Zhu,
1367 J., Chen, J., Morozov, A., Pickrell, A. M., Theus, M. H., & Xie, H. (2019). EGR1 recruits TET1 to
1368 shape the brain methylome during development and upon neuronal activity. *Nat Commun*,
1369 *10*(1), 3892. https://doi.org/10.1038/s41467-019-11905-3
1370 Syal, C., Seegobin, M., Sarma, S. N., Gouveia, A., Hsu, K., Niibori, Y., He, L., Wondisford, F. E.,
1371 Frankland, P. W., & Wang, J. (2018). Ectopic expression of aPKC-mediated phosphorylation in
1372 p300 modulates hippocampal neurogenesis, CREB binding and fear memory differently with
1373 age. *Sci Rep*, *8*(1), 13489. https://doi.org/10.1038/s41598-018-31657-2
1374 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
1375 a negative regulator of osteogenesis. *Biol Open*, *11*(3). https://doi.org/10.1242/bio.058534
1376 Tycko, B. (2010). Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet*, *19*(R2), R210-
1377 220. https://doi.org/10.1093/hmg/ddq376
1378 Veyrac, A., Besnard, A., Caboche, J., Davis, S., & Laroche, S. (2014). The transcription factor
1379 Zif268/Egr1, brain plasticity, and memory. *Prog Mol Biol Transl Sci*, *122*, 89-129.
1380 https://doi.org/10.1016/B978-0-12-420170-5.00004-0
1381 Wang, J., Weaver, I. C., Gauthier-Fisher, A., Wang, H., He, L., Yeomans, J., Wondisford, F., Kaplan, D.
1382 R., & Miller, F. D. (2010). CBP histone acetyltransferase activity regulates embryonic neural
1383 differentiation in the normal and Rubinstein-Taybi syndrome brain. *Dev Cell*, *18*(1), 114-125.
1384 https://doi.org/10.1016/j.devcel.2009.10.023
1385 Wei, Z., Zheng, S., Spangler, S. A., Yu, C., Hoogenraad, C. C., & Zhang, M. (2011). Liprin-mediated large
1386 signaling complex organization revealed by the liprin-alpha/CASK and liprin-alpha/liprin-beta
1387 complex structures. *Mol Cell*, *43*(4), 586-598. https://doi.org/10.1016/j.molcel.2011.07.021
1388 Wickham, H. (2016). ggplot2. In: Springer International Publishing:.
1389 Wulfridge, P., Langmead, B., Feinberg, A. P., & Hansen, K. D. (2019). Analyzing whole genome
1390 bisulfite sequencing data from highly divergent genotypes. *Nucleic Acids Res*, *47*(19), e117.
1391 https://doi.org/10.1093/nar/gkz674

1392 You, L., Yan, K., Zou, J., Zhao, H., Bertos, N. R., Park, M., Wang, E., & Yang, X. J. (2015). The lysine
1393 acetyltransferase activator Brpf1 governs dentate gyrus development through neural stem
1394 cells and progenitors. *PLoS Genet*, *11*(3), e1005034.
1395 https://doi.org/10.1371/journal.pgen.1005034
1396 Zhang, J., Ji, F., Liu, Y., Lei, X., Li, H., Ji, G., Yuan, Z., & Jiao, J. (2014). Ezh2 regulates adult hippocampal
1397 neurogenesis and memory. *J Neurosci*, *34*(15), 5184-5199.
1398 https://doi.org/10.1523/JNEUROSCI.4129-13.2014
1399 Zhang, W., Macias-Velasco, J. F., Zhuo, X., Belter, E. A., Jr., Tomlinson, C., Garza, J., Tekkey, N., Li, D.,
1400 & Wang, T. (2025). methylGrapher: genome-graph-based processing of DNA methylation
1401 data from whole genome bisulfite sequencing. *Nucleic Acids Res*, *53*(3).
1402 https://doi.org/10.1093/nar/gkaf028
1403 Zhou, Q., Obana, E. A., Radomski, K. L., Sukumar, G., Wynder, C., Dalgard, C. L., & Doughty, M. L.
1404 (2016). Inhibition of the histone demethylase Kdm5b promotes neurogenesis and
1405 derepresses Reln (reelin) in neural stem cells from the adult subventricular zone of mice. *Mol*
1406 *Biol Cell*, *27*(4), 627-639. https://doi.org/10.1091/mbc.E15-07-0513

1407

Figure 1

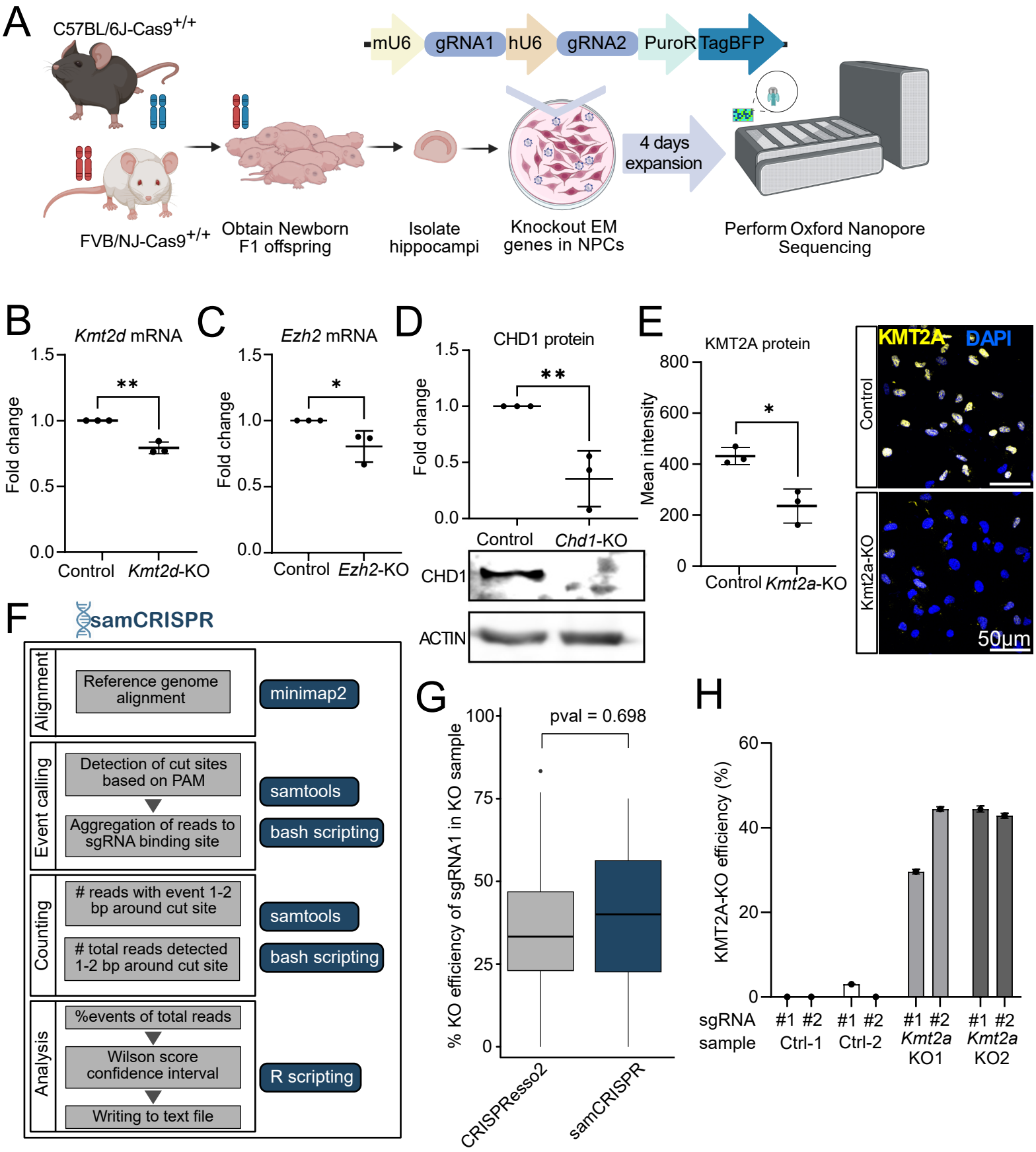

A
C57BL/6J-Cas9+/+
FVB/NJ-Cas9+/+
Obtain Newborn F1 offspring
Isolate hippocampi
mU6
gRNA1
hU6
gRNA2
PuroR
TagBFP
Knockout EM genes in NPCs
4 days expansion
Perform Oxford Nanopore Sequencing
B
Kmt2d mRNA
Fold change
Control
Kmt2d-KO
C
Ezh2 mRNA
Control
Ezh2-KO
D
CHD1 protein
Control
Chd1-KO
CHD1
ACTIN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

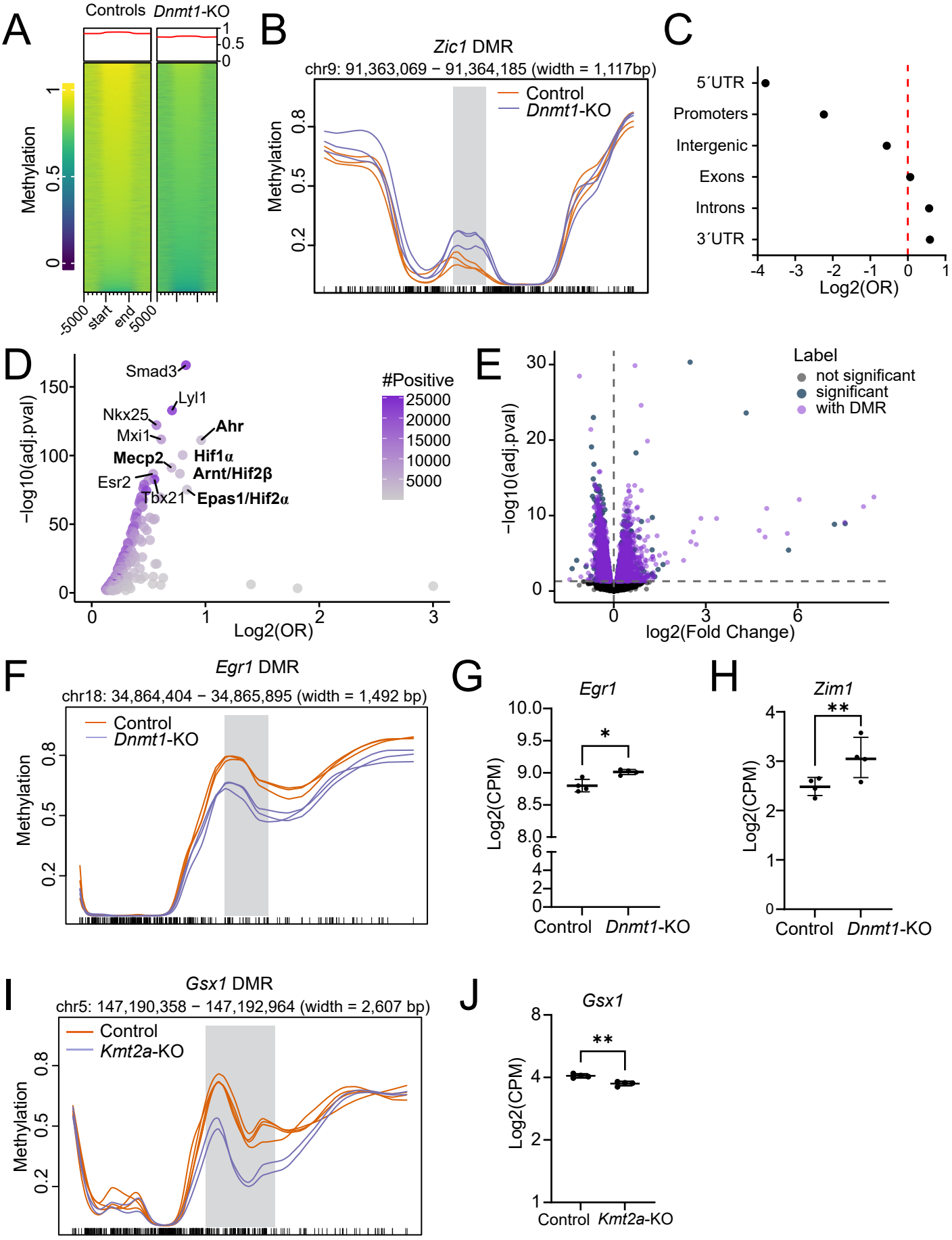

A
Controls
Dnmt1-KO
Methylation
-5000
start
end
5000
B
Zic1 DMR
chr9: 91,363,069 – 91,364,185 (width = 1,117bp)
Control
Dnmt1-KO
Methylation
C
5´UTR
Promoters
Intergenic
Exons
Introns
3´UTR
Log2(OR)
D
Smad3
Lyl1
Nkx25
Mxi1
Ahr
Mecp2
Hif1α
Arnt/Hif2β
Esr2
Tbx21
Epas1/Hif2α
#Positive
25000
20000
15000
10000
5000
−log10(adj.pval)
Log2(OR)
E
Label
not significant
significant
with DMR
−log10(adj.pval)
log2(Fold Change)
F
Egr1 DMR
chr18: 34,864,404 – 34,865,895 (width = 1,492 bp)
Control
Dnmt1-KO
Methylation
G
Egr1
Log2(CPM)
Control
Dnmt1-KO
H
Zim1
Log2(CPM)
Control
Dnmt1-KO
I
Gsx1 DMR
chr5: 147,190,358 – 147,192,964 (width = 2,607 bp)
Control
Kmt2a-KO
Methylation
J
Gsx1
Log2(CPM)
Control
Kmt2a-KO

Figure 3

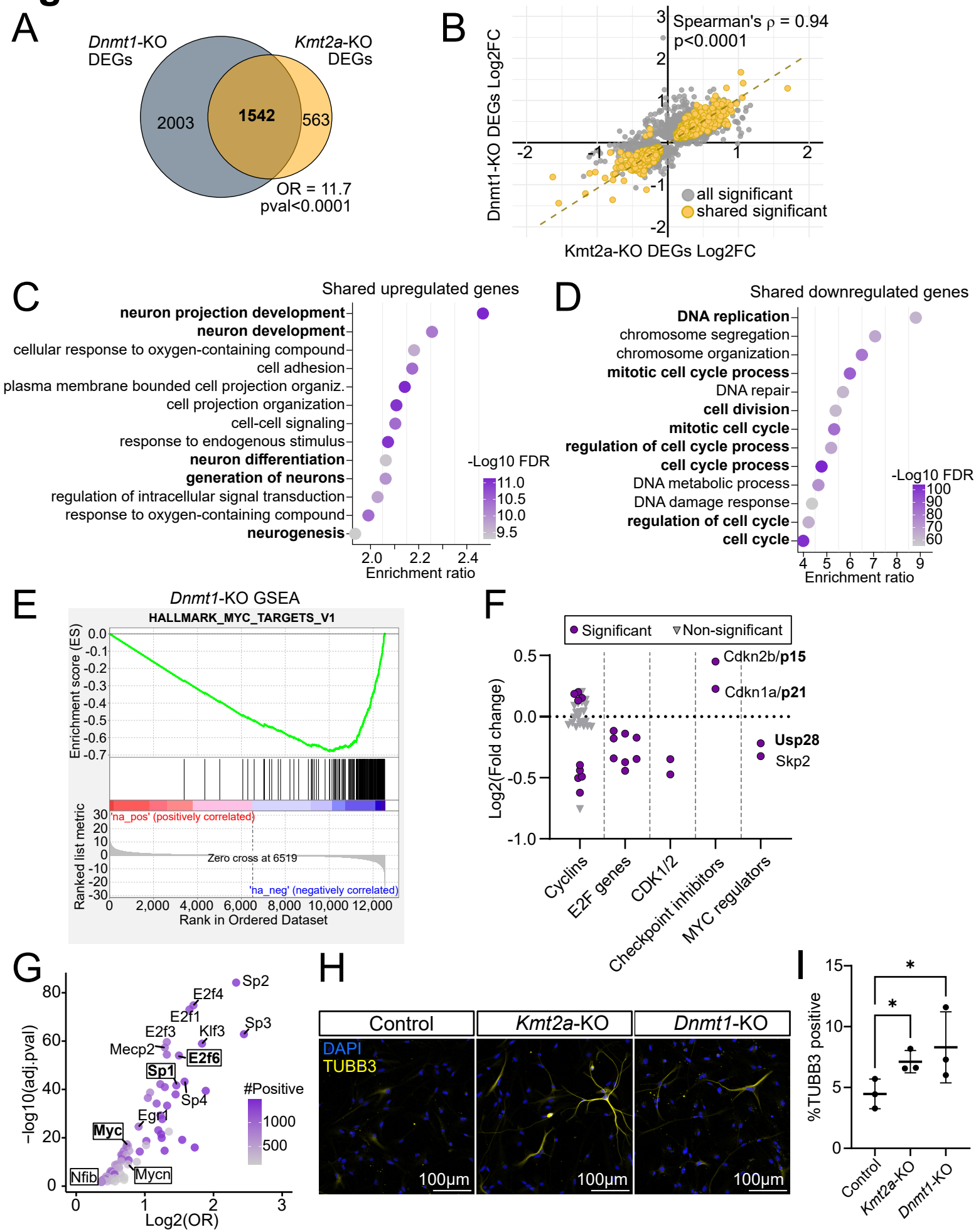

A
Dnmt1-KO DEGs
Kmt2a-KO DEGs
2003
1542
563
OR = 11.7
pval<0.0001
B
Spearman's ρ = 0.94
p<0.0001
Dnmt1-KO DEGs Log2FC
Kmt2a-KO DEGs Log2FC
all significant
shared significant
C
Shared upregulated genes
neuron projection development
neuron development
cellular response to oxygen-containing compound
cell adhesion
plasma membrane bounded cell projection organiz.
cell projection organization
cell-cell signaling
response to endogenous stimulus
neuron differentiation
generation of neurons
regulation of intracellular signal transduction
response to oxygen-containing compound
neurogenesis
-Log10 FDR
11.0
10.5
10.0
9.5
Enrichment ratio
D
Shared downregulated genes
DNA replication
chromosome segregation
chromosome organization
mitotic cell cycle process
DNA repair
cell division
mitotic cell cycle
regulation of cell cycle process
cell cycle process
DNA metabolic process
DNA damage response
regulation of cell cycle
cell cycle
-Log10 FDR
100
90
80
70
60
Enrichment ratio
E
Dnmt1-KO GSEA
HALLMARK_MYC_TARGETS_V1
Enrichment score (ES)
'na_pos' (positively correlated)
Zero cross at 6519
'na_neg' (negatively correlated)
Ranked list metric
Rank in Ordered Dataset
F
Significant
Non-significant
Log2(Fold change)
Cdkn2b/p15
Cdkn1a/p21
Usp28
Skp2
Cyclins
E2F genes
CDK1/2
Checkpoint inhibitors
MYC regulators
G
−log10(adj.pval)
Log2(OR)
Sp2
E2f4
E2f1
Sp3
E2f3
Klf3
Mecp2
E2f6
Sp1
Sp4
Egr1
Myc
Nfib
Mycn
#Positive
1000
500
H
Control
Kmt2a-KO
Dnmt1-KO
DAPI
TUBB3
100µm
I
%TUBB3 positive
Control
Kmt2a-KO
Dnmt1-KO

Figure 4

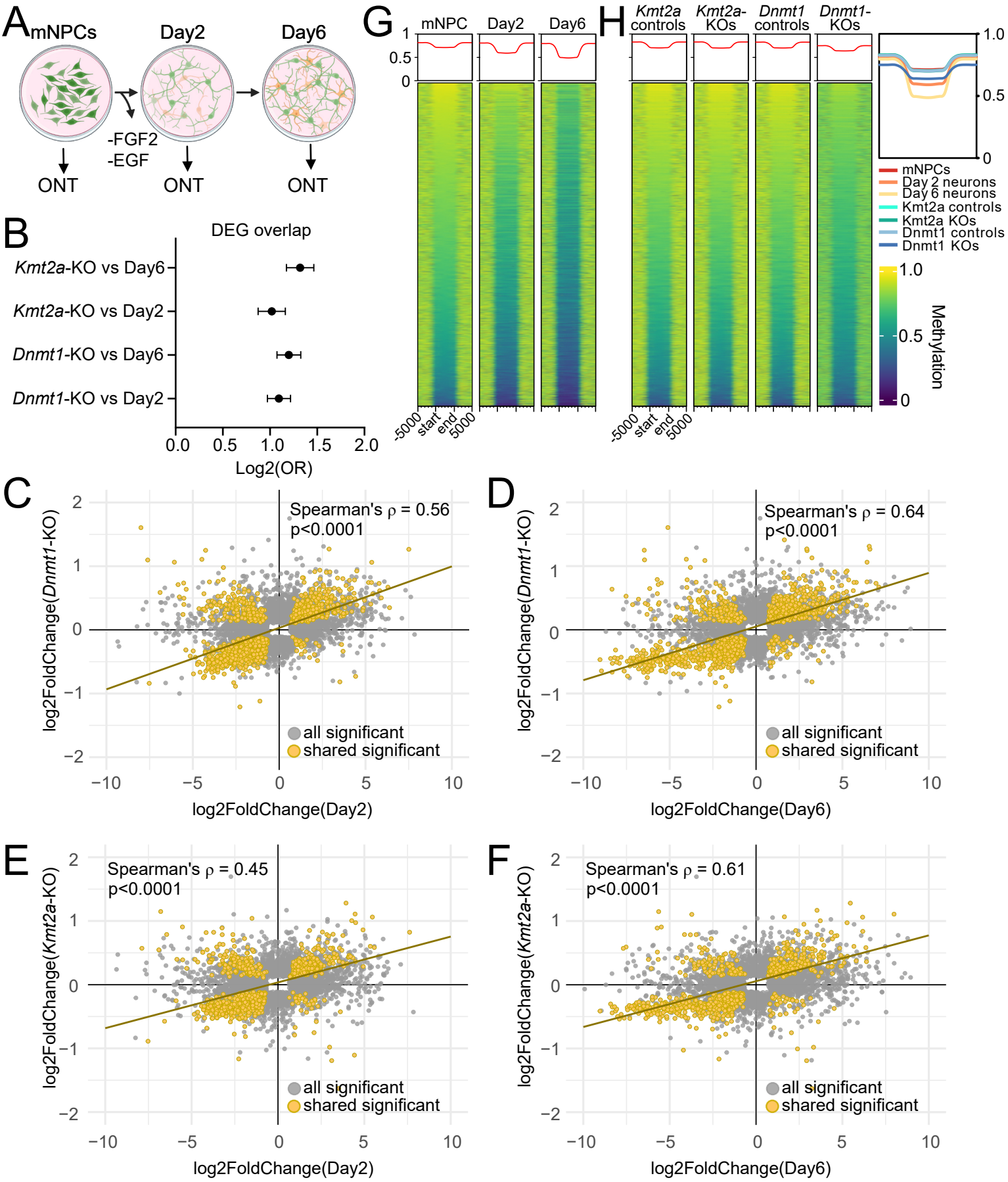

A
mNPCs
Day2
Day6
-FGF2
-EGF
ONT
ONT
ONT
B
DEG overlap
Kmt2a-KO vs Day6
Kmt2a-KO vs Day2
Dnmt1-KO vs Day6
Dnmt1-KO vs Day2
0.0
0.5
1.0
1.5
2.0
Log2(OR)
G
mNPC
Day2
Day6
-5000
start
end
5000
H
Kmt2a controls
Kmt2a-KOs
Dnmt1 controls
Dnmt1-KOs
mNPCs
Day 2 neurons
Day 6 neurons
Kmt2a controls
Kmt2a KOs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

Figure 5

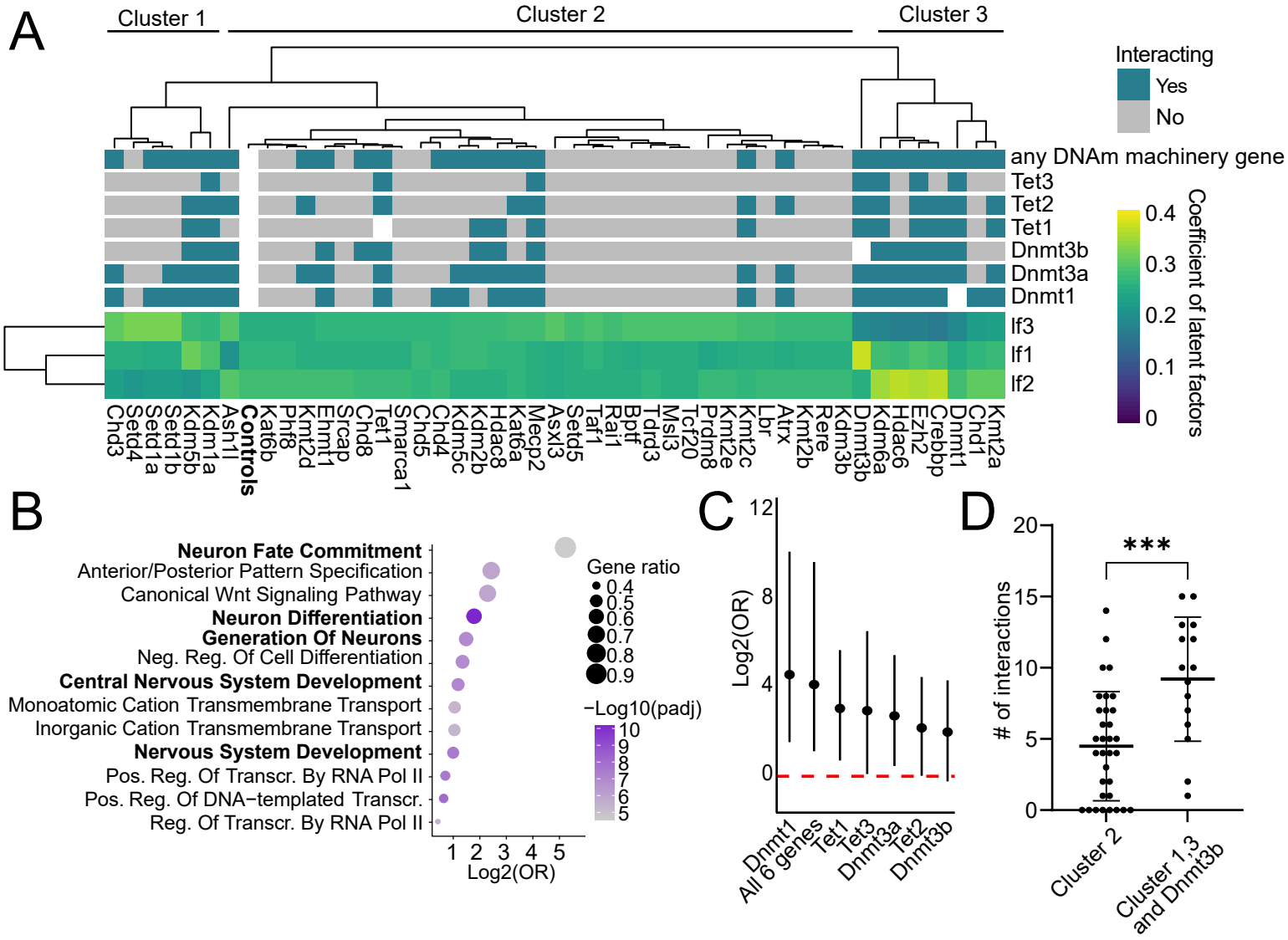
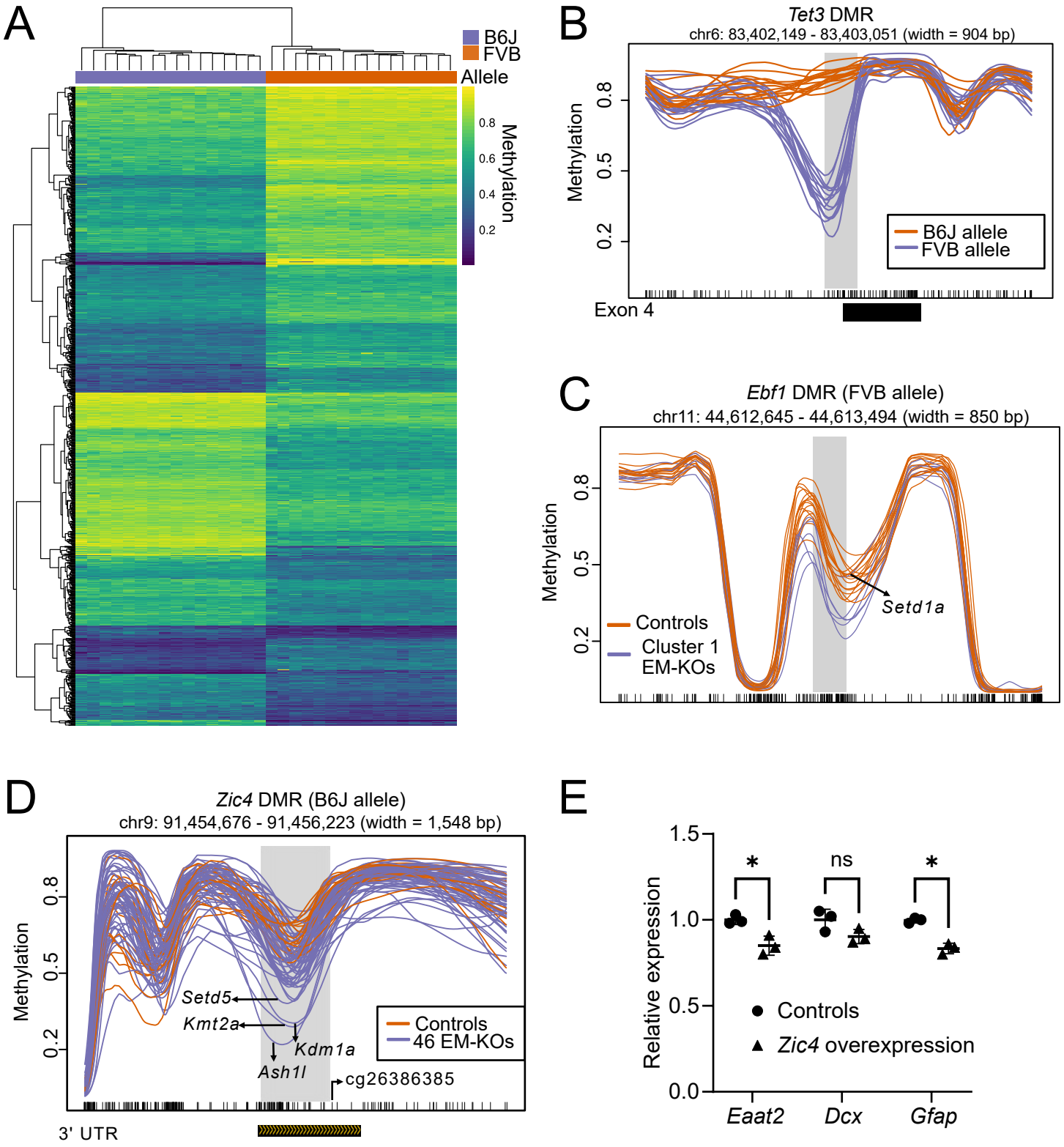

A
Cluster 1
Cluster 2
Cluster 3
Interacting
Yes
No
any DNAm machinery gene
Tet3
Tet2
Tet1
Dnmt3b
Dnmt3a
Dnmt1
If3
If1
If2
Coefficient of latent factors
0.4
0.3
0.2
0.1
0
Chd3
Setd4
Setd1a
Setd1b
Kdm5b
Kdm1a
Ash1l
Controls
Kat6b
Phf8
Kmt2d
Ehmt1
Srcap
Chd8
Tet1
Smarca1
Chd5
Chd4
Kdm5c
Kdm2b
Hdac8
Kat6a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

Figure 6


A
B6J
FVB
Allele
Methylation
0.8
0.6
0.4
0.2
B
Tet3 DMR
chr6: 83,402,149 - 83,403,051 (width = 904 bp)
Methylation
0.8
0.5
0.2
B6J allele
FVB allele
Exon 4
C
Ebf1 DMR (FVB allele)
chr11: 44,612,645 - 44,613,494 (width = 850 bp)
Methylation
0.8
0.5
0.2
Setd1a
Controls
Cluster 1
EM-KOs
D
Zic4 DMR (B6J allele)
chr9: 91,454,676 - 91,456,223 (width = 1,548 bp)
Methylation
0.8
0.5
0.2
Setd5
Kmt2a
Kdm1a
Ash1l
cg26386385
Controls
46 EM-KOs
3' UTR
E
Relative expression
1.5
1.0
0.5
0.0
*
ns
*
Controls
Zic4 overexpression
Eaat2
Dcx
Gfap