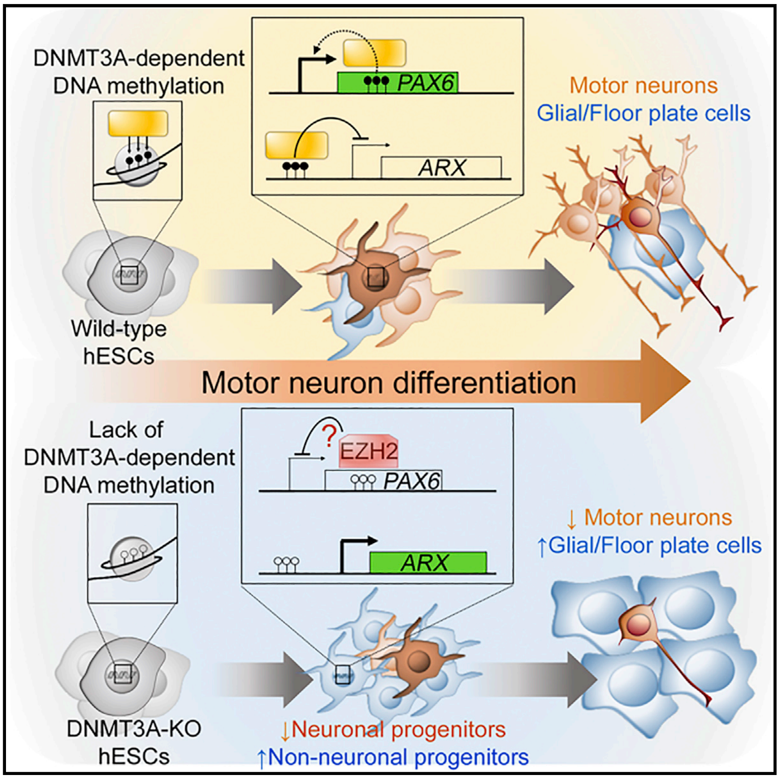
Article

Cell Stem Cell

# Dissecting the Functional Consequences of *De Novo* DNA Methylation Dynamics in Human Motor Neuron Differentiation and Physiology

## Graphical Abstract



## Highlights

- DNMT3A KO alters the fate of ESC-derived neural cells by modulating transcription factor expression
- Single-cell RNA-seq reveals induction of floor plate cells in DNMT3A KO cultures
- Targeted DNA methylation editing of PAX6/ARX gene regulatory elements rescues differentiation defects
- Lack of DNMT3A impairs morphology and physiological activity of MNs

## Authors

Michael J. Ziller, Juan A. Ortega, Katharina A. Quinlan, ..., C.J. Heckman, Alexander Meissner, Evangelos Kiskinis

## Correspondence

meissner@molgen.mpg.de (A.M.),
evangelos.kiskinis@northwestern.edu (E.K.)

## In Brief

Kiskinis and colleagues demonstrate that DNA methylation dynamics play a central role in the differentiation of human pluripotent stem cells toward highly specialized motor neurons. Through a combination of molecular and functional analysis they identify key transcriptional mediators of these effects and link DNA methylation to neuronal patterning and function.

Ziller et al., 2018, Cell Stem Cell *22*, 559–574
April 5, 2018 © 2018 Elsevier Inc.
https://doi.org/10.1016/j.stem.2018.02.012

Cell Stem Cell
Article

# Dissecting the Functional Consequences of *De Novo* DNA Methylation Dynamics in Human Motor Neuron Differentiation and Physiology

Michael J. Ziller,[1,9] Juan A. Ortega,[2,9] Katharina A. Quinlan,[3] David P. Santos,[2] Hongcang Gu,[4] Eric J. Martin,[2] Christina Galonska,[5] Ramona Pop,[6] Susanne Maidl,[1] Alba Di Pardo,[2] Mei Huang,[7] Herbert Y. Meltzer,[7,8] Andreas Gnirke,[4] C.J. Heckman,[8] Alexander Meissner,[4,5,6,*] and Evangelos Kiskinis[2,8,10,*]
[1]Department of Translational Psychiatry, Max Planck Institute of Psychiatry, Munich 80804, Germany
[2]The Ken & Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[3]Department of Biomedical and Pharmaceutical Sciences, George and Anne Ryan Institute for Neuroscience, University of Rhode Island, Kingston, RI 02881, USA
[4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany
[6]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[7]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[8]Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: meissner@molgen.mpg.de (A.M.), evangelos.kiskinis@northwestern.edu (E.K.)
https://doi.org/10.1016/j.stem.2018.02.012

## SUMMARY

**The somatic DNA methylation (DNAme) landscape is established early in development but remains highly dynamic within focal regions that overlap with gene regulatory elements. The significance of these dynamic changes, particularly in the central nervous system, remains unresolved. Here, we utilize a powerful human embryonic stem cell differentiation model for the generation of motor neurons (MNs) in combination with genetic mutations in the *de novo* DNAme machinery. We quantitatively dissect the role of DNAme in directing somatic cell fate with high-resolution genome-wide bisulfite-, bulk-, and single-cell-RNA sequencing. We find defects in neuralization and MN differentiation in DNMT3A knockouts (KO) that can be rescued by the targeting of DNAme to key developmental loci using catalytically inactive dCas9. We also find decreased dendritic arborization and altered electrophysiological properties in DNMT3A KO MNs. Our work provides a list of DNMT3A-regulated targets and a mechanistic link between *de novo* DNAme, cellular differentiation, and human MN function.**

## INTRODUCTION

DNA methylation (DNAme) is a widespread and essential epigenetic modification predominantly targeted to CpG dinucleotides in normal mammalian development (Smith and Meissner, 2013). Establishment and maintenance of DNAme patterns are carried out by two distinct groups of DNA methyltransferase (DNMT) enzymes. DNMT1 is mainly responsible for genome-wide maintenance of DNAme marks after replication, while DNMT3A (3A) and DNMT3B (3B) operate as *de novo* methyltransferases (Jeltsch, 2006). The somatic DNAme landscape, once established in the early postimplantation embryo, is relatively stable on a genome-wide scale. However, focal regions with predominantly regulatory function remain highly plastic over the course of development and differentiation (Ziller et al., 2013). Conditional knockout (KO) studies of DNMT enzymes have highlighted the importance of DNAme in the control of lineage-fate decision, even after embryonic development has been completed (Bröske et al., 2009; Dhawan et al., 2011; Feng et al., 2010; Trowbridge et al., 2009). These studies suggest a link between DNAme and the regulation of cellular fate during differentiation (Smith and Meissner, 2013). Yet, while properly directed DNAme patterns are required for lineage-fate choices during development, the importance, as well as the specific targets of *de novo* DNAme in establishing cellular identity and controlling cellular traits, remains poorly defined.

*Dnmt3a* and *Dnmt3b* exhibit a dynamic expression pattern in the central nervous system (CNS), where evidence from mouse models suggests they play diverse and critical roles during neural development and function (Feng et al., 2005, 2010; Nguyen et al., 2007; Wu et al., 2010). Mice with a targeted deletion of *Dnmt3a* in the nervous system are born healthy but degenerate in adulthood and die prematurely (Nguyen et al., 2007). They exhibit prominent defects in motor movement, fewer motor neurons (MNs), and an accumulation of fragmented endplates in neuromuscular junctions. This phenotype suggests that *Dnmt3a* might play an important role in the development and function of the neuromuscular circuit. However, besides these broad phenotypes, little is known about the underlying mechanisms and the molecular targets of the *de novo* DNAme in human MN differentiation. In particular, it remains unclear whether the lack of properly established DNAme patterns during cell differentiation can directly impair the physiological properties of the resulting

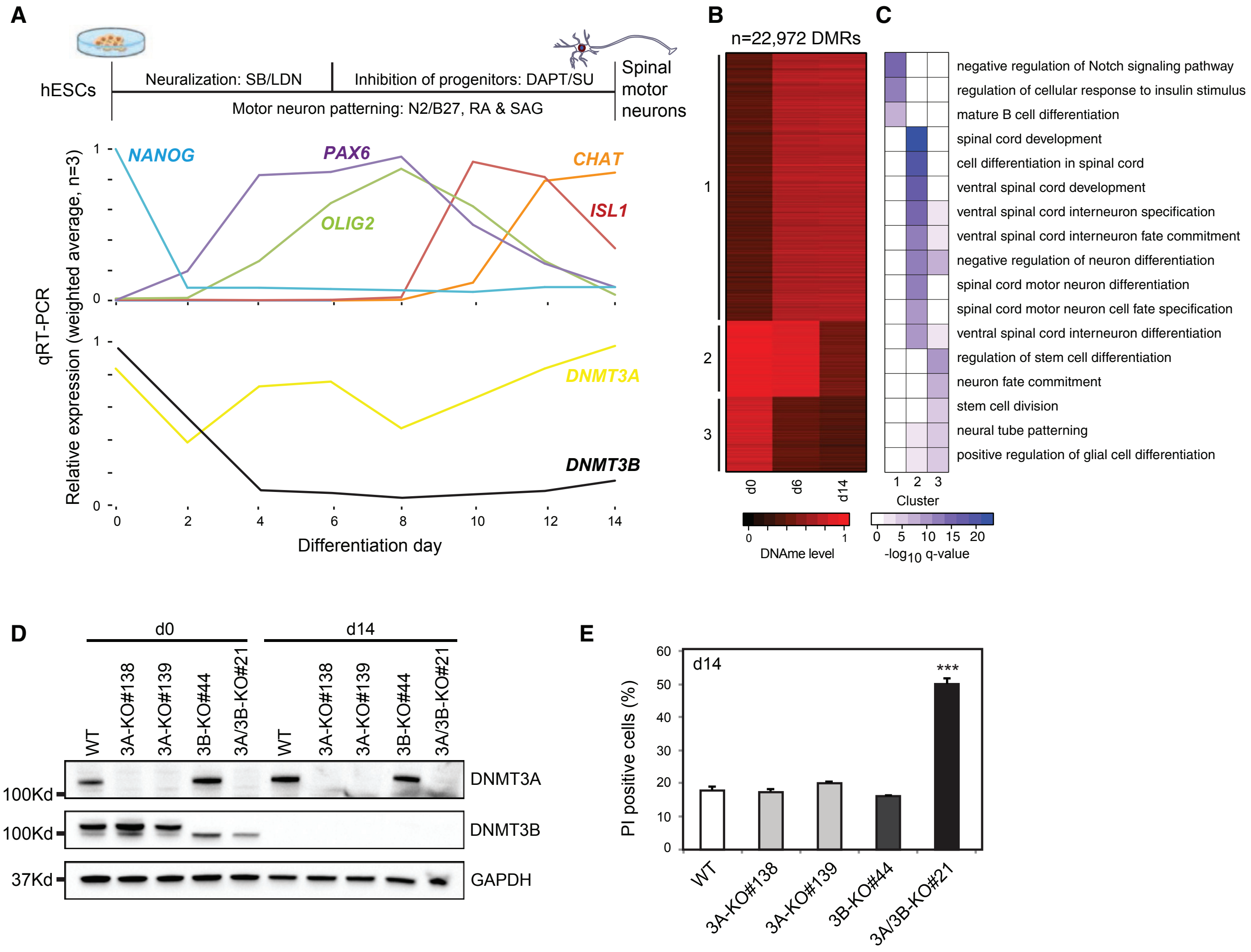


**Figure 1. MN Differentiation from Human ESC Offers a Platform for Delineating the Regulatory and Functional Consequences of *De Novo* DNAme**

(A) Top: schematic of 14-day differentiation protocol indicating the timing and use of different small molecules (see STAR Methods for details). Bottom: qRT-PCR analysis for selected genes. Results represent the averages of 3 independent differentiation experiments that each include 3 technical replicates per time point. Values are normalized against the time point of highest expression for each of the genes.

(B) Clustering of all regions that exhibit significant differential DNAme during MN differentiation. DMRs were defined based on a posterior probability $\leq 0.05$ and an absolute DNAme difference $\geq 0.2$, requiring a minimum of 4 CpGs. Numbers and black bars indicate different DMR clusters. See Table S2.

(C) Subset of gene set enrichment analysis for each cluster in (B). See Table S3.

(D) Western blot analysis of 3A and 3B in the parental HUES-64 line (WT), two 3A-KO lines (#138 and #139), one 3B-KO line (#44), and the double-3A/3B-KO line (#21) in undifferentiated stem cells (d0) and differentiated MN cultures (d14). The catalytically active 3B isoform is abolished in 3B-KO cells. GAPDH used as a loading control.

(E) Percentage of propidium iodide $(PI)^+$ apoptotic cells measured by FACS on d14 of MN differentiation presented as means ± SEM; ANOVA (***$p < 0.001$).

neural populations, in addition to modulating lineage-fate decisions or inducing proliferation or apoptosis. What also remains unclear is the functional redundancy between the two *de novo* DNMT3 enzymes and whether they serve overlapping or distinct purposes in cellular systems.

## RESULTS

### Differentiation of ESCs toward MNs Is Coupled to Focal DNAme Dynamics

We used the human embryonic stem cell (ESC) line HUES-64 (Chen et al., 2009) to derive ventral spinal cord cells, enriched for MNs (Figure 1A, top). To characterize the precise molecular events that lead to the generation of MNs, we examined the gene expression pattern of key developmental markers (Figure 1A, bottom). Pluripotency-associated genes, such as *NANOG*, are rapidly downregulated, while the early neuronal progenitor transcription factor (TF) *PAX6* is rapidly induced, followed by transient upregulation of the MN progenitor TF *OLIG2* (Figure 1A, bottom). These changes are trailed by a second wave of induction of more terminal MN markers such as *ISL1* and *CHAT*. Notably, the two *de novo* DNMT display opposing dynamics, with *3B* being rapidly downregulated and *3A* exhibiting a dynamic expression pattern with sustained high levels throughout, particularly when compared to other somatic cell types (Figure S1A). To determine the 3A/B protein levels, we performed a high-resolution time course western blot (WB) analysis, which confirmed the rapid downregulation of 3B and a transient reduction followed by steady expression of 3A after 40 days *in vitro* (Figure S1B).

To define all regions with dynamic DNAme during MN differentiation, we generated whole genome bisulfite sequencing (WGBS) data and focused on three time points: day 0, which is the undifferentiated ESC state; day 6, which represents the peak of the generation of neural progenitors; and day 14, which represents the peak of the generation of postmitotic neurons. We generated a median of 500 million 100-bp paired-end reads per time point (Table S1) and identified n = 22,987 differentially methylated regions (DMRs, posterior probability $\leq$0.05, absolute methylation difference $\geq$0.2) between any of these three states (Figure 1B; Table S2). Out of these DMRs, three major clusters emerge. The first and largest cluster consists of DMRs that gain DNAme (63.4%), while clusters two and three represent DMRs that lose DNAme after 6 and 0 days, respectively. The DMRs within these clusters represent highly specific remodeling of the DNAme landscape, with the associated targets matching well with terms of CNS and spinal cord development (Figure 1C, right, Table S3). Genes and gene regulatory elements (GREs) associated with alternative lineages, such as "mature B cell differentiation" (cluster 1), are targeted for DNAme early on. In contrast to these terms and in agreement with the MN differentiation protocol, the terms "neural tube patterning" (cluster 3), "ventral spinal cord development," and "spinal cord MN differentiation" (cluster 2) are sequentially demethylated, presumably to provide access to GREs for key TFs or as a consequence of their activity. Our data show widespread, focal DNAme changes during human MN differentiation, validating it as a model to dissect the regulatory link between DNAme dynamics and cell specification. Our results also show that 3A is the predominant *de novo* methyltransferase expressed during MN differentiation, and it is therefore reasonable to assume it plays a role in the observed DNAme examined at putative GREs (Figure 1B, cluster 1).

It is difficult to ascertain without functional studies whether these dynamics are directly involved in the generation of ventral spinal cord cells and MNs specifically, or simply reflect downstream effects of upstream regulators. To assess their functional relevance, we took advantage of previously generated homozygous *3A*, *3B*, as well as *3A*/*3B* double-KO HUES-64 ESCs (Liao et al., 2015). We validated the precise protein deletions in undifferentiated ESCs and MN-enriched cultures relative to the parental HUES-64 isogenic control (wild-type [WT]) (Figure 1D) and asked whether blocking the acquisition of DNAme would compromise cell viability during MN differentiation. Deletion of both *3A*/*B* caused dramatic apoptosis with more than 50% of the differentiated cells incorporating propidium iodide by day 14, while deletion of *3A* or *3B* had no substantial effect relative to the isogenic control (Figures 1E and S1C).

## Lack of DNMT3A Protein Impairs Neuralization and the Development of MNs

To further probe the significance of the observed *de novo* DNAme in directing ESCs toward ventral spinal cord identities, we differentiated in parallel two *3A-KO* clones (#138 and #139), a *3B-KO*, a double *3A*/*B*-KO, and the parental WT control cell line down the MN lineage, as described above. We evaluated the efficiency of MN generation through immunocytochemistry (ICC) for ISL LIM homeobox gene (ISL)$^+$ (ISL1/2$^+$) and choline acetyltransferase (ChAT)$^+$ MNs on day 16 (Figure 2A). We found a significantly reduced percentage of MNs in the *3A-KOs* (Figure 2B; 30% reduction in ISL$^+$ cells, from 44% in WT to 31% in *3A-KO*; and 40% reduction in the number of ChAT$^+$ cells, from 56% in WT to 33% in *3A-KO*, n = 8, $p < 0.0001$). We also found a reduction in the overall production of neurons (35% reduction in TUJ1$^+$ cells, ~83% in WT to 55% in KOs, n = 8, $p < 0.0001$) (Figure 2B). In contrast, *3B-KO* cells were relatively unaffected, while the few surviving *3A*/*3B* double-KO cultures exhibited *3A-KO*-like defects (Figures 2A and 2B). We validated these findings in the *3A-KO* cultures by measuring protein levels of ChAT and TUJ1 (Figures 2C and S2A), as well as by quantifying the amount of the secreted MN neurotransmitter acetylcholine (Figure 2D). These defects were not temporal, as we consistently detected reduced protein levels of ChAT and TUJ1 between days 14 and 40 in the *3A-KO* cultures (Figure 2E). Taken together, we observe a robust impairment in the generation of neurons and MNs from *3A-KO* ESCs. These differences were consistent for both *3A-KO* clones and indicate that 3A and/or at least a subset of its targets are involved in the MN differentiation process. Previous work showed that mouse neural stem cells lacking *Dnmt3a* exhibited an increased proliferation rate (Wu et al., 2010). To determine whether this defect was consistent in human cells, we generated neural progenitor cells (NPCs) through manual rosette selection and found that *3A-KO* NPCs exhibited increased proliferation relative to their isogenic controls (Figures S2B–S2E).

## Loss of DNMT3A Blocks *De Novo* Methylation and Results in Selective Transcriptional Changes

The differentiation phenotype presents an opportunity to dissect the underlying molecular defects caused by the lack of *de novo* DNAme activity during MN generation. Our results indicate that *3A-KO* ESCs exhibit a shift or block in their differentiation trajectory (Figure 2). Visually, we observed an apparent qualitative and quantitative difference in 3A-KO cultures, that involves the appearance of cells with very prominent nuclei (Figures S2F and S2G). Importantly, however, there was no difference in ectodermal lineage commitment, evaluated by the percentage of cells that stained positive for the neural cell adhesion molecule 1 (NCAM1) on day 14 (Figure S2K).

To complement the phenotypic characterization, we generated WGBS libraries and analyzed DNAme in the *3A-KO* and isogenic control samples (days 0, 6, 14; n = 2–4) (Figure 3A). We did not observe any significant change ($p = 0.12$, two-tailed t test) in the global mean DNAme levels in the undifferentiated and differentiated cells (Figure S3A), confirming that we are investigating localized interference with the *de novo* machinery. We identified 7,763 DMRs arising over the course of MN differentiation between the WT and KO cells (Figures 3B and 3C; Table S2). Of these, 37.4% correspond to differentiation-associated DMRs (from the ones shown in Figure 1B) that are *not* faithfully recapitulated in the KO (Figures 3B, diff DMRs, and 3C, left panel). The rest represent DMRs that arise exclusively in the *3A-KO* and do not correspond to regions that change their

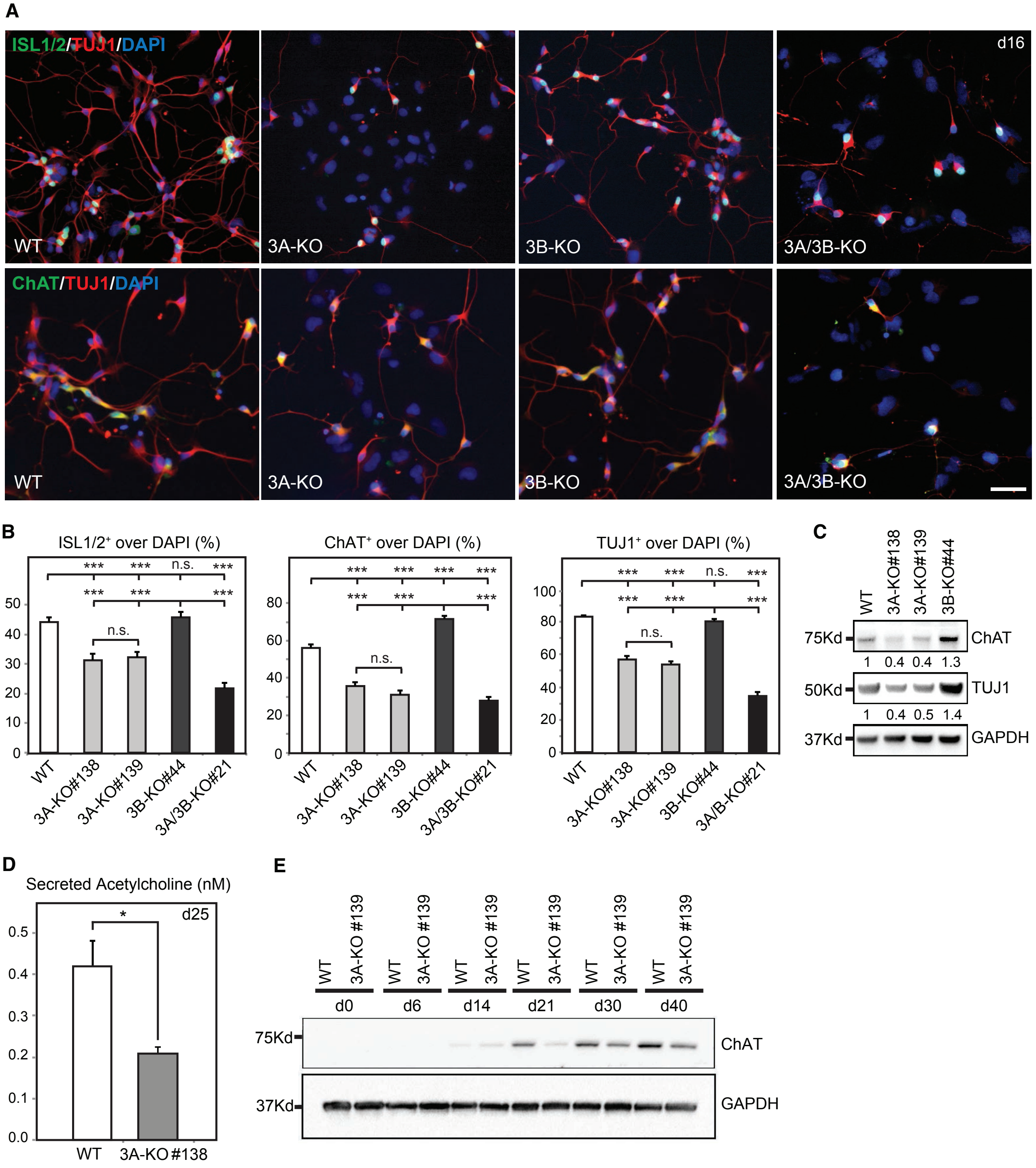


**Figure 2. Deletion of DNMT3A Protein Results in Reduced Numbers of Differentiated Neurons and MNs**

(A) Representative ICC images of WT, 3A-KO, 3B-KO, and 3A/3B-KO cultures at d16 of MN differentiation with the neuronal marker TUJ1 (red) and the MN markers ISL and ChAT (green). Nuclei stained with DAPI (blue). Scale bar, 50 μm.

(B) Quantification of ISL⁺, ChAT⁺, and TUJ1⁺ MN percentages across cell lines and the parental isogenic control at d16 presented as means ± SEM; ANOVA (***$p < 0.001$), n.s: not significant.

(C) Western blot of ChAT and TUJ1 protein in 3A, 3B-KO and isogenic WT d14 differentiation cultures. GAPDH used as a loading control and fold increase values relative to WT are annotated under each lane.

(D) High-performance liquid chromatography (HPLC) quantification of secreted acetylcholine in WT and 3A-KO d25 differentiation cultures presented as means ± SEM; 2-tailed t test (*$p < 0.05$).

(E) Western blot of CHAT protein in 3A-KO and isogenic WT MN cultures over 40 days of differentiation. GAPDH was used as a loading control.

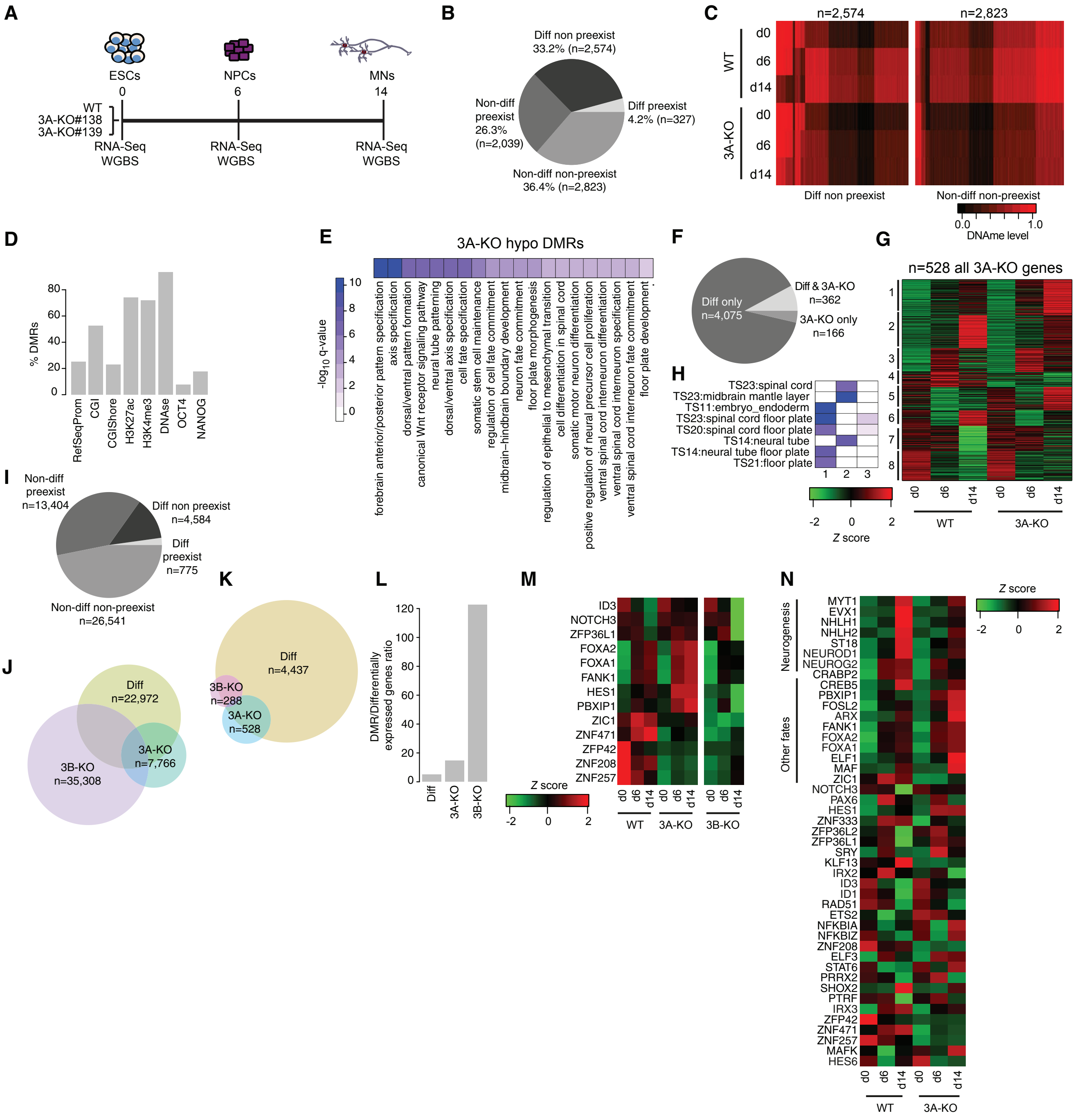


**Figure 3. Loss of DNMT3A Impairs *De Novo* Methylation and Causes Transcriptional Changes**

(A) Cell lines were differentiated in parallel and samples collected for WGBS and RNA-seq analysis on d0, 6, and 14.

(B) DMRs ($p \leq 0.05$, methylation difference $\geq 0.2$) that arise between the WT and 3A-KO during MN differentiation broken down into four groups: differentiation-associated (diff) DMRs that occur between individual time points in the WT differentiation but are impaired in the KO, as well as non-differentiation-associated (non-diff) DMRs that exclusively occur in *3A*-KO. These two groups are subdivided in DMRs that pre-exist in the pluripotent state and ones that occur only over the differentiation.

(C) DNAme levels based on WGBS of non-preexisting DMRs that arise during WT differentiation and are de-regulated in the *3A*-KO (left) and DMRs that arise exclusively in the *3A*-KO differentiation (right).

(D) Percentage of diff DMRs that fail to gain DNAme in the *3A*-KO overlapping with different genomic features: CGI (CpG islands), CGIshore (CpG island shore), H3K27ac, H3K4me3, DNase hypersensitivity (each defined as the union of H3K27ac or H3K4me3 peaks), DNase hypersensitivity sequencing peaks across more than 70 distinct cell and tissue types from the REMC and ENCODE projects, and TF binding sites in HUES-64 for OCT4 and NANOG.

(E) Subset of gene sets enriched in the non-preexisting hypomethylated DMRs. See Table S3.

*(legend continued on next page)*

DNAme state in the WT differentiation (Figures 3B, non-diff DMRs, and 3C, right panel). In line with the loss of a *de novo* DNAme, the vast majority (87.6%) of DMRs are hypomethylated compared to the WT. In addition to these KO hypomethylated DMRs, we also detect 956 (12.3%) regions that arise through the lack of proper de-methylation upon differentiation of the KO cells and likely represent indirect effects of the loss of 3A. Last, a subset of the DMRs that fail to gain DNAme upon differentiation (30.5%) exhibit pre-existing lower DNAme levels in the pluripotent *3A-KO* cells (Figure 3B, preexisting).

To identify genes that fail to acquire proper DNAme during the differentiation, we focused our analysis on the subset of non-preexisting *3A-KO* DMRs that gain DNAme in the WT (non-preexisting differentiation DMRs) (Figure 3C, left, n = 2,574). In line with general DNAme dynamics, only a small fraction is associated with promoter elements, and the majority (>80%) overlaps with putative GREs, as defined by DNase I hypersensitivity, H3K4me3 or H3K27ac positive regions (Figure 3D). A subset of these DMRs likely represent GREs linked to pluripotency, as 20% overlap with binding sites of pluripotency-associated TFs in human ESC (Figure 3D). In addition, there is significant enrichment for GREs associated with genes implicated in axis patterning, alternative neuronal fates such as forebrain and floor plate, and mature MN gene sets (e.g., somatic MN differentiation) (Figure 3E; Table S3).

To get further insight on the functional consequences of deregulated DNAme in the *3A-KO* cells, we performed gene expression profiling using bulk RNA sequencing (RNA-seq) on the differentiating WT and *3A*-KO cells (days 0, 6, 14; n = 3–4) (Table S1). This analysis revealed 4,603 differentially expressed genes upon differentiation (Figure 3F; Table S4). In contrast to the DNAme changes, only a small fraction of these (Figures 3F and 3G, n = 528/11.5%) can be attributed to the lack of 3A (Figure 3G; Table S4), with even fewer genes being differentially expressed during MN differentiation (n = 362). This very specific set of genes was particularly enriched for floor plate genes (cluster 1, upregulated genes), or spinal-cord-associated genes (cluster 2, downregulated genes) (Figure 3H; Table S4).

Given the observation that *3B* is rapidly downregulated during the first 4 days of the differentiation (Figure 1A), we asked to what extent these molecular changes were specific to the *3A-KO*. To address this, we performed WGBS for *3B-KO* cells during MN differentiation (days 0, 6, 14; n = 2). Surprisingly, this analysis revealed a large number of DMRs (n = 35,308), far exceeding the count in the *3A-KO* (Figure 3I). In light of the large number of *3B* DMRs, their substantial overlap with differentiation DMRs (Figure 3J), but lack of robust MN differentiation defects (Figure 2), we investigated the functional consequences of the lack of *3B* on the transcriptome by RNA-seq. We found a minor impact on gene expression (n = 288 genes, Table S4), with only a small fraction overlapping with genes differentially expressed during WT differentiation, or with genes affected in the *3A-KO* (Figures 3K and S3B, n = 79). These genes did not exhibit any substantial enrichment of any gene sets, aside from generic GO categories such as extracellular matrix (Table S4). Against this background of frequent, focal DNAme changes, but lack of large functional consequences (low DMR/differential gene expression ratio, Figure 3L), we tested the hypothesis that many genomic regions affected by *3B-KO* might be relevant in other cell types and lineages. Thus, we compared the DNAme levels of these regions across a panel of 25 cell types with available WGBS data (Figure S3C). This analysis showed low and intermediate DNAme levels in several other non-neural tissues, suggesting that a substantial subset of regions that fail to gain DNAme in the *3B-KO* are likely relevant in other cellular contexts and lineages, but less so for neuronal differentiation. To better understand the molecular differences between the *3A* and *3B-KO* that could explain the clear phenotypic differences observed, we investigated the TFs differentially expressed in both genetic backgrounds. While the intersection is small, most of the overlapping TFs show opposing trends in *3A* and *3B-KO*s (Figure 3M). Similarly, most other TFs uniquely impaired in the *3B-KO* exhibit reduced expression during MN differentiation compared to the WT and are not directly implicated in MN differentiation (Figure S3D).

The TFs that are de-regulated in the *3A-KO* represent many with well-defined roles in CNS development. In particular, we find a striking lack of induction of pro-neuronal factors in the *3A*-KO cell population on day 14, including *MYT1*, *NEUROD1*, *EVX1*, and *NEUROG2* (neurogenesis cluster in Figure 3N). Furthermore, TFs associated with other neural cell fates, such as *ARX*, *FOXA1*, and *FOXA2* that are expressed in floor plate cells, were significantly upregulated (other fates cluster in Figure 3N). We also find that *PAX6*, a marker of neuronal progenitors, and *HES1*, which marks gliogenic progenitors, exhibit divergent profiles. *3A*-KO cells express low *PAX6* levels and fail to turn off *HES1*.

---

(F) Fraction of differentially expressed genes ($q \leq 0.01$, absolute $\log_2$ fold change $\geq 1$, minimum expression $\geq 10$ reads per kilobase per million mapped reads [RPKM]) arising in the differentiation only or between the WT and KO differentiation stages. See Table S4.

(G) Clustering of all genes differentially expressed between the WT and KO across the differentiation time course (*Z* scores). See Table S4.

(H) Subset of gene set enrichment analysis for the Mouse Genome Informatics (MGI) database of mouse tissues across development. Results are shown as odds-ratios significant at a Benjamini-Hochberg (BH)-corrected p value below 0.05 (Fisher's exact test) for expression clusters 1–3 in (G). See Table S4.

(I) All DMRs ($p \leq 0.05$, methylation difference $\geq 0.2$) that arise between the WT and 3B-KO during MN differentiation broken down into four groups: diff DMRs that occur between individual time points in the WT differentiation but are impaired in the KO as well as non-differentiation-associated DMRs (non-diff) that exclusively occur in *3B*-KO. These two groups are subdivided in DMRs that pre-exist in the pluripotent state and ones that occur only over the differentiation.

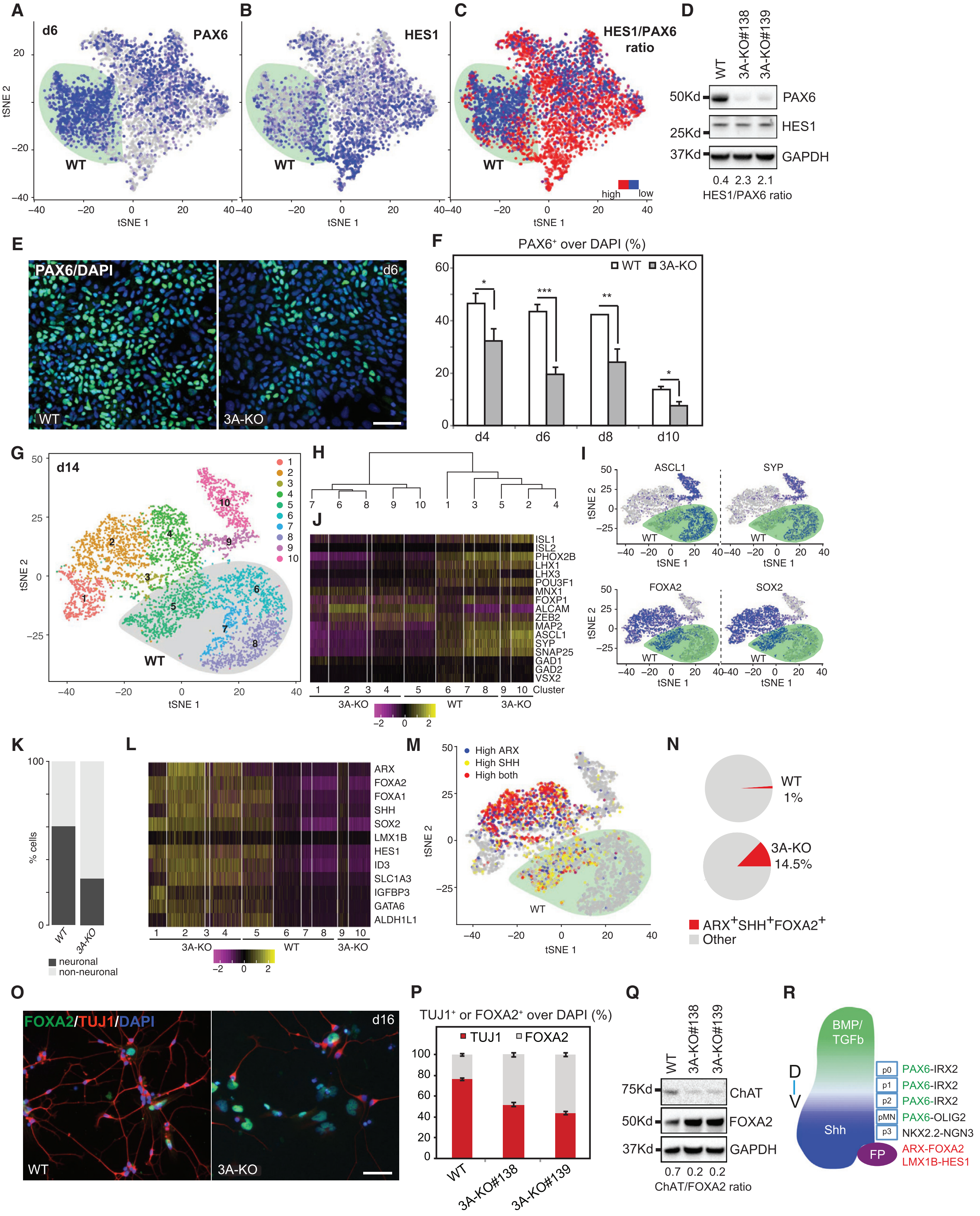
(J) Overlap of all DMRs detected in the WT differentiation (diff), between the WT and the *3A*-KO differentiation (3A-KO) and the WT and *3B*-KO differentiation (3B-KO).

(K) Overlap of all differentially expressed genes detected in the WT differentiation (Diff), between the WT and the *3A*-KO differentiation (3A-KO) and the WT and the *3B*-KO differentiation (3B-KO). See Table S4.

(L) Ratio of the number of DMRs and differentially expressed genes detected in WT, *3A*-KO, and *3B*-KO differentiations.

(M) Expression levels (*Z* scores) of TFs that are differentially expressed in the *3A*-KO and *3B*-KO differentiation compared to the WT.

(N) Expression levels (*Z* scores) of all TFs that are differentially expressed between the WT and *3A*-KO differentiation.



*(legend on next page)*

In summary, these data indicate that the KO of *3B*, which is downregulated during the first stages of MN differentiation, affects a large number of DMRs, but with apparently little functional consequences in our differentiation. In contrast, our analysis point toward a role of 3A in silencing GREs associated with pluripotency and alternative neural cell lineages. The failure to properly regulate them via DNAme translates into transcriptional changes and suggests a shift in the differentiation trajectory of the *3A-KO* cells toward non-neurogenic cell fates.

### DNMT3A KO Impairs Neurogenesis and Leads to the Generation of Gliogenic Progenitors and Floor Plate Cells

To validate these phenotypes in individual cells, we performed single-cell RNA-seq analysis and used ICC and WB after 6 and 16 days of differentiation in *3A*-KO and isogenic control cells. We examined the ratio of HES1/PAX6 in progenitors to determine the balance between a gliogenic (higher) and neurogenic (lower) fate (Sugimori et al., 2007). Single-cell analysis of day 6 progenitors (>5,000 cells, Figure S4A) showed a clear separation between the WT and the 3A populations, with *PAX6* and *HES1* mRNA levels demarcating this segregation (Figures 4A, 4B, S4B, and S4C). The majority of WT cells were *PAX6*$^+$, while *3A*-KOs where more diverse, with PAX6$^+$ cells representing only a sub-population. In contrast, *HES1* expression was evenly distributed among both genotypes. Thus, the majority of *3A*-KO cells had a high *HES1*/*PAX6* mRNA (Figure 4C) and protein ratio (Figure 4D). This indicates that in the absence of 3A there is a shift in the neurogenic potential toward a more gliogenic fate. PAX6 marks all neuronal progenitor domains in the developing spinal cord, including MN progenitors (pMNs), which eventually give rise to MNs (Figure 4R) (Ericson et al., 1997). Using qRT-PCR, we found that, while *PAX6* maintains a normal developmental pattern of early induction followed by downregulation, it is consistently reduced in KOs (Figure S4D). We next used ICC to measure PAX6 protein at single-cell resolution and found that significantly fewer PAX6 cells were generated as early as day 4 and throughout the differentiation (50% reduction on day 6) (Figures 4E, 4F, and S4E–S4G). As a semiquantitative measure of PAX6 protein, we evaluated the level of intensity of PAX6 staining and found that KO progenitors consistently exhibited lower intensity than their isogenic control counterparts (Figure S4H). This deregulation also impacted the number of MN progenitors, evident by a reduction in OLIG2$^+$ and PAX6/OLIG2$^+$ cells in the *3A*-KO cultures (Figures S4E–S4G).

We next analyzed single-cell gene expression on day 14 and found a clear segregation between WT and *3A*-KO cells in a 2D scatterplot (Figure 4G). Although we examined a roughly equal number of WT and KO cells (Figure S4I), the KOs were more heterogeneous and comprised six sub-clusters (1–4, 9, 10), while the WTs only four (5–8) (Figure 4H). The clusters separated into two branches: a neuronal one, defined by the high expression levels of *ASCL1* and synaptic vesicle protein *SYP*, and a neural and non-neuronal one, discernable by the expression of *FOXA2* and the neural progenitor marker *SOX2* (Figures 4I and S4K). The neuronal cells were represented by two closely related clusters in the KO population (9, 10), and three closely related clusters in the WT population (6, 7, 8). These neuronal clusters widely expressed markers of MN columns, while interneuron markers, such as *GAD1* and *GAD2*, were absent (Figure 4J). In contrast, the non-neuronal cells were twice as many (Figure 4K), and in line with their proliferative potential, we find increased numbers of cycling Ki67$^+$ cells in the *3A*-KO cultures (Figure S4J). The non-neuronal clusters expressed high levels of floor plate markers (*SHH*, *FOXA2*, *ARX*, *T-negative*), early gliogenic genes (*SLC1A3*, *IGFBP3*, *GATA6*, *ALDH1L1*), and neural progenitor markers (*SOX2*, *HES1*, *ID3*) (Figures 4L and S4L). Although the only WT non-neuronal cluster (5) was similar to the ones produced in the KO cultures (1, 2, 3, 4), the induction of a population of triple positive cells for *SHH*, *ARX*, and *FOXA2* (14.5%), likely representing genuine floor plate, was unique for the *3A-KO* (Figures 4M and 4N). We verified by ICC that the *3A-KO* cells exhibited a high proportion of FOXA2$^+$/TUJ1$^-$ cells (Figures 4O and 4P), as well as a dramatic shift in the ratio of ChAT and FOXA2 protein (Figure 4Q). These findings are in concordance with the bulk and single-cell RNA-seq data, as well as the ICC and biochemical validation of reduced neuronal production (Figure 2). We analyzed the correlation

**Figure 4. Lack of Functional DNMT3A Protein Impairs Neurogenesis and Promotes the Generation of Neural Floor Plate and Gliogenic Cells**

(A–C) Single-cell RNA-seq showing (A) HES1$^+$, (B) PAX6$^+$ cells, and (C) HES1/PAX6 mRNA ratio along the transcriptionally segregated WT (circled) and *3A*-KO cultures at d6 of MN differentiation.

(D) Western blot of HES1 and PAX6 protein in *3A*-KO and WT cultures. GAPDH used as a loading control. The ratio of the signal was quantified using Fiji.

(E and F) Percentage of (E) PAX6$^+$ cells in *3A*-KO and WT cultures quantified by ICC over MN differentiation (F) presented as means ± SEM, 2-tailed t test (*p < 0.05, **p < 0.01, ***p < 0.001). Scale bar, 50 μm.

(G) t-distributed stochastic neighbor embedding (tSNE) analysis of single-cell transcriptional profiles of WT and *3A*-KO d14 cultures grouped into 10 distinct cell sub-populations. WT (circled) segregated from *3A*-KO cells.

(H) Dendrogram of hierarchical clustering of the 10 existing subpopulations in WT and *3A*-KO d14 cultures.

(I) Distribution of neuronal (*ASCL1*, *SYP*) and non-neuronal (*FOXA2*, *SOX2*) genes in the tSNE space across WT and *3A*-KO cultures.

(J) Single-cell expression levels of key neuronal markers among the WT and *3A*-KO subpopulations identified by single-cell RNA (scRNA)-seq shown as *Z* scores.

(K) Fraction of cells classified as neuronal and non-neuronal in the WT and *3A*-KO condition at d14.

(L) Expression pattern (*Z* scores) of floor plate and glial genes across the subpopulations defined in (G).

(M) tSNE plot showing the distribution of WT (circled) and *3A*-KO cells expressing high levels of *ARX*, *SHH*, or both at d14.

(N) Pie charts indicating the percentages of *ARX*$^+$/*SHH*$^+$/*FOXA2*$^+$ floor plate cells in *3A*-KO and WT cultures.

(O) Representative ICC images of WT and *3A*-KO cultures for FOXA2$^+$ (green) and TUJ1$^+$ (red) cells. Nuclei were stained with DAPI (blue). Scale bar, 50 μm.

(P) Percentages of TUJ1$^+$ neurons and FOXA2$^+$ floor plate cells in WT and *3A*-KOs presented as means ± SEM, ANOVA *p* values 3.565e-41 (% FOXA2$^+$ cells), 1.462e-22 (%TUJ1$^+$ cells).

(Q) Ratio of ChAT/FOXA2 protein levels analyzed by western blot. GAPDH used as a loading control.

(R) Schematic showing the combination of TFs guiding distinct progenitor domains in the developing spinal cord. The dorso-ventral axis (D-V) is shown on the left. Sonic hedgehog (Shh) and BMP/TGFbeta signaling drive this patterning. FP, floor plate.

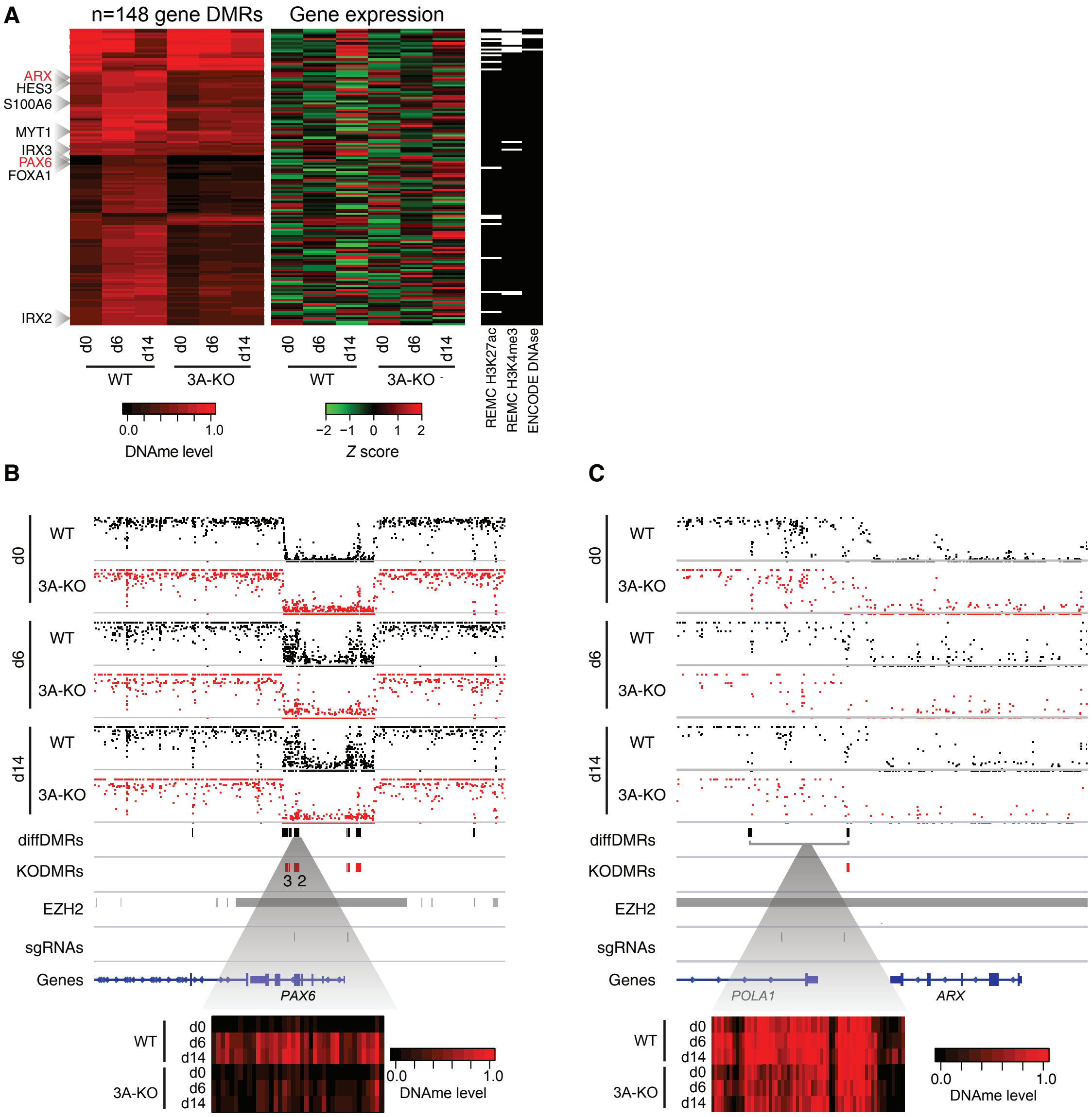


**Figure 5. *PAX6* and *ARX* Fail to Get Methylated Properly in the DNMT3A KO Cells**

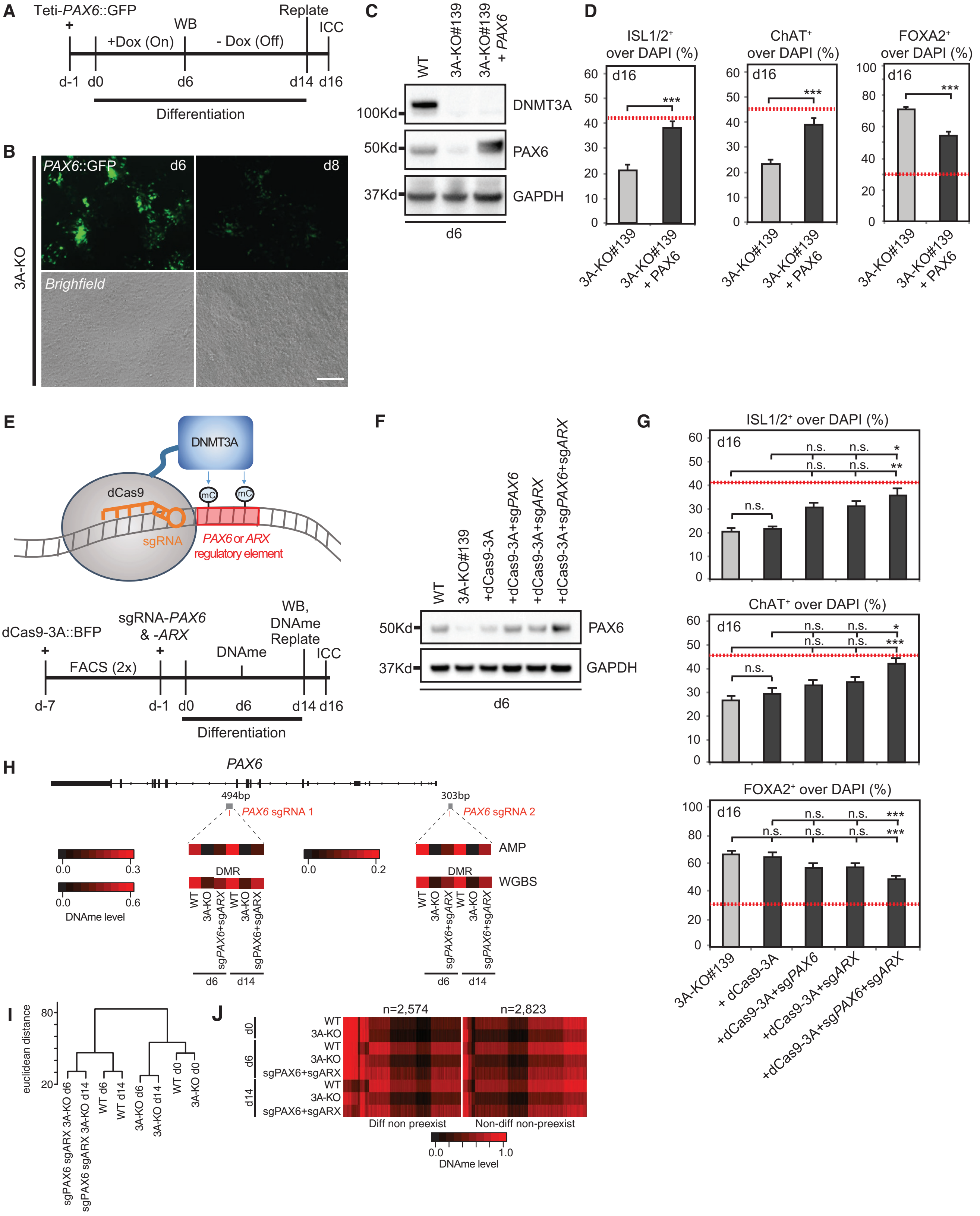
(A) Integrated analysis results of DMRs located within 100 kb of a gene differentially expressed between WT and *3A*-KO cells. Heatmaps of DNAme (left) and gene expression (middle) levels shown as *Z* scores. Selected TFs are highlighted on the left. If multiple DMRs are associated with the same gene, the average DNAme level was plotted. Right: associated gene expression and genomic feature annotation, black bars indicate DMR overlapping with an H3K27ac, H3K4me3, or DNase-seq peak in at least one of 70 distinct cell and tissue types profiled by the REMC or ENCODE consortium.

(B and C) High-resolution view of the (B) *PAX6* (chr11: 31,751,210–31,898,098) and the (C) *ARX* locus (chrX: 25,001,926–25,040,681) with DNAme levels shown as black (WT) and red (KO) dots. Each dot represents one CpG, with the position on the y axis indicating the DNAme level (bottom: 0%, top, 100%). Black boxes indicate DMRs arising during the WT differentiation process (diffDMRs); red boxes indicate DMRs identified DMRs between WT and KO. Gray bars in the EZH2 track indicate EZH2 binding sites in ESC; gray bars in sgRNA track indicate target sites of sgRNAs. Gray box and heatmap shows the DNAme level of all CpGs covered with at least 5 reads in the indicated DMR across all conditions.

between cellular identity and nuclear size and found that FOXA2$^+$ cells exhibit significantly larger nuclei than TUJ1$^+$ neurons, illuminating our early observation that *3A-KO* cultures contained more cells with prominent nuclei (Figures S2H–S2J). This difference largely contributes to the genotypic difference between *3A-KO*s and controls, as the KOs have significantly more FOXA2$^+$ floor plate cells. Collectively, we find that in the absence of 3A-dependent *de novo* DNAme, there is a dramatic shift in cellular-fate determination from neurogenesis/motor neurogenesis, toward gliogenic progenitors and the generation of floor plate cells (Figure 4R).

## Direct Targets of DNMT3A Include Key Developmental TFs

To identify primary targets of 3A during MN differentiation, we cross-correlated DMRs with associated gene expression



(legend on next page)

changes in an integrative analysis (Figure 5A; Table S5). We identified a list of 148 genes that were significantly affected at the level of DNAme and gene expression. The majority represented genes that were gradually methylated (80.4%) and downregulated in the WT case but failed to gain DNAme and maintained considerable expression levels in the absence of 3A (64.1%, Figure 5A, left). The associated DMRs highly overlap with putative GREs defined by H3K4me3 and H3K27ac positive regions (Figure 5A, right). These GREs are likely direct targets for *de novo* DNAme by 3A and affect the expression of ''primary effectors'' that are responsible for the phenotypic changes. Among these, we identified the master neurogenic factor *PAX6* (Ericson et al., 1997), and *ARX*, that has a known regulatory role in the spinal cord floor plate (Cho et al., 2014). We used our WGBS data to find a set of particular DMRs that fail to gain DNAme in the later stages of the KO cellular population (Figures 5B and 5C). In the case of *ARX*, lack of DNAme correlates with sustained high mRNA expression, which might be central to the formation of floor plate. In the case of *PAX6*, lack of DNAme correlates with a lack of sustained expression, which might be mediating defective neurogenesis. Of note, both sites are bound by EZH2 in human ESCs, which in the case of *PAX6* may need to be displaced in order to induce expression. To interrogate whether the *PAX6* DMRs could act as GREs, we sub-cloned them in tandem with a minimal promoter and a luciferase reporter using a CpG-free vector system. We evaluated the capacity of this potential GRE to regulate expression in a DNAme-sensitive fashion and found a strong DNAme-dependent activity with no activity in the methylated state (Figure 5B, DMRs 2 and 3, Figure S5). These data indicate that the *PAX6* DMR has the capacity to operate as a GRE that is repressed by EZH2 in ESCs and subsequently DNA methylated during MN differentiation.

### Expression of PAX6 and Targeted Methylation of the PAX6 and ARX Loci Rescue Differentiation Defects in DNMT3A KO ESCs

Given the well-established and critical role of PAX6 in early neurogenesis (Ericson et al., 1997), we interrogated whether its re-expression would lead to a phenotypic rescue. To recapitulate the endogenous temporal expression pattern of the gene, we used a lentiviral plasmid driven by a reverse tetracycline-controlled transactivator (rtTA). After optimizing the lentiviral dose and timing (Figures S6A–S6E), we infected *3A-KO* ESCs with Tet-On-PAX6::GFP and treated the cultures with doxycycline for the first 6 days of the differentiation (Figure 6A). We monitored PAX6 protein levels by WB and GFP fluorescence and found that PAX6 was effectively induced by day 2 and almost completely disappeared 2 days after the removal of doxycycline (Figures 6B, 6C, and S6E). This led to a significant, but not complete rescue of motor neurogenesis and repression of floor plate induction, as measured by ICC on day 16. We specifically found increased numbers of ISL$^+$ and ChAT$^+$ MNs, as well as fewer FOXA2$^+$ cells, although these fell short of the levels in the parental isogenic control (Figure 6D).

In order to directly assess the role of DNAme and the significance of the DMRs at the *PAX6* and *ARX* loci, we performed targeted DNAme editing using a protein fusion between a catalytically inactive dCas9 and 3A, in combination with single guide RNAs (sgRNAs) (Liu et al., 2016) (Figure 6E, top). We infected the 3A-KO ESCs with dCas9-3A-BFP lentivirus and generated a stable cell line by serial fluorescence-activated cell sorting (FACS) (Figure S6F). We then acutely infected this line with lenti-sgRNAs targeting *PAX6*, *ARX*, or both DMRs simultaneously and differentiated these into MNs (Figure 6E, bottom). Targeting the *PAX6* locus alone, or targeting both genes resulted in a prominent induction of PAX6 protein, to levels similar or higher to the ones observed in the WT control, respectively (Figure 6F). Unexpectedly, we find that the dCas9-3A alone, as well as targeting the *ARX* DMR alone, also resulted in some protein induction probably due to indirect or off-target effects (Figure 6F). We next asked whether this targeted DNAme restores the differentiation trajectory of the ESCs. We find that, while targeting either the *PAX6* or the *ARX* loci alone did not yield a complete rescue, simultaneous introduction of guides for both loci has a significant effect on motor neurogenesis and repression of floor plate induction, as measured by ICC on day 16. We specifically find increased generation of ISL$^+$ and ChAT$^+$ MNs, as well as a decreased percentage of FOXA2$^+$ cells, similar to what is typically observed with the parental WT control (Figure 6G). To validate that this effect was mediated by the restoration of DNAme, we performed targeted DNAme analysis. In the presence of both guides, the condition where we observed the strongest functional rescue, the DNAme levels are restored to levels that are close to the ones observed in WT cells (Figures 6H and S6G).

**Figure 6. Expression of PAX6 and Targeted Methylation of the *PAX6* and *ARX* Loci Rescue Differentiation Defects in DNMT3A KO ESCs**

(A) Experimental schematic: PAX6::GFP was re-expressed from d0 to d6 (mimicking PAX6 in WT cells) using a Doxycycline-inducible system in *3A*-KO cultures.

(B) Top panels show green fluorescent live cell images, and bottom panels show bright-field images of d6 and d8 MN cultures infected with PAX6:GFP. Scale bar, 50 μm.

(C) Western blot of PAX6 protein in *3A*-KO cells infected with TETi-PAX6::GFP and rtTA lentivirus.

(D) Quantification of ISL$^+$, ChAT$^+$, and FOXA2$^+$ cells over (DAPI) in PAX6::GFP transduced and non-transduced *3A*-KO cells on d16. Red lines indicate the average percentage in WT cultures. Data are presented as means ± SEM; ANOVA (***$p < 0.001$).

(E) Top, illustration of the targeted DNAme editing strategy based on a catalytically inactive dCas9 fused to 3A. Bottom, experimental schematic.

(F) Western blot analysis of PAX6 in WT and *3A*-KO lines non-transduced and transduced with dCas9-3A-BFP and *PAX6* or *ARX*-sgRNA-RFP viruses. GAPDH used as a loading control.

(G) Quantification of ISL$^+$, ChAT$^+$, and FOXA2$^+$ cells over (DAPI). Red lines indicate the average percentage in WT cultures. Data are represented as means ± SEM; ANOVA (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$), n.s, not significant.

(H) *PAX6* loci. Top: gray bars indicate regions targeted by amplicon sequencing; red bars indicate binding sites of sgRNAs. Bottom: heatmaps indicate average DNAme level of each targeted DMR based on amplicon sequencing (AMP) and WGBS for the WT, *3A*-KO, and *3A*-KO with dCas9-3A and two sgRNAs against the *PAX6* and *ARX* DMRs on d6 and d14.

(I) Dendrogram based on hierarchical clustering of DNAme levels across all DMRs in WT and *3A*-KO differentiating cultures.

(J) DNAme levels based on WGBS of non-preexisting DMRs that arise during WT differentiation and are de-regulated in the *3A*-KO (left) and DMRs that arise exclusively in the *3A*-KO differentiation (right) for the WT, *3A*-KO, and the *3A*-KO together with dCas9-3A dual locus targeting (sg*PAX6*+sg*ARX*).

Given previously noted off-target activity (Galonska et al., 2018), we analyzed the genome-wide DNAme effects of the dCas9-3A. We performed WGBS in the cells infected with the dCas9-3A alone, and when it was co-expressed with both sets of *PAX6* and *ARX* guides. Global clustering analysis on the DNAme levels of all DMRs detected in the WT and KO differentiation reveals a clear separation between the KO samples on one side and the WT and dCas9 rescue condition on the other side (Figure 6I). However, subclustering of the WT and rescue cells by condition and not by day indicates that substantial differences remain. This observation is confirmed upon investigation of the key DMR sets arising between WT and *3A-KO* cells (Figure 6J). Surprisingly, almost all DMRs exhibit a substantial increase in DNAme, approaching WT levels, indicating partial rescue of the DNAme phenotype and consistent with the functional rescue results. We next evaluated whether or not this global DNAme level rescue at differentiation-associated DMRs is specific or rather a global off-target effect. To address this, we compared the DNAme level and DNAme level change at the *3A-KO* differentiation-associated DMR set with the change at regions that are hypomethylated in the WT at each time point, and hence any DNAme gain represents a true off-target effect (Figures S6H and S6I). While this analysis indicates the presence of off-target effects, the effect sizes in DNAme gain vary substantially. In case of general off-target effects, the average DNAme change is only between 0.06 (SD 0.16) at day 6, and 0.04 (SD 0.15) at day 14 (Figure S6I). In contrast, the DNAme level change at differentiation-associated DMRs amounts 0.19 (SD 0.14, day 6) and 0.21 (SD 0.14, day 14), indicating specific targeting of ectopic dCas9-3A to these regions (Figure S6I). In summary, these results show efficient targeted DNAme of the *PAX6* and *ARX* locus, concomitant with a global rescue of the DNAme landscape specifically at the *3A-KO* differentiation-associated DMR set. While the DNAme levels do not fully reach WT levels, the observed change in DNAme levels in the dual *PAX6* and *ARX* targeting condition is consistent with the functional rescue of the *3A-KO*-associated phenotype.

## DNMT3A KO Impairs MN Morphology and Function

While the efficiency of MN production was impaired in the *3A-KO* background, we still obtained ISL/TUJ1$^+$ postmitotic cells. Moreover, *3A* mRNA and protein are continuously present during postmitotic stages of MN maturation *in vitro* (days 14–40; Figures 1A, S1A, and S1B), and *3A* mRNA remains significantly expressed in mature FACS-purified MNs (Figure S1A). Integrated analysis of gene expression and DNAme data indicate that a large number of genes affected in the KOs are involved in neuronal function, including ion channels, synaptic proteins, and genes involved in neurotransmitter regulation (Figure S7A; Table S5). These observations prompted us to investigate whether *3A-KO*-derived MNs are physiologically normal. We initially analyzed MN survival using a live/dead assay and found that *3A-KO* MN cultures exhibited a slight (1.3-fold), but highly significant ($p < 0.005$) increase in the proportion of apoptotic cells relative to the WT (Figure S7B). We then analyzed the morphological characteristics of WT and *3A-KO* neurons (Figure 7A). To minimize variability, we used a lentiviral GFP reporter driven by the *CamKII* promoter, which in these cultures labels mature MNs with high correspondence (>80% of GFP cells are ISL/CHAT$^+$ on day 40) (Figures S7C and S7D). While there were no differences in soma size, we found that KO MNs exhibited a significantly reduced length of neurites, as well as a strongly reduced complexity of neuronal branching (Figures 7A–7C). We also find reduced levels of PSD95, indicative of a reduced synaptic network in *3A-KO* cultures (Figure 7D).

Next, we evaluated the electrophysiological properties of the MN-enriched cultures using a multi-electrode array (MEA) platform to collect recordings from neuronal populations. We plated an equal number of differentiated KO and isogenic control cells and measured the pattern of spontaneous firing continuously through days 14–40 (Figures 7E and 7F). As neurons matured in both genotypes the mean firing rate increased significantly with time in culture ($p < 0.0001$). While there was difference in the number of active electrodes, *3A-KO* cultures exhibited pronounced hyper-excitability with a significantly higher mean firing rate ($p = 0.0008$). At the same time, KO neurons displayed prominent bursting, with a significantly extended duration for each burst ($p < 0.0001$). To determine whether these population-based physiological differences arise as a result of intrinsic defects in *3A-KO* MNs, we performed whole-cell current-clamp recordings. We analyzed at least $n = 59$ *CamKII*-GFP$^+$ cells per genotype from three independent differentiations, after 40 days in culture (Figures 7G–7J; Table S6). MNs from both genotypes had stable resting membrane potentials and were capable of robust bouts of activity including repetitive firing during depolarizing current steps and ramps. While both displayed prominent spike frequency adaptation in response to depolarizing current steps, *3A-KO* MNs were not able to achieve the same action potential (AP) firing frequencies as their control counterparts (Figure 7G). They exhibited a reduced maximum initial rate of firing (top dotted line, $p = 0.009$), and a reduced steady-state firing frequency (bottom dotted line, $p < 0.001$). They were also hyper-excitable, as APs were elicited with significantly less depolarizing current ($p = 0.008$) (Figure 7H). The persistent inward current (PIC), a voltage dependent inward Na$^+$ current that is well-characterized in MNs, was activated at a significantly more hyperpolarized voltage ($p = 0.01$) suggesting a shift in Na$^+$ channel activation parameters (Figure 7I) (Kuo et al., 2006). *3A-KO* MNs also exhibited different AP profiles, with a faster rate of rise and a larger AP size (measured as mV from threshold to peak), ($p < 0.001$), indicating altered Na$^+$ channel activity (Figure 7J).

## DISCUSSION

DNAme is considered a critical process in regulating lineage choice and commitment. Here we present a high-resolution molecular, phenotypic, and functional characterization of the role that *de novo* DNAme plays during the *in vitro* differentiation of human ESCs into a major neuronal subtype of the CNS, the spinal MN. We found that loss of *3A*, but not *3B*, impairs the ability of ESCs to efficiently generate MNs. This defect was caused by focal changes in DNAme patterns affecting a discrete number of uncharacterized regions that are normally targeted for DNAme during differentiation, and in particular GREs such as putative enhancers and CpG islands. Our findings suggest that DNAme plays a highly coordinated role in delineating neural/neuronal diversity and function during the development of spinal cord cellular lineages.

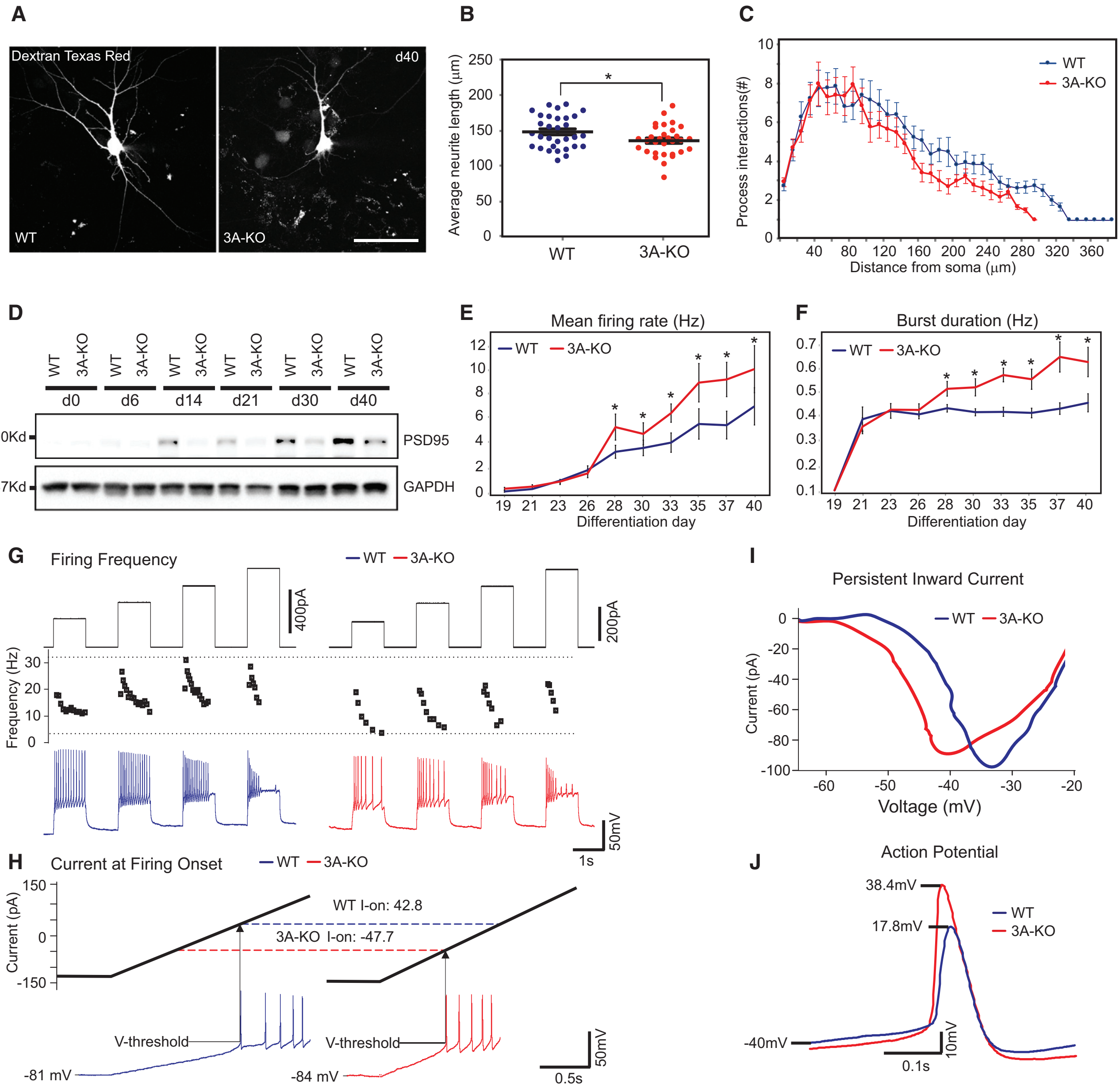


**Figure 7. Functional Defects in DNMT3A KO MNs**

(A) 2-photon microscopy images of d40 MNs derived from *3A*-KO and isogenic control ESCs. MNs are labeled by Dextran Texas Red for morphological and electrophysiological studies. Scale bar, 50 µm.

(B) Analysis of average neurite length in d40 MNs. Data are means of independent differentiations (n = 4, m = 31 ± SEM), 2-tailed t test (p = 0.0298).

(C) Sholl analysis of d40 MNs. Data are mean of independent differentiations (n = 4, m = 31 ± SEM), Mann-Whitney test (p < 0.05).

(D) Western blot analysis of PSD95 in d0–d40 *3A*-KO and WT MN cultures. GAPDH used as a loading control.

(E and F) MEA-based activity analysis of the mean firing rate and the average burst duration in *3A*-KO and isogenic control cultures over 40 days *in vitro*. Data are mean of independent differentiations (n = 3, m = 12 ± SEM); 2-way ANOVA (E, p = 0.0008; F, p < 0.0001).

(G) Activity of a typical WT (left panel, blue trace) and *3A*-KO (right panel, red trace) MN recorded in current clamp are shown below depolarizing current steps that evoked activity (top traces) and the frequency plot of the evoked action potentials (middle traces). Scale bar in control current trace, 400 pA. Scale bar in KO current trace, 200 pA. Scale bars for all voltage traces: horizontal, 1 s; vertical, 50 mV.

(H) Current at firing onset in WT and *3A*-KO MNs. Typical voltage threshold of action potentials and corresponding I-on values are shown for a WT (blue trace, left panel) and *3A*-KO (red trace, right panel) MNs. Horizontal scale bar, 0.5 s; vertical, 50 mV apply to both traces.

(I) Persistent inward current in KO cell line (red trace) was evoked at a significantly hyperpolarized voltage compared to the WT (blue trace). The average onset voltage (voltage at which the downward deflection in the current traces is first initiated) was −55 mV in KO and −51 in WT, similar to the examples shown.

(J) The AP height (size) was greater in KO respect to WT MNs (example trace in red). This was driven by a faster rate of rise in the upswing. The rate of fall was unchanged. Scale bar vertical, 10 mV; horizontal, 0.1 s. See Table S6 for measurements and statistics.

The fact that in the 3A-KO cells some regions still gain DNAme raises the possibility that these DMRs are specific targets of 3B. Indeed, our attempts to differentiate double-3A/3B-KO cells resulted in dramatic apoptosis suggesting that 3B-mediated DNAme that takes place during the initial differentiation stages (when the protein is still present) is critical. On the other hand, while functional deletion of 3B had pronounced effects at the global DNAme level, it resulted in very few changes at the gene expression level during MN differentiation. Moreover, the overlap between differentially expressed genes in *3A* and *3B-KO* cells during the differentiation was minimal. These results are in alignment with the contrasting expression profiles of these two enzymes during MN differentiation and suggest that they have distinct roles, and very minimal overlap in this context as previously shown in other lineages (Challen et al., 2014).

Integrative molecular analysis identified multiple TFs that are differentially regulated in the differentiating KO cultures. In particular, the lack of *PAX6* induction coupled with the sustained expression of *ARX* might be responsible for shifting the neurogenic fate of *3A-KO* cells toward the induction of floor plate. This suggests that DNAme and 3A specifically play a critical role during the patterning of differentiating ESCs toward spinal cord lineages, and under the context of this protocol, in the specification of MNs. This would be in line with the role that DNAme frequently plays during cellular differentiation, in repressing default cell fates, as well as the enabling of appropriate TF binding to allow cell specification and determination (Bröske et al., 2009; Dhawan et al., 2011; Feng et al., 2010; Trowbridge et al., 2009).

While our experiments demonstrate the functional relevance of DNAme changes at specific loci, they also raise important new questions. In particular, we observe not only significant DNAme rescue at the targeted genomic regions, but also at most *3A-KO*-induced DMRs, specifically in the dual *PAX6/ARX* targeting experiments. Since the WT 3A protein is absent, at least two mechanisms mediating this phenomenon are possible: given that most of the DMRs rescued in the dCas9-3A condition gain DNAme by the first 6 days of differentiation, it is possible that the efficient induction of *PAX6* and repression of *ARX* triggers compensatory mechanisms. These may lead to the recruitment of 3B to these DMRs during the first 2 days of differentiation, while the protein is still expressed. Alternatively, a transcriptional cascade that is only activated by simultaneous targeting of *PAX6* and *ARX* may induce recruitment of the entire dCas9-3A complex specifically to the *3A-KO* DMRs.

Several of the DMRs we identified, such as the *ARX* loci show concordance between reduced DNAme levels and increased expression. In contrast, the GREs within the *PAX6* locus fail to gain DNAme, coinciding with reduced PAX6 expression, while targeted DNAme resulted in protein induction. This suggests a dual role of this specific *PAX6* regulatory element. While the GRE itself can operate as an enhancer in other cellular contexts (supported by VISTA *in vivo* enhancer data), it is bound by the Polycomb complex in human ESCs, similarly to the rest of the *PAX6* locus, which exhibits a bivalent chromatin architecture (Ku et al., 2008). Thus, our observations suggest that the recruitment of 3A and the gain of DNAme during differentiation might block the binding of a repressor, or is required for efficient repressor displacement at the *PAX6* locus, supporting increased expression levels. In the future, it will be interesting to dissect the specific opposing roles of DNAme at the individual locus level, its interplay with other layers of epigenetic regulation, as well as the broader relevance of this regulatory mechanism in cellular differentiation.

Despite the observed reduction in neuronal progenitors and MNs, *3A-KO* ESCs consistently generated MNs. Thus, it is worth considering that the germline and conditional KO mice must also be able to generate MNs (Nguyen et al., 2007; Okano et al., 1999). Indeed, mice devoid of *Dnmt3a* in the nervous system are born healthy but die prematurely, exhibiting motor movement defects. In accordance with this *in vivo* evidence, we also find that the human *in-vitro*-derived *3A-KO* MNs are physiologically defective. These findings are consistent with the increased incidence of *3A* mutations in individuals with autism (Sanders et al., 2015), and the neurodevelopmental disorder Tatton Brown Rahman Syndrome (TBRS) (Tatton-Brown et al., 2014). Future studies can illuminate whether 3A plays overlapping or distinct roles in discrete CNS subtypes such as MNs and cortical neurons. Whether these physiological defects occur because of improperly configured DNAme patterns during the differentiation stage, or rather as a consequence of a lack of *de novo* DNAme during the maturation of postmitotic MNs, remains to be elucidated.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Stem cell culture
  - MN differentiation
- METHOD DETAILS
  - RNA preparation and qRT-PCR
  - Preparation and sequencing of WGBS libraries
  - Amplicon sequencing
  - Luciferase activity assay
  - Preparation and sequencing of RNA-Seq libraries
  - Single-cell RNA-sequencing
  - Immunocytochemistry and image analysis
  - Western blot analysis
  - FACS and NCAM analysis
  - Cell viability assays
  - Quantification of acetylcholine
  - Electrophysiological recordings
  - Characterization of neural progenitor cells
  - Rescue of PAX6 expression
  - Targeted DNAme using dCas9-DNMT3A
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - WGBS data processing
  - Identification of DMRs
  - DMR gene set enrichment analysis (GSEA)
  - Integrated analysis
  - Annotation with genomic features
  - RNA-Sequencing data processing and analysis
  - Single-cell RNA-Seq data processing and analysis
  - Statistical testing
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.stem.2018.02.012.

## ACKNOWLEDGMENTS

The pLV-TETi-PAX6-2A-GFP plasmid is a gift of Danwei Huangfu. The Kiskinis lab acknowledges support by the Les Turner ALS Foundation, the Muscular Dystrophy Association, and institutional funds from Northwestern Feinberg School of Medicine; M.J.Z., by the BMBF eMed (01ZX1504) and the Max Planck Society; K.A.Q., by Target ALS; and A.M., by the New York Stem Cell Foundation, NIH grants (1P50HG006193, P01GM099117, R01DA036898), and the Max Planck Society. A.M. is a New York Stem Cell Foundation Robertson Investigator. E.K. is a Les Turner ALS Research and Patient Center Investigator.

## AUTHOR CONTRIBUTIONS

M.J.Z. conceptualized the project, performed all bioinformatics analysis, interpreted the data, and wrote the manuscript. J.A.O. performed differentiations; generated RNA and DNA for molecular analysis, ICC, WBs, single-cell RNA-seq, morphological analysis, and rescue experiments; interpreted data; and wrote the manuscript. K.A.Q. performed electrophysiology experiments. D.P.S. performed differentiations, generated RNA and DNA for molecular analysis, and performed MEA and qPCR experiments. E.J.M. performed rescue experiments. A.D.P., M.H., and H.Y.M. performed Ach analysis. C.G. and R.P. generated the RNA-seq libraries. S.M. performed reporter assays. H.G. and A.G. generated the WGBS libraries. C.J.H. interpreted electrophysiology experiments. A.M. and E.K. conceptualized the project, designed experiments, interpreted the data, provided guidance and supervision, and wrote the manuscript. All authors helped to edit the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2017
Revised: December 18, 2017
Accepted: February 23, 2018
Published: March 15, 2018

## REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate—A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bröske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat. Genet. *41*, 1207–1215.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., et al. (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. Cell Stem Cell *15*, 350–364.

Chen, A.E., Egli, D., Niakan, K., Deng, J., Akutsu, H., Yamaki, M., Cowan, C., Fitz-Gerald, C., Zhang, K., Melton, D.A., and Eggan, K. (2009). Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. Cell Stem Cell *4*, 103–106.

Cho, G., Lim, Y., Cho, I.T., Simonet, J.C., and Golden, J.A. (2014). Arx together with FoxA2, regulates Shh floor plate expression. Dev. Biol. *393*, 137–148.

Dhawan, S., Georgia, S., Tschen, S.I., Fan, G., and Bhushan, A. (2011). Pancreatic β cell identity is maintained by DNA methylation-mediated repression of Arx. Dev. Cell *20*, 419–429.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., Jessell, T.M., and Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. Cell *90*, 169–180.

Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. J. Neurosci. Res. *79*, 734–746.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat. Neurosci. *13*, 423–430.

Galonska, C., Charlton, J., Mattei, A.L., Donaghey, J., Clement, K., Gu, H., Mohammad, A.W., Stamenova, E.K., Cacchiarelli, D., Klages, S., et al. (2018). Genome-wide tracking of dCas9-methyltransferase footprints. Nat. Commun. *9*, 597.

Jeltsch, A. (2006). Molecular enzymology of mammalian DNA methyltransferases. Curr. Top. Microbiol. Immunol. *301*, 203–225.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Kuo, J.J., Lee, R.H., Zhang, L., and Heckman, C.J. (2006). Essential role of the persistent sodium current in spike initiation during slowly rising inputs in mouse spinal neurones. J. Physiol. *574*, 819–834.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA methylation in the mammalian genome. Cell *167*, 233–247.

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

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., and Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. Dev. Dyn. *236*, 1663–1676.

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

Park, Y., and Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics *32*, 1446–1453.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al.; Autism Sequencing Consortium (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron *87*, 1215–1233.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: An open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: Roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Sugimori, M., Nagao, M., Bertrand, N., Parras, C.M., Guillemot, F., and Nakafuku, M. (2007). Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord. Development *134*, 1617–1629.

Sun, D., Xi, Y., Rodriguez, B., Park, H.J., Tong, P., Meong, M., Goodell, M.A., and Li, W. (2014). MOABS: Model based analysis of bisulfite sequencing data. Genome Biol. *15*, R38.

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al.; Childhood Overgrowth Consortium (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat. Genet. *46*, 385–388.

Topol, A., Tran, N.N., and Brennand, K.J. (2015). A guide to generating and using hiPSC derived NPCs for the study of neurological diseases. J. Vis. Exp. Published online February 21, 2015. https://doi.org/10.3791/52495.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449.

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature *518*, 344–349.

Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat. Genet. *50*, 83–95.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science *329*, 444–448.

Xi, Y., and Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron *78*, 785–798.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481.

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L., and Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J. Virol. *72*, 9873–9880.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| ACTIN | Santa Cruz Biotechnology | Cat# sc-1615; RRID:AB_630835 |
| Alexa Fluor Secondary antibodies | ThermoFisher Scientific | https://www.thermofisher.com/us/en/home/life-science/antibodies/secondary-antibodies/fluorescent-secondary-antibodies/alexa-fluor-secondary-antibodies.html |
| ChAT | Millipore | Cat# AB144; RRID:AB_90650 |
| DNMT3A | Cell Signaling Technology | Cat# 2160S; RRID:AB_2263617 |
| DNMT3B | Santa Cruz Biotechnology | Cat# sc-20704; RRID:AB_2094125 |
| FOXA2 | Santa Cruz Biotechnology | Cat# sc-101060; RRID:AB_1124660 |
| GAPDH | Cell Signaling Technology | Cat# 2118; RRID:AB_561053 |
| HES1 | Abcam | Cat# ab49170; RRID:AB_2117856 |
| HRP Donkey Anti-goat | Santa Cruz Biotechnology | Cat# sc-2020; RRID:AB_631728 |
| HRP Goat Anti-mouse | LI-COR | Cat#:926-80010; RRID:AB_2721263 |
| HRP Goat Anti-rabbit | LI-COR | Cat#:926-80011; RRID:AB_2721264 |
| ISL1 | DHSB | Cat# 40.2D6; RRID: AB_528315 |
| ISL1/2 | DHSB | Cat# 39.4D5; RRID: AB_2721271 |
| Ki67 | ThermoFisher Scientific | Cat# RM-9106, RRID:AB_2335745 |
| Alexa Fluor 647 Mouse Anti-Human CD56 (NCAM) | BD Biosciences | Cat# 557711; RRID:AB_396820 |
| NESTIN | Abcam | Cat# ab92391; RRID:AB_10561437 |
| OLIG2 | Millipore | Cat# AB9610; RRID:AB_570666 |
| PAX6 | DHSB | Cat# AB_528427; RRID:AB_528427 |
| PH3 | Cell Signaling Technology | Cat# 9701; RRID:AB_331535 |
| PSD95 (7E3) | Cell Signaling Technology | Cat# 36233; RRID:AB_2721262 |
| TUJ1 | BioLegend | Cat# 801201; RRID:AB_2313773 |
| TUJ1 | Sigma-Aldrich | Cat# T2200; RRID:AB_262133 |
| SOX2 | Abcam | Cat# ab79351, RRID:AB_10710406 |
| Bacterial and Virus Strains | | |
| One Shot TOP10 Chemically Competent *E. coli* | ThermoFisher Scientific | Cat# C404010 |
| HIV-1-based lentivirus | Northwestern University DNA RNA Delivery Core | Zufferey et al., 1998 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Ascorbic Acid | Sigma-Aldrich | Cat# A4403 |
| BDNF | R&D Systems | Cat# 248-BD |
| CNTF | R&D Systems | Cat# 257NT/CF |
| DAPT | DNSK International | Cat# 208255-80-5 |
| Doxycycline | Sigma-Aldrich | Cat# D9891 |
| GDNF | R&D Systems | Cat# 212-GD |
| Laminin | Life Technologies | Cat# 23017-015 |
| LDN-193189 | DNSK International | Cat# 1062368-24-4 |
| Matrigel | CORNING | Cat# BD354277 |
| Poly-D-Lysine | Sigma-Aldrich | Cat# P6407 |
| Retinoic Acid | Sigma-Aldrich | Cat# R2625 |
| ROCK inhibitor (Y-27632) | DNSK International | Cat# 129830-38-2 |
| SB431542 | DNSK International | Cat# 301836-41-9 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Smoothened-Agonist (SAG) | DNSK International | Cat# 364590-63-6 |
| SU5402 | DNSK International | Cat# 215542-92-3 |
| Critical Commercial Assays | | |
| TRIzol | ThermoFisher Scientific | Cat# 15596026 |
| Script Reverse Transcription Supermix | BIO-RAD | Cat# 1708841 |
| iTaq Universal SYBR Green Supermix | BIO-RAD | Cat# 172-5124 |
| Direct-zol RNA MiniPrep | Zymo Research | Cat# R2051 |
| TruSeq RNA Library Preparation Kits | Illumina | Cat# RS-122-2001 |
| DNA Clean & Concentrator kit | Zymo Research | Cat# D4005 |
| EZ DNAme-Gold kit | Zymo Research | Cat# D5005 |
| Accel-NGS Methyl-Seq DNA library kit | Swift Biosciences | Cat# 30096 |
| TaKaRa EpiTaq HS | Clontech | Cat# R110B |
| Chromium Single Cell 3′ Library & Gel Bead Kit v2 | 10x genomics | Cat# PN-120267 |
| Chromium Single Cell A Chip Kit | 10x genomics | Cat# PN-1000009 |
| Chromium i7 Multiplex Kit | 10x genomics | Cat# PN-120262 |
| CytoTox 96 Non-Radioactive Cytotoxicity Assay | Promega | Cat# G1780 |
| LIVE/DEAD assay | ThermoFisher Scientific | Cat# L3224 |
| Deposited Data | | |
| Raw and processed data | GEO: GSE90553 | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE90553 |
| Analyzed data | Tables S1, S2, S3, S4, S5, S6, and S7 | N/A |
| Experimental Models: Cell Lines | | |
| HUES-64 cell line | Harvard University | RRID:CVCL_B199 |
| SH-SY5Y cell line | ATCC | Cat# ATCC CRL-2266; RRID:CVCL_0019 |
| Oligonucleotides | | |
| Primers for qPCR, See Table S7 | This paper | N/A |
| Primers for Amplicon sequencing, See Table S7 | This paper | N/A |
| Guide RNAs for targeted DNAme, See Table S7 | This paper | N/A |
| GRE sequences for reporter assay, See Table S7 | This paper | N/A |
| Recombinant DNA | | |
| Fuw-dCas9-DNMT3A-P2A-BFP | Addgene | Cat# 84569 |
| pgRNA-humanized-mCherry | Addgene | Cat# 44248 |
| pLV-TETi-PAX6-2AGFP | (Verma et al., 2018) | N/A |
| FUW-rtTA | (Zhang et al., 2013) | N/A |
| pCpGfree-promoter-Lucia | InvivoGen | Cat# pcpgf-promlc |
| Software and Algorithms | | |
| GraphPad Prism 7 | GraphPad Software | http://www.graphpad.com/scientific-software/prism/ |
| FlowJo v10 | FlowJo LLC. | https://www.flowjo.com/ |
| Fiji | (Schindelin et al., 2012) | https://imagej.net/Fiji |
| MetaMorph | Molecular Devices | https://www.moleculardevices.com/systems/metamorph-research-imaging/metamorph-microscopy-automation-and-image-analysis-software |
| Adobe Illustrator CC 2015 | Adobe | Version 16.0.0 |

(*Continued on next page*)

**Continued**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Adobe Photoshop CC 2015 | Adobe | Version 19.0.0 |
| bsMap version 2.7 | (Xi and Li, 2009) | https://code.google.com/archive/p/bsmap/ |
| MOABS version xx | (Sun et al., 2014) | https://github.com/sunnyisgalaxy/moabs |
| DSS version 2.14.0 | (Park and Wu, 2016) | https://bioconductor.org/packages/release/bioc/html/DSS.html |
| methylKit version 0.9.5 | (Akalin et al., 2012) | https://bioconductor.org/packages/release/bioc/html/methylKit.html |
| GREAT version 3.0.0 | (McLean et al., 2010) | http://great.stanford.edu/public/html/index.php |
| Trimmomatic version 0.33 | (Bolger et al., 2014) | http://www.usadellab.org/cms/?page=trimmomatic |
| Salmon version 0.8.2 | (Patro et al., 2017) | https://combine-lab.github.io/salmon/ |
| DESeq2 version 1.14.1 | (Love et al., 2014) | https://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| Seurat version 2.0.0 | (Satija et al., 2015) | https://github.com/satijalab/seurat |
| R version 3.3.3 | | https://www.r-project.org/ |
| Tximport version 1.4.0 | (Soneson et al., 2015) | https://bioconductor.org/packages/release/bioc/html/tximport.html |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact, Evangelos Kiskinis (evangelos.kiskinis@northwestern.edu).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Stem cell culture

Human embryonic stem cells (hESC), HUES-64 (Harvard University) and HUES-64 mutated lines (Liao et al., 2015), were maintained on Matrigel (BD Biosciences) with mTeSR1 media (Stem Cell Technologies) and passaged on a weekly basis using 1mM EDTA or Accutase (Sigma). All cell cultures were maintained at 37°C and 5% $CO_2$. All ESC/iPSC cultures are tested on a monthly basis for mycoplasma. All lines used were mycoplasma-free.

### MN differentiation

hESC colonies were dissociated using Accutase and plated at a density of 74,000 cells/$cm^2$ with 10$\mu$M ROCK inhibitor (Y-27632, DNSK International) in mTeSR1 for 24 hours. Media was replaced on day 0 with N2B27 medium (50% DMEM:F12, 50% Neurobasal, supplemented with NEAA, Glutamax, N2 and B27; GIBCO, Life Technologies) containing 10$\mu$M SB431542 (DNSK International), 100nM LDN-193189 (DNSK International), 1 $\mu$M Retinoic Acid (RA, Sigma) and 1$\mu$M of Smoothened-Agonist (SAG, DNSK International). Culture medium was changed daily for 6 days and then was switched to N2B27 medium supplemented with 1$\mu$M RA, 1$\mu$M SAG, 5$\mu$M DAPT (DNSK International) and 4$\mu$M SU5402 (DNSK International). Cells were fed daily until day 14 of differentiation. On day 14, MNs were dissociated using TrypLE Express (GIBCO, Life Technologies) supplemented with DNase I (Worthington) and plated onto pre-coated Poly-D-Lysine/Laminin surfaces (BD Biosciences) and cultured in Neurobasal medium supplemented with NEAA, Glutamax, N2, B27, Ascorbic acid (0.2 $\mu$g/ml; Sigma-Aldrich) and BDNF, CNTF and GDNF (10ng/mL, R&D systems).

## METHOD DETAILS

### RNA preparation and qRT-PCR

Cells were harvested by scraping from 24-well plates at the relevant time points of the MN differentiation. Pellets were resuspended in TRIzol Reagent (Life Technologies) and RNA was isolated following manufacturers protocol. 0.5 to 1 $\mu$g of RNA was treated with DNase I (Invitrogen) and subsequently used for the generation of cDNA using iSCRIPT Reverse Transcription Supermix (Bio-Rad) following manufacturer's instructions. RT-PCR was performed using iTaq Universal SYBR® Green Supermix on the CFX system (Bio-Rad). All assays were performed in triplicates. The 5′ → 3′ sequences of the forward (F) and reverse (R) primers used in this study can be found in Table S7. The averaged cycle of threshold (Ct) value of two housekeeping genes (*ACTIN/GAPDH*) was subtracted from the Ct value of the gene of interest to obtain the $\Delta$Ct. Relative gene expression was determined as $^{2-}\Delta$Ct ($\Delta\Delta$Ct) and expressed relative to the control sample or the highest expressed sample in the experiment.

### Preparation and sequencing of WGBS libraries

100-200 ng of genomic DNA was fragmented using a Covaris S2 for 6 min according to the following program: duty cycle 5%; intensity 10; cycle per burst 200. The sheared DNA was purified using the DNA Clean & Concentrator kit from Zymo Research per the manufacturer's recommendations. Bisulfite conversion of DNA was then conducted using the EZ DNAme-Gold kit (Zymo Research) followed the instructions, eluting to 15 μl low TE buffer. To minimize continued loss during storage, bisulfite-converted DNA was immediately processed for generating WGBS libraries using the Accel-NGS Methyl-Seq DNA library kit (Swift Biosciences) following the manufacturer's protocol. The libraries were sequenced on an Illumina HiSeq 2500 sequencer as 100-bp paired-end reads.

### Amplicon sequencing

Bisulfite conversion of DNA was conducted using the EZ DNAme-Gold kit (Zymo Research) following manufacturer's instructions. Individual amplicons were PCR amplified using TaKaRa EpiTaq HS (Clontech). The PAX6 and ARX amplicons were amplified using the primer pairs listed in Table S7. Amplicons were sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (300-cycles). The genomic coordinates for each amplicon are: PAX6 Amplicon 1: chr11: 31,821,416-31,821,910; PAX6 Amplicon 2: chr11: 31,840,497-31,840,800; ARX Amplicon 1: chrX: 25,011,817-25,012,198; ARX Amplicon 2: chrX: 25,017,310-25,017,775.

### Luciferase activity assay

PAX6 GRE fragments (Region 2:chr11:31,818,705-31,820,467, Region 3:chr11:31,840,673-31,841,636) (Table S7) were synthesized by GeneScript and cloned into the pCpGfree-promoter-Lucia (pcpgf-promlc / InvivoGen) vector upstream of a minimal promoter. Subsequently, the empty vector, as well as the two GRE containing vectors were methylated *in vitro* using SssI (M0226M / New England Biolabs) incubating the plasmids with SssI for 4h at 37°C, followed by 20 min on ice, followed by 20 min at 65°C and another 10min on ice. Proper methylation of the plasmids was tested by HpaII (R0171S / New England Biolabs) restriction. Subsequently, 2ug of unmethylated GFP control vector, the empty vector in a methylated and unmethylated version as well as each GRE methylated/unmethylated were each nucleofected into 800,000 SHSY5Ycells using a Amaxa Nucleofector II program G-004 and the Lonza Nucleofector Kit V (VCA-1003). Following nucleofection, cells were plated in 96 well plates. On the next day, 24 h later, the media was replaced with 100 μl media. Two days post nucleofection, the renilla activity was measured following manufacturers protocol (Quanti-Luc / rep-qlc1 / Invivogen). The renilla activity from the GRE elements was then normalized to a luciferase acitivity (BrightGlo Luciferase Assay / E2610 / Promega). Each experiment was performed in 3 independent biological replicates. For each replicate, four distinct luciferase and renilla readouts were conducted and averaged.

### Preparation and sequencing of RNA-Seq libraries

RNA was purified using the Direct-zol RNA MiniPrep (Zymo Research Inc) and RNA-seq libraries generated using the Illumina TruSeq RNA Library Preparation Kits. Pools were sequenced on the Illumina HiSeq 2500 sequencer as 100bp paired-end reads.

### Single-cell RNA-sequencing

WT and *3A*-KO cells were dissociated on days 6 and 14 of differentiation with Accutase for 10 min. Cells were collected in DMEM:F12 (day 6) or Neurobasal (day 14) + 2% FBS, and centrifuged in order to quench and remove the Accutase. Cell pellets were re-suspended in PBS + 0.04% Bovine Serum Albumin (Millipore) and filtered through a 40μm Nylon Mesh-Cell strainer (Fisher Scientific). Next, we quantified cell number and cell viability using the Nexcelom Cell Counter (Nexcelom Bioscience). To generate Cell 3′ Libraries we used a Single Cell 3′ Reagent Kit v2 (10xGenomics, Chromium). In brief, 10K cells/sample with the RT master mix were loaded in a Single Cell 3′ Chip along with the RT Single Cell 3′ Gel Beads and the Partitioning oil, and run in a Chromium Controller to generate Single Cell Gel Bead-In-Emulsions (GEMs). RT reaction of the resulting single cell-beads suspension was run in order to obtain 10x Barcoded cDNA. Silane magnetic beads were used to clean the barcoded cDNA, which was amplified by PCR to generate sufficient mass for library construction. Enzymatic Fragmentation, End Repair and A-tailing Double Sided Size Selection were used to incorporate R2, P5, P7, 10bp randomer as well as i7 index read sequences that results in Illumina-ready sequencing libraries. To QC post cDNA Amplification reaction and libraries, we loaded 1-2 μl on Us d5000 and d1000 tapes, respectively, and run in a Tapestation 4200 (Agilent Technologies). To quantify DNA concentrations in the libraries we utilized a Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen). For sequencing of Single Cell 3′ libraries, we used standard Illumina sequencing primers and were run using paired-end sequencing with dual indexing following manufacturer's recommendations. Subsequently each library was sequenced on a HiSeq4000 lane.

### Immunocytochemistry and image analysis

Cells were fixed with 4% paraformaldehyde and blocked for 1 h in PBS containing 10% normal donkey serum (Jackson ImmunoResearch) and 0.1% Triton (Sigma-Aldrich). Samples were then incubated overnight at 4°C with primary antibodies: ChAT (goat, 1:250, Millipore), FOXA2 (mouse, 1:500; Santa Cruz Biotechnology), ISL1 (1:250, Development Studies Hybridoma Bank, DSHB, University of Iowa), ISL1/2 (1:250, Development Studies Hybridoma Bank, DSHB, University of Iowa), Ki67 (1:500, Thermo Scientific), NESTIN (1:500, Abcam), PAX6 (1:500; DSHB), OLIG2 (1:500; Millipore), TUJ1 (1:1000; BioLegend), TUJ1 (1:1000; Sigma-Aldrich), SOX2 (1:500; Abcam). After several washes with PBS + 0.1% Triton, samples were incubated for 1 h with the appropriate secondary antibodies conjugated with Alexa488, Alexa555 or Alexa647 fluorophores (1:500, Molecular Probes).

Cell nuclei were stained using Hoechst (Life Technologies). Immunolabeled samples were imaged using a Leica DMi8 microscope (Leica, Buffalo Grove, IL), photographed using a C10600-ORCA-R2 digital CCD camera (Hamamatsu Photonics, Japan) and processed with MetaMorph software (Molecular Devices). The images were assembled in Adobe Photoshop CC, with consistent quality adjustments for contrast, brightness and color balance. For image analyses were performed from immunolabeled cells from 9–12 pre-designated adjacent optical fields, and from a minimum of three independent differentiations per experimental condition. Nuclear size, staining intensity as well as nuclei and distinct cell types countings were analyzed using MetaMorph (Molecular Devices) and Fiji software (ImageJ, NIH Image). Neuronal morphology analysis was performed with the Fiji software on images from neurons labeled with Texas Dextran Red 3000MW captured with an Olympus BX-51WIF microscope fitted with an Olympus 40x/0.8NA water-dipping objective lens. Maximum-reconstructions were calibrated and adjusted for brightness and contrast for a subsequent semi-automatic tracing process using Simple Neurite Tracer (SNT) from Fiji. This tool allows for the skeletonization of all neuronal processes and measurement of neurite length. These traced-processes images were used to analyze the mophological complexity of the neurons by the Fiji plugin Sholl analysis, which creates a series of concentric circles around the neuron cell body, and calculates the number of neuronal processes crossing the different circles. The graphical representation of these data indicates the branching complexity along the neuronal arbor of one specific neuron.

### Western blot analysis

The cells were homogenized in RIPA buffer (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1mM EDTA, 0.1% SDS. 1% Triton X-100, Sigma Aldrich) with protease inhibitors (Millipore) and phosphatase inhibitors (Abcam) cocktail. Cell homogenates were kept on ice for 30 min and then sonicated to completely disrupt the cell membranes and DNA. Protein extracts were separated by SDS-PAGE at 120 V for 90 min and electro-transferred at 100V for 60 min to a nitrocellulose membrane (Bio-Rad), which were subsequently blocked and incubated overnight at 4°C with primary antibodies: ACTIN (1:1000; Santa Cruz Biotechnology), ChAT (1:250; Millipore), DNMT3A (1:1000; Cell Signaling Technology), DNMT3B (1:500; Santa Cruz Biotechnology), FOXA2 (1:500; Santa Cruz Biotechnology), GAPDH (1:1000; Cell Signaling), HES1 (1:500; Abcam), PAX6 (1:1000; DSHB), PH3 (1:1000; Cell Signaling Technology), PSD95 (1:1000; Cell Signaling Technology), TUJ1 (1:1000; BioLegend). After several washes with Tris-buffered saline (TBS, 50mM Tris, 150mM NaCl, HCl to pH 7.6) 0.1% Tween (Sigma-Aldrich), membranes were incubated with their corresponding secondary HRP-conjugated antibodies. Protein signals were detected by a ChemiDoc$^{TM}$ XRS+ (Bio-Rad), using the SuperSignal West Pico chemilluminescent system (Thermo Fisher Scientific). Densitometry analysis of the bands from a minimum of 3 independent biological replicates was performed using Fiji software.

### FACS and NCAM analysis

On day 14 of the MN differentiation protocol, cells were dissociated with TrypLE Express and DNase I. We collected 2 million cells, fixed for 30 min at 4°C with 4% PFA and stained with Alexa Fluor 647 Mouse Anti-Human CD56 (1:100, BD Biosciences) antibody for 30 min at 4°C. Next, cells were washed with PBS and resuspended in Neurobasal medium minus Phenol Red (Life Technologies) + 2% fetal bovine serum (GE Healthcare Lifesciences) for analysis on either BD FacsAria SORP 5-Laser or BD LSRFortessa. Flow cytometry data from a minimum of 3 independent experiments was further analyzed using FlowJo software.

### Cell viability assays

For assessment of cell viability, we used 3 different methods: **1**) Propidium Iodide (PI) incorporation in day 14 MNs was measured to assess differential cell death between cell lines utilized in this study. Cell cultures were dissociated with TrypLE Express and DNase I and re-suspended in Neurobasal (without Phenol Red) media + 2% fetal bovine serum. PI (5μg/ml) was added to the cell suspension right before flow cytometry analysis on BD LSRFortessa. **2**) The Live/Dead viability/cytotoxicity kit (Molecular Probes) was used following manufacturer instructions. Briefly, live cells exhibit intracellular esterase activity, determined by enzymatic conversion of the non-fluorescent cell-permeant calcein AM (2μM) to the intensely green fluorescent calcein, whereas dead cells with damage membranes incorporate ethidium homodimer (EthD-1, 4μM) that binds to nucleic acid producing a bright red fluorescence. Immunolabeled cells from 9 to 12 predesigned adjacent optical fields were captured with a Leica DMi8 microscope and analyzed by Fiji software. **3**) CytoTox 96® Non-Radioactive Cytotoxicity Assay (Promega) is a colorimetric assay that quantitatively measures lactate dehydrogenase (LDH), a stable cytosolic enzyme that is released upon cell lysis. We collected cell media at distinct time points during the MN differentiation and we measured the extracellular levels of LDH enzyme by quantifying the conversion of a tetrazolium salt into a red formazan product, read by 490-492 nm absorbance, to assess the level of cell death in the different cell genotypes. The analyses were done in at least 3 independent differentiations with a minimum of 3 technical replicates per condition.

### Quantification of acetylcholine

To quantify the levels of acetylcholine released in MN cultures, cells were rinsed with PBS and incubated for 20 minutes at 37°C with artificial cerebrospinal fluid (aCSF; 10mM glucose, 3 mM KCl, 124 mM NaCl, 1.25 mM $KH_2PO_4$, 2 mM $CaCl_2$, 1mM $MgSO_4$, 26mM $NaHCO_3$; pH 7.4). Artificial CSF was collected, centrifuged at 3000 g for 7 min and supernatant was isolated. The concentration of acetylcholine in the medium was determined by High-Performance Liquid Chromatography (HPLC).

### Electrophysiological recordings

MN cultures were plated on Poly-D-Lysine/Laminin-coated M678-GL1-30Pt200 12-well plates (Axion BioSystems) at a density of 30K cells per well. The next day 100K WT mouse glia were plated per well. Recordings were done 3 times per week (Monday/Wednesday/Friday) from day 19 until day 40 using Axion Integrated Studio (AxIS, Axion Biosystems) and data processing was done using AxIS Metric Plotting Tool (Axion BioSystems). Whole-cell patch clamp was performed on day 40-50 MNs using 2-4 MΩ glass electrodes pulled from glass capillary tubes with a Flaming-Brown P-97. Electrodes were positioned using a Sutter Instrument MP-285 motorized micromanipulator (Sutter Instrument Company). Whole-cell patch clamp measurements were performed at room temperature using the Multiclamp 700B amplifier (Molecular Devices) and Winfluor software. Briefly, coverslips with cells were perfused with a modified Ringer's solution containing (in mM): 111 NaCl, 3.09 KCl, 25.0 $NaHCO_3$, 1.10 $KH_2PO_4$, 1.26 $MgSO_4$, 2.52 $CaCl_2$, and 11.1 glucose for an hour before recording commenced. The solution was oxygenated with 95% $O_2$ and 5% $CO_2$ and the perfusion rate was 1.5 – 2.0 ml/min. Patch electrodes contained (in mM) 138 K-gluconate, 10 HEPES, 5 ATP-Mg, 0.3 GTP-Li and Texas Red dextran (75 μM, 3000MW, Life Technologies). In voltage-clamp mode, fast and slow capacitance transients as well as whole-cell capacitance was compensated using the automatic capacitance compensation on the Multiclamp. Holding potential was set at −90 mV in voltage clamp. To measure persistent inward currents, MNs were subjected to slow, depolarizing voltage ramps of 22.5mV/s, bringing the cell to 0mV in 4 s, and then back to the holding potential in the following 4 s. In current clamp, hyperpolarizing current was injected to maintain all neurons at −80 mV. Neurons were then subjected to a range of 1 s depolarizing steps to test for the maximum firing rates (both instantaneous and steady state). Hyperpolarizing current steps were used to test Ih. Depolarizing current ramps (4 s duration) were also used for testing *I*on (the current level at firing onset), *I*off (the current level at cessation of firing), deltaI (Ion-Ioff) and the frequency–current (*F–I*) relationship. Action potential properties were all measured from the first spike of an episode of repetitive firing elicited from a depolarizing current ramp. Neurons were included for analysis if 1) the membrane potential was more hyperpolarized than −45mV at 0 injected current (measured from descending ramps in voltage clamp); 2) the series resistance was less than 20MΩ; 3) repetitive firing could be elicited; and 4) voltage threshold was less than −20mV. Statistical analysis was performed using multivariate analysis with genotype, days in culture, and week as fixed factors with post hoc Bonferroni corrected multiple comparisons tests (SPSS statistical software).

### Characterization of neural progenitor cells

NPCs were differentiated from ESCs as described previously (Topol et al., 2015). Briefly, 70% confluent ESC cultures are cultured in matrigel coated plates with N2B27 medium containing dual SMAD inhibitors (SB431542 and LDN-193189) for 12 days, having the media also supplemented with laminin (Life Technologies) from day 5 to 12. Next, neural rosettes are dissociated with neural rosette selection reagent (STEMCELL Technologies) to obtain NPCs, which are expanded in N2B27 medium supplemented with bFGF (Millipore). The proliferation of WT and *3A*-KO NPCs was determined by three different methods. First, we performed a cell-growth assay: NPCs were dissociated with accutase, resuspended in DMEM/F12 media and cell numbers were quantified with Countess™ (Thermo-Fisher) every 24 h for 4 consecutive days. Second, we performed cell cycle analysis by flow cytometry: NPCs were dissociated with accutase, washed twice with PBS and fixed with iced cold ethanol for 2 h. Next, cells were washed with PBS twice, DNA was stained with propidium iodide (PI, 1mg/ml; Sigma-Aldrich) for 20 min at 37°C and cell PI intensities were quantified with a BD LSRFortessa 6-Laser analyzer. Third, we measured protein levels of the mitotic marker, Phospho-Histone H3 (PH3), by WB.

### Rescue of PAX6 expression

DNMT3A KO cells were co-transduced with FUW-rtTA (Zhang et al., 2013) and pLV-TETi-PAX6-2A-GFP, 24hrs before initiating MN differentiations. Viruses were prepared by the Northwestern University DNA RNA Delivery Core. Differentiation media was supplemented with doxycycline (4 μg/ml) from days 0-6 and removed from days 7-14 to induce and silence PAX6 expression respectively. WB analysis was performed from cells collected at day 6 and ICC from cells fixed at day 16.

### Targeted DNAme using dCas9-DNMT3A

Lentivirus for Fuw-dCas9-DNMT3A-P2A-BFP (Addgene plasmid # 84569) (Liu et al., 2016) and pgRNA-humanized-mCherry (Addgene plasmid # 44248) (Qi et al., 2013) plasmids were packaged in transfected 293T cells using 2nd generation packaging vectors psPAX2 and pMD2.G as described previously (Zufferey et al., 1998) by the Northwestern University DNA RNA Delivery Core. The coordinates of the targeted GREs are: *PAX6* region 1/2 (one element split into 2 pieces) chr11:31,821,694-31,822,222/chr11:31,818,705-31,820,467; *PAX6* region 3; chr11:31,840,585-31,840,607; *ARX* region 1; chrX:25,011,708-25,011,730; *ARX* region 2; chrX:25,017,503-25,017,525. Sequences for the guide RNAs can be found in Table S7. The *3A*-KO cell line (#139) was transduced with the Fuw-dCas9-DNMT3A-P2A-BFP lentivirus and we performed two serial FACS-sorting protocols of BFP+ cells to generate a stable cell line. We then infected this line with distinct combinations of PAX6 and ARX GRE guides RNAs using the pgRNA-humanized-mCherry virus 24 h before the induction of MN differentiation. Cells were collected on days 6 and 14 for DNAme analysis and WB analysis, and fixed at day 16 for ICC.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### WGBS data processing

Whole genome bisulfite sequencing raw reads and amplicon sequencing data were aligned using bsMap version 2.7 (Xi and Li, 2009) with default parameters against the human genome build GRCh37. Next, we used the Picard suite to remove all duplicate reads. Subsequently, we carried out methylation using the mcall module implemented in the MOABS suite (Sun et al., 2014).

### Identification of DMRs

DMRs were identified using the R package DSS (Park and Wu, 2016). To that end, we performed all pairwise comparisons between the WT time course samples and between matching time points of the WT and KO samples using all replicates for each condition listed in Table S1. For each of these pairwise comparisons, we applied the following 3 functions from the DSS package to the appropriate biological replicates. First, utilized dmlTest with smoothing = T and smoothing.span = 200. Next, we identified differentially methylated CpGs using the callDML function with a threshold of p = 0.001. Finally, we identified DMRs using the callDMR function directly using the posterior probability that the methylation difference exceeds a certain value. More specifically, we set the parameters delta = 0.15, p.threshold = 0.05, minCG = 4 and dis.merge = 500. All other parameters were left at their default settings. Next, we merged all DMRs identified across the pairwise comparisons into one DMR set, collapsing DMRs that overlap by at least one base pair into a single DMR. For further downstream analysis and visualization, we then employed the methylKit package (Akalin et al., 2012). In particular, we computed the DNAme level and coverage of each DMR in each sample. We then only retained those DMRs that were covered by more than 10 reads in all samples. Next, we average the DNAme levels of each DMR across replicates assigned the DMRs to different groups based on their methylation difference between the coverage weighted replicate average DMR level methylation values. DMRs that exhibited an absolute methylation difference $\geq$ 0.2 between any pair of the WT samples were defined as differentiation DMRs and are shown in Figure 1B. Next, we classified the DMRs that arise between the KO and the WT samples into 4 groups. The first group contains DMRs that normally arise over the course of the WT differentiation, gaining methylation as the cells progress, but fail to do so in KO. This group is the further split into those regions that exhibit already differential methylation in the pluripotent state (pre-existing hypo DMRs) and those that do not (non-pre-existing hypo DMRs). In addition, we also classified the group of DMRs that fail to lose DNAme in the KO (hyper DMRs), also splitted into those that already exist in the pluripotent state and those that arise over the course of differentiation (non-preexisitng non-diff DMRs). These 4 groups are shown in Figures 3B and 3I and are listed in Table S2. Please note, that a small number of DMRs can be in multiple groups, e.g., hypermethylated at one stage and hypomethylated at another.

### DMR gene set enrichment analysis (GSEA)

In order to annotate different sets of differentially methylated regions we clustered the DMRs that arise over the course of WT HUES-64 differentiation into 3 clusters and performed gene set enrichment analysis using the GREAT toolbox (McLean et al., 2010) separately for each cluster. DMRs and cluster memberschip can be found in Table S2. Furthermore, we performed GSEA using GREAT for all non-prexisting DMRs that arise between the WT and the *3A*-KO over the course of MN differentiation and are hypomethylated in the 3A-KO. The full list of significantly enriched regions (using the GREAT standard parameter settings for significance) is provided in Table S3 and a selected subset is shown in Figures 1B and 3E.

### Integrated analysis

For the integrated analysis, we associated each differentially expressed gene between the WT and the *3A*-KO or the WT and the *3B*-KO with all *3A*-KO DMRs within 100kb upstream or downstream of the genes TSS. We report the entire list in Table S5. For visualization purposes, we collapsed all DMRs that were associated with the same gene and average their DNAme levels. The results are shown in Figure 5A.

### Annotation with genomic features

In order to gain a better understanding of the genomic context of differential methylation, we overlapped the DMRs set with a collection of genomic features, including 6 ChIP-Seq based peaks sets for TFs mapped in WT HUES-64 embryonic stem cells (Tsankov et al., 2015). The feature was defined as Promoters: +- 1kb of non-overlapping ENSEMBL gene TSS, ENCODE TFBS represents a collection of over 1000 ChIP-Seq experiments for various TF in different cell lines and was obtained from the ENCODE/UCSC website (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegTfbsClustered/). CpG islands (Table S3), using CgiHunter (http://cgihunter.bioinf.mpi-inf.mpg.de/), requiring a minimum CpG observed versus expected reatio of 0.6, a minimum GC content of 0.5 and a minimum length of 700bp. CpG island shores were denoted as 2kb regions directly upstream and downstream of CpG islands. For feature overlap analysis, we obtained DNase HS I, H3K4me3 and H3K27ac ChIP-Seq for a large collection of cell and tissue types from the REMC (Kundaje et al., 2015) and ENCODE (Bernstein et al., 2012). We counted features as overlapping if they overlapped by at least 1 base pair. The results of this analysis are shown in Figures 3D and 5A.

### RNA-Sequencing data processing and analysis

Raw RNA-Seq reads were trimmed to a total length of 45 bp, removing 5bp from the 5′ end and removing all bases beyond bp 50 at the 3′ end using trimmomatic version 0.33 (Bolger et al., 2014). Subsequently, gene expression levels were quantified using Salmon

(Patro et al., 2017) against the cDNA sequences of all ENSEMBL transcripts release 89 of genome assembly GRCh37 using default parameters. Sample quality metrics, group assignment and GEO numbers are listed in Table S1. Next, gene expression levels were quantified by collapsing distinct transcript read counts onto the corresponding ENSEMBL genes using the R packages tximport (Soneson et al., 2015) and tximportData with the salmon import option. Following quantification, differential gene expression analysis was performed using the DESeq2 framework (Love et al., 2014) with default parameter settings. Finally, all genes differentially expressed at an Benjamini-Hochberg corrected p value $\leq 0.01$, a minimal log2 fold change $\geq 1$ and a minimal expression level $\geq$ 10 FPKM in at least one of the samples in each pairwise comparison were defined as the set of differentially expressed genes between the various comparisons. For visualization purposes, sequencing depth normalized gene expression levels were averaged across replicates for each sample group transformed to row-wise z-scores. These values were also used for clustering analysis using k-means clustering with 100 random initializations. The number of clusters was estimated based on the drop in the ratio of the between cluster standard squared error and the total standard squared error for a cluster number range of 5 to 20 clusters. The clustering of all differentially expressed genes along with their differential expression status for various pairwise comparisons can be found in Table S4. For the gene set enrichment analysis reported in the text, we obtained all developmental gene set definitions at different stages (TSxx) of mouse development from the MGI (http://www.informatics.jax.org/) database. We then converted all mouse gene identifiers to human using the ENSEMBL cross-species mapping table using biomaRt (Durinck et al., 2009). Non-mappable gene symbols were discarded. Next, we performed gene set enrichment analysis for each of the 10 KO gene clusters for each MGI gene set. We determined gene set enrichment using Fisher's exact test, defining all genes expressed greater than 10 FPKM in at least one of the conditions as background. The resulting p values for where then corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). The full list is provided in Table S4. In addition, we performed gene set enrichment analysis using the enrichR software suite (Kuleshov et al., 2016) for genes up or downregulated separately, retaining only gene sets enriched at a Benjamini-Hochberg corrected p value $\leq 0.05$. We further filtered this list for genes that were annotated as TFs and report them in Figures 3M and 3N, as well as genes that were annotated as neurotransmitter, synaptic or channel in the GO database and highlight them in Figure S7B.

### Single-cell RNA-Seq data processing and analysis

Raw reads were then processed using the 10X Genomics cell ranger pipeline (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) with standard settings. Following these pre-processing steps, the resulting UMI tables were analyzed for day 6 and day 14 separately using the R package Seurat version 2.0.0 (Satija et al., 2015). In Seurat, cells were first filtered for number of genes per cell between 200 and 10000 as well as having less than 6% of reads aligning to mitochondrial RNA. Subsequently, UMI counts were coverage and log normalized. Following normalization, we performed variable gene detection using the FindVariableGenes function with the following parameters: function = ExpMean, dispersion.function = LogVMR, x.low.cutoff = 0.125, x.high.cutoff = 3, y.cutoff = 0.5. We then continued the analysis focusing only these variable genes. Next, we performed cell cycle phase analysis of all cells using the CellCycleScoring function and cell cycle gene sets obtained from Seurat. For day 6 samples, we accounted for confounding factors by regressing on the total number of genes per cell, the percentage of reads aligning to mitochondrial RNA, the S-phase cell cycle scores and the G2M cell cycle scores using the ScaleData function. We then worked with resulting scaled values in subsequent analyses. For day 14 single cell expression profiles, we regressed instead on the differences in the S and G2M phase scores. In order to detect clusters of single cells, we next determined the number of principal components capturing meaningful biological information using the JackStraw function and the variance explained per PC. For day 6 samples, we selected the first 17 PCs, whereas for the day 14 we focused on the first 10 PCs. We then utilized PCs to identify clusters of cells in the reduced principal component space using the FindClusters function with the resolution parameter set to 0.9 for day 6 samples and 0.75 for day 14 samples. For visualization purposes, we then performed a t-SNE embedding on the reduced PC space and labeled the cells according to identified clusters. Heatmaps, violin plots and feature plots for selected genes shown in Figure 4 were generated using Seurat's visualization functiona PCHeatmap VlnPlot and FeaturePlot. For the latter, we capped the maximum expression level at a log normalized expression value of 1. Lastly, we plot the cumulative distribution function of the scaled single cell expression levels for *PAX6* and *HES1* in Figure S3C. In order to quantify the number of neuronal and non-neuronal cells depicted in Figure 4K, we used the single cell clustering for day 14 data. All cells in clusters 6-10 were designated neuronal based on appropriate marker expression (see Figures 4J and 4L). All cells in clusters 1-5 were designated as non-neuronal.

### Statistical testing

Statistical analysis from a minimum of 3 independent differentiations displayed in Figures 1, 2, 4, 6, and 7 and Figures S1, S2, S4, S6, and S7 were performed using GraphPad Prism software. Normality test was used to determine the application of the proper statistical approaches. To analyze data with a normal distribution, paired Student's t tests were used to compare two experimental conditions, and ANOVA followed by post hoc test for comparisons of three or more experimental groups. For data without a normal distribution, Mann-Whitney test was applied. For the comparison of the global DNAme levels between WT and *3A*-KO based on the WGBS data, we used the global mean DNAme levels of each sample, averaged across all covered CpGs, grouping all replicates for day day 6 and day 14 into a WT (n = 8) and *3A*-KO (n = 6) groups. These values are also reported in the column mean Table S1. We then performed a two-tailed t test using the r function t.test giving rise to the reported p value of 0.12.

The exact number of biological and/or technical replicates, the specific statistical test applied and the exact p values from main figures are:

Figure 1A: Statistical differences among the 5 groups determined by ANOVA (differences between groups, $p$ = 3.249e-4) and Bonferroni's Multiple Comparison Test (***$p < 0.001$).

Figure 2B: A total of n = 4-9 fields in n = 3-14 independent differentiations were analyzed. Data are represented as means ± SEM. Statistical significance determined by ANOVA (differences between groups $p$: ISL1/2 = 7.029e-21; ChAT = 2.354e-34; TUJ1 = 2.109e-66) and Bonferroni's Multiple Comparison Test (***$p < 0.001$).

Figures 4E and 4F: Quantified n = 45,467 DAPI cells in n = 6 fields, and n = 2 dishes per differentiation, in n = 3 independent differentiation experiments. Data are represented as means ± SEM. Statistical significance determined by a 2-tailed t test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 6D: Statistical significance determined by ANOVA and Bonferroni's Multiple Comparison Test (***$p < 0.001$ between non-transduced and PAX6::GFP-transduced *3A*-KO cells; n = 3 independent differentiations).

Figure 6G: Statistical significance determined by ANOVA (differences between groups noted by $p$ of 1.314e-8, 7.421e-5 and 5.38e-7 for ISL1/2, ChAT and FOXA2% respectively) and Bonferroni's Multiple Comparison Test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$), n = 3 independent differentiations.

Statistical information from supplemental figures is provided in the respective figure legends.

## DATA AND SOFTWARE AVAILABILITY

The accession number for the data reported in this paper is GEO: GSE90553. All software utilized is listed in the methods sections and publically available.