

# Single-cell epigenomic reconstruction of developmental trajectories from pluripotency in human neural organoid systems

In the format provided by the  
authors and unedited

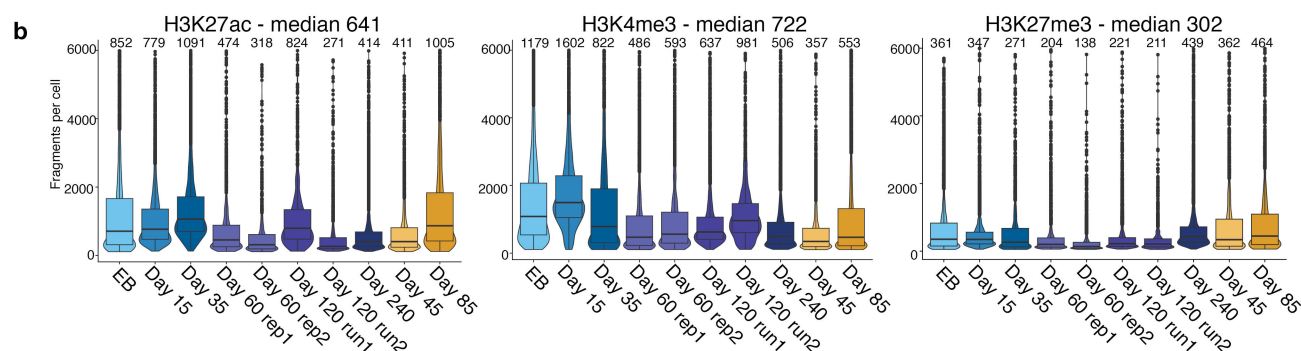
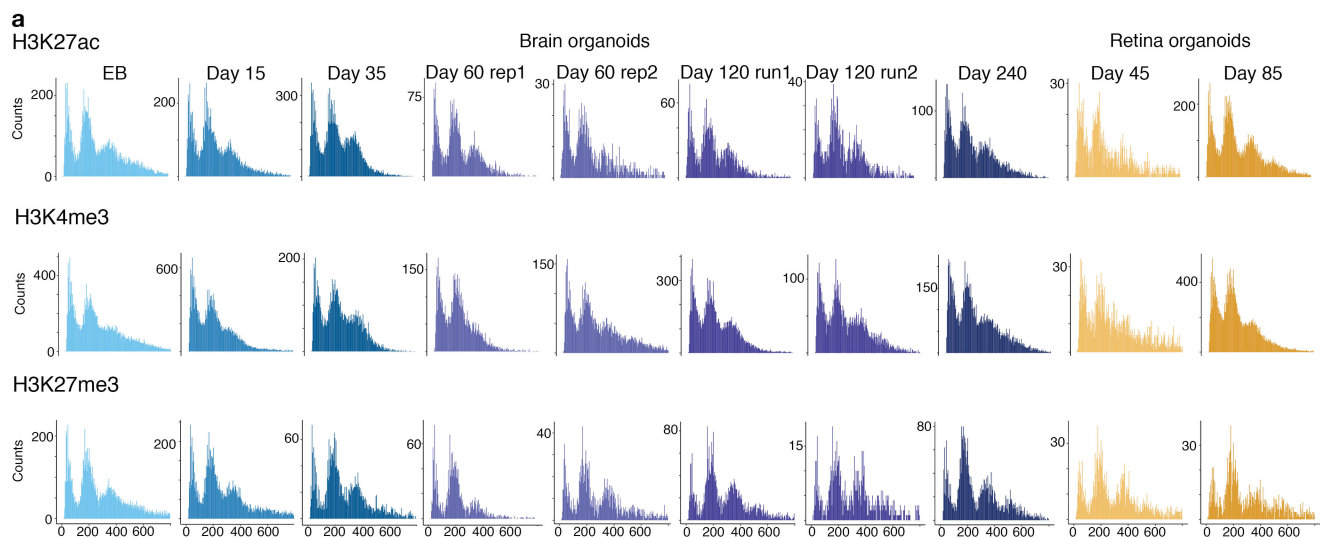
## **Supplementary Information**

### **Regionalization within the glial lineage**

We investigated regional variance also at the chromatin level when we associate region-specific peaks of the active marks H3K27ac and H3K4me3 to the closest gene (Supplementary Fig. 2j). Interestingly, H3K27me3 at regional marker genes (LHX1, FOXG1, LHX2) showed higher gene repression variance in NPCs compared to astrocytes indicating that gene repression in astrocytes is less predictive of regional identity. Altogether, our analysis reveals that the interplay of chromatin modifications induces and stabilizes brain region diversification events.

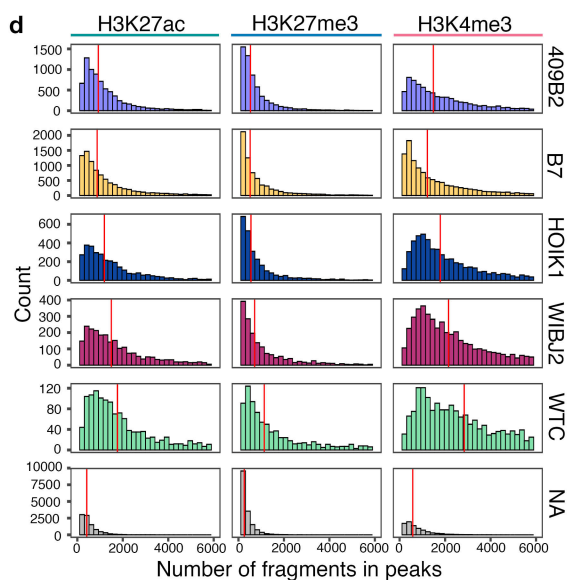
### **Differential regulation of regional diversification and neuronal differentiation**

We analyzed each gene across all modalities to determine how histone modifications might be differentially involved in brain region diversification versus neuronal differentiation. (Supplementary Fig. 7a-d, Supplementary Table 16, see Methods Analysis of pseudotemporal and regional variance). Overall, RNA expression and histone modifications (Supplementary Fig. 7e) showed high concordance. Variance in gene expression and H3K27ac enrichment was mostly explained by regional diversification (Supplementary Fig. 7f-g), whereas H3K27me3 showed similar variance across regional branches and over pseudotime. H3K4me3 showed lower variance across regional branches than over pseudotime (Supplementary Fig. 7g). We characterized the genome-wide distribution of branch-specific peaks. This revealed that the majority (~75%) of H3K4me3 peaks was located at promoter regions, while for H3K27ac and H3K27me3, a lower percentage (60%) of branch-specific peaks was found at promoter regions and instead, 40% of these peaks were found in distal regions (Supplementary Fig. 7h).



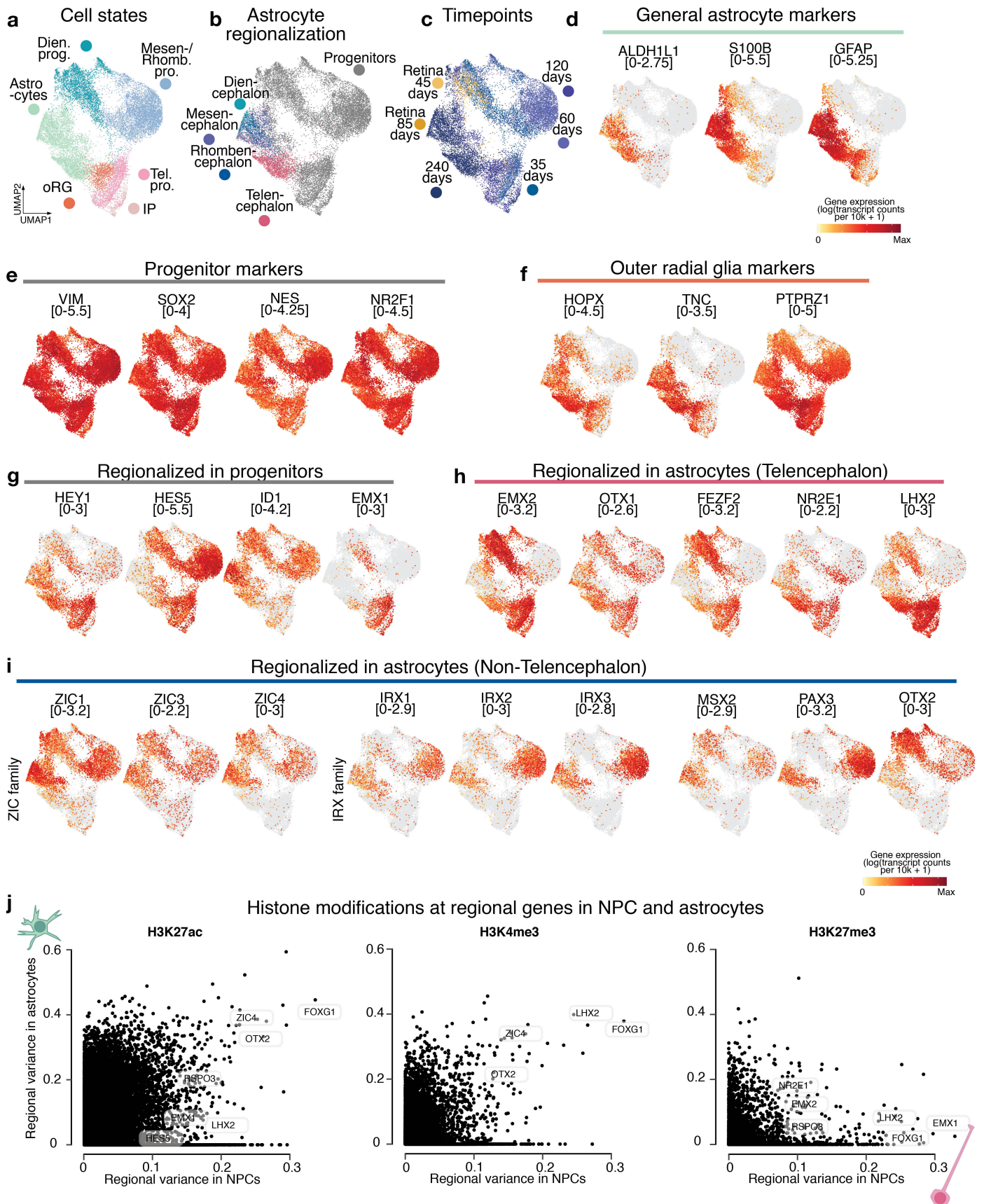
**c**

| Citation                                   | Method        | Tissue/Cell lines                                                                                                                                                                                           | Median fragments/cell                                |
|--------------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|
| Zenk & Fleck et al. 2023 (presented study) | scCUT&Tag     | Human brain and retina organoids                                                                                                                                                                            | H3K27ac - 641 (range: 271 - 1005)                    |
|                                            |               |                                                                                                                                                                                                             | H3K4me3 - 722 (range: 357 - 1602)                    |
|                                            |               |                                                                                                                                                                                                             | H3K27me3 - 302 (range: 138 - 464)                    |
| Zhang et al. 2022 Nat. Biotech.            | scCUT&Tag pro | Human peripheral blood mononuclear cells (PBMC)                                                                                                                                                             | H3K27ac - 199<br>H3K4me3 - 316<br>H3K27me3 - 501     |
| Bartosovic et al. 2021 Nat. Biotech.       | scCUT&Tag     | Postnatal mouse brain (FACS sorted) and mix of mouse embryonic stem cells (mESC, C57Bl/6J origin), mouse embryonic fibroblasts (NIH-3T3) (ATCC), mouse oligodendrocyte progenitor model cell line (Oli-neu) | H3K27ac - 453<br>H3K4me3 - ~420<br>H3K27me3 - 597    |
| Gopalan et al. 2021 Mol. Cell.             | multi-CUT&Tag | E14 mouse embryonic stem cell line<br>5-4 mouse trophoblast stem cell line                                                                                                                                  | not reported, cut-off of 200 cut sites per cell used |
| Wu et al. 2021 Nat. Biotech.               | scCUT&Tag     | Human peripheral blood mononuclear cells (PBMC)                                                                                                                                                             | H3K27me3 - 1110 (range: 954 - 1230)                  |
|                                            |               |                                                                                                                                                                                                             | H3K27ac - 1047                                       |



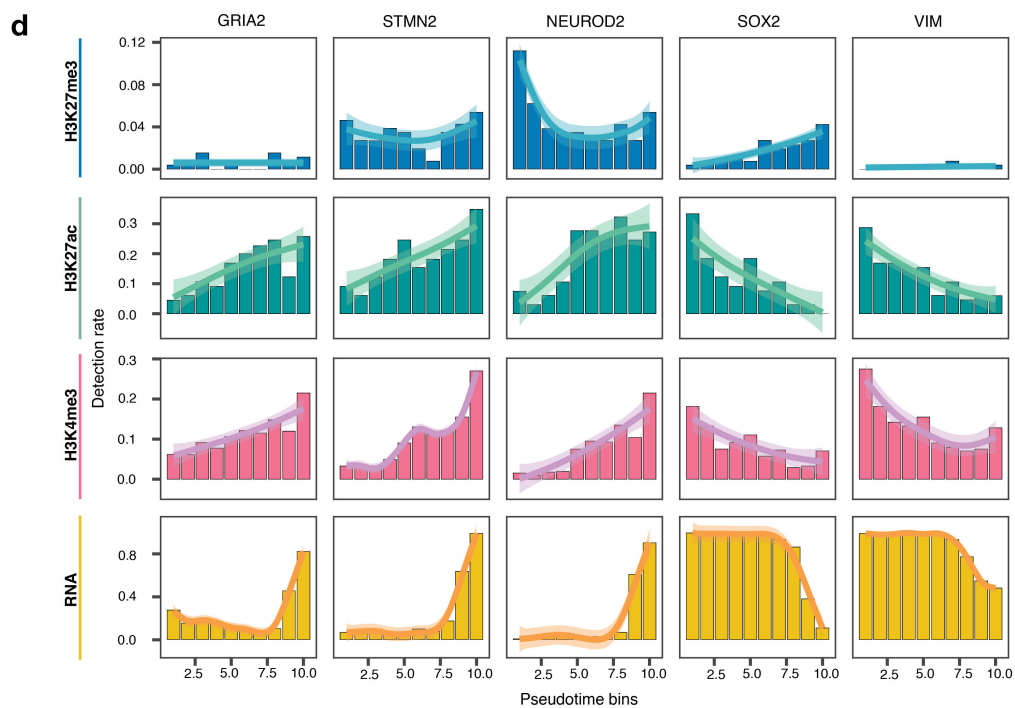
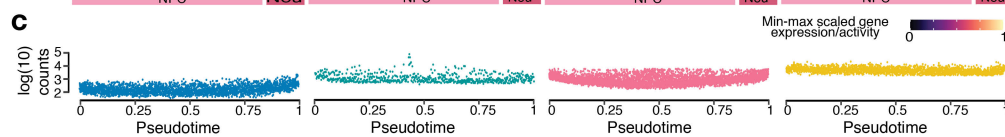
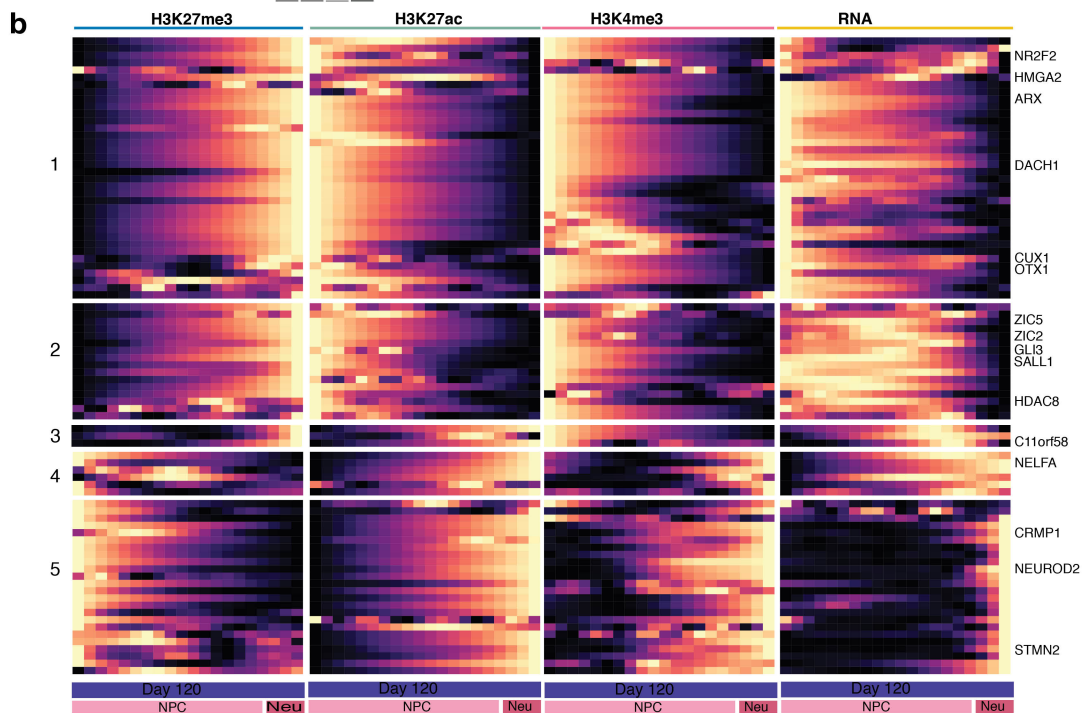
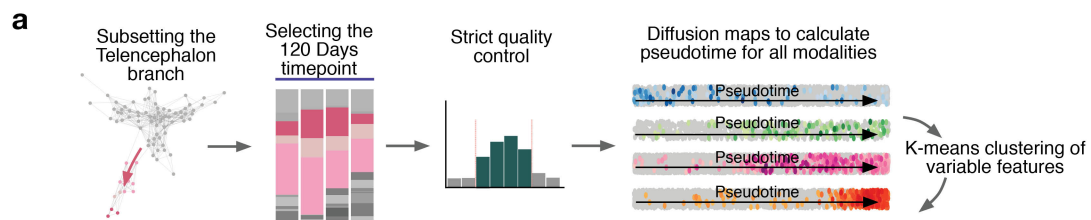
### **Supplementary Fig. 1 Quality metrics of the scCUT&Tag dataset**

a, Histograms showing fragment length distribution for all samples (different timepoints including replicates) and modalities (H3K27ac, H3K4me3, H3K27me3), showing a clear nucleosome size pattern b, Violin plots showing the distribution of detected fragments per cell for each time point and chromatin modality (H3K27ac: n=33533, H3K27me3: n=34357, H3K2me3: n=42057, 8 independent timepoints each). The median value of each boxplot is indicated at the top (median +/- Q1/Q3). c, Table comparing median read per cell as a measure of data quality between different scCUT&Tag publications and the presented study (Zenk & Fleck et al. 2023). d, Histograms comparing the distribution of peak fragments per cell for each histone modification in all cell lines in the presented dataset. The red line demarcates the median.



## **Supplementary Fig. 2 Astrocyte regionalization signatures in human brain organoids**

a-c, UMAP embedding of NPCs and astrocytes from the 30-240 day time points colored by cell state (a), brain region (b), and time point (c). d, UMAP embedding colored by gene expression ( $\log(\text{transcript counts per } 10k + 1)$ ) of general astrocyte marker genes. e, UMAP embedding colored by gene expression of general progenitor marker genes. f, UMAP embedding colored by gene expression of general outer radial glia marker genes. g-i, UMAP embedding colored by expression of genes that show a brain region specific expression pattern in NPCs (e), are specifically expressed in forebrain astrocytes and NPCs (f) are specifically expressed in non-forebrain astrocytes and NPCs (g). j, Scatter plot comparing the regional variance ( $R^2$ ) of histone modification enrichment (for H3K27ac, H3K4me3, H3K27me3) between astrocytes and NPCs. This suggests that gene repression (H3K27me3) is less predictive of regional identity in astrocytes compared to NPCs (see Methods Analysis of pseudotemporal and regional variance for details).



### **Supplementary Fig. 3 Pseudotemporal reconstruction of dorsal telencephalic neuron differentiation using diffusion maps**

a, Schematic of the analysis. We subset the telencephalon branch and isolated only cells of the 120 days time point. The cells were again filtered for quality and pseudotime for all modalities was calculated using diffusion maps (see Methods Reconstruction of the neurogenesis trajectory for the 4 months time point for details). b, Heatmap showing scaled and smoothed gene activity scores ( $\log(\text{fragment counts per } 10k + 1)$ ) on gene body +2kb promoter region) for H3K27me3, H3K27ac and H3K4me3 as well as RNA expression ( $\log(\text{transcript counts per } 10k + 1)$ ) over the telencephalic neuron differentiation trajectory from pluripotency. Pseudotime was binned and genes were K-means clustered based on average expression/activity of all marks and RNA in all bins (Supplementary Table 11 contains all clusters). Using this strictly refined analysis we could again identify a neuronal cluster that clearly showed priming with active chromatin modifications. c, Jitter plot showing even UMI count and fragment counts versus pseudotime for all modalities. d, Barplot showing fragment detection rate in pseudotime bins at selected example genes for all modalities. The line indicates smoothing with generalized additive models.

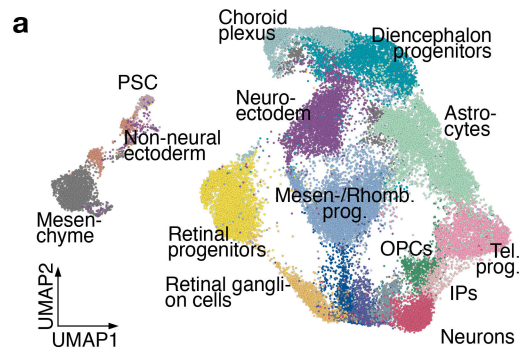


**Supplementary Fig. 4 Global analysis of neuronal genes in organoids and pseudotime reconstruction of neurogenesis in the primary developing cortex reveal widespread priming.**

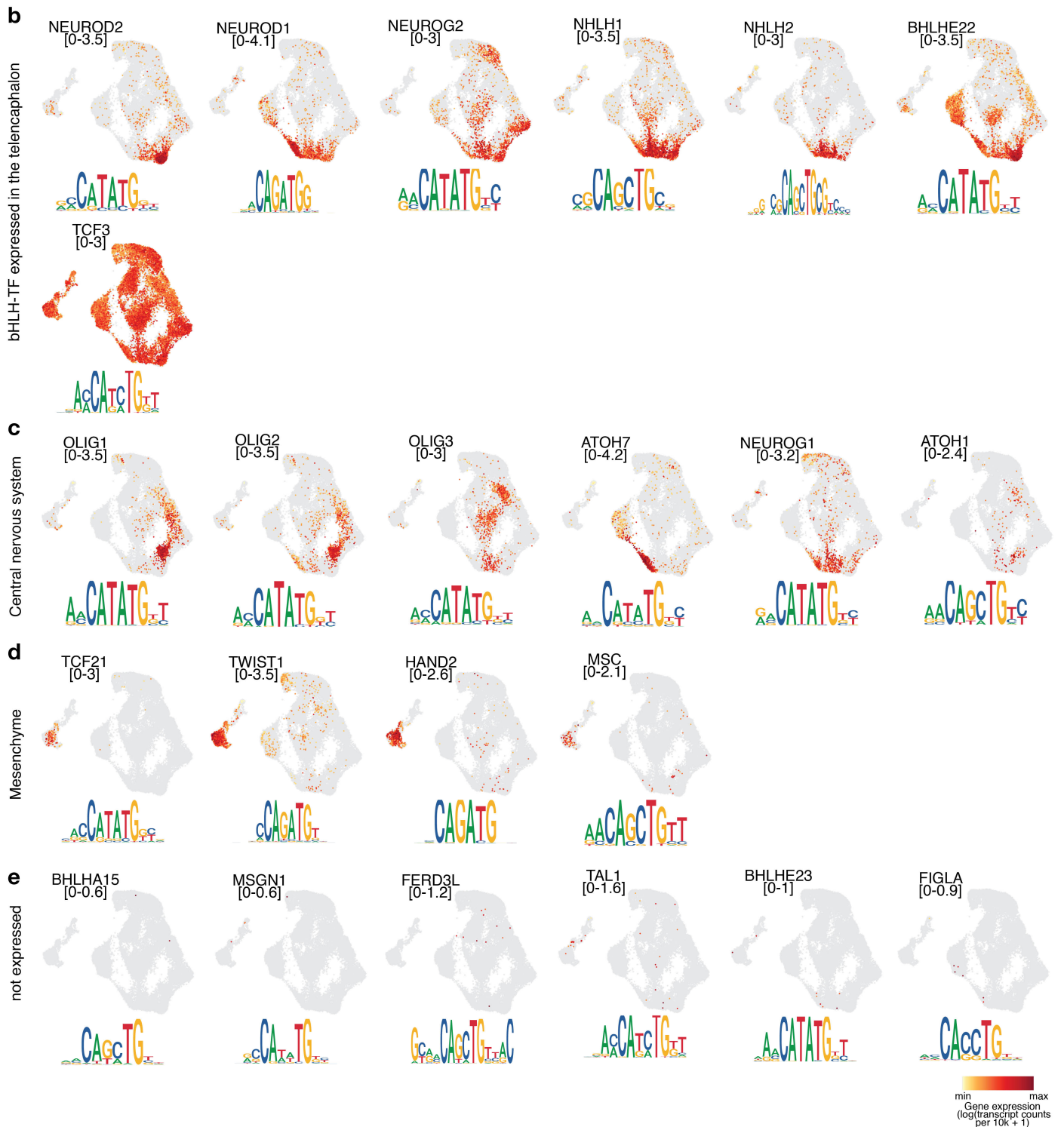
a, Transcription factor motif enrichment in peaks on primed neuronal genes (gene body +2kb promoter region, see Methods Transcription factor motif enrichment)(cluster 6, Fig. 4a) plotted against their expression fold change in cortical neurons. Colored dots indicate a significant motif enrichment by Fisher exact test  $FDR < 0.01$ . b, Ridge plots showing the distribution of pseudotime lag between active histone modifications (H3K27ac and H3K4me3) and RNA expression. Positive values indicate histone modifications preceding RNA, negative values the opposite. (see Methods Detection of inflection point for neuronal genes). Neuronal genes were selected based on having maximal expression in neurons over NPCs and detection of H3K27ac in more than 5% of cells in any neuronal high-resolution cluster. Primed genes show an enrichment of GO-terms related to neuron projection and axonogenesis, while co-transcriptionally marked genes are enriched for terms related to synapse organization. c, Boxplot (median  $\pm$  Q1/Q3) showing the shift in pseudotime between the establishment of active histone modifications, chromatin accessibility and RNA expression on the same set of neuronal genes as b ( $n=357$  genes). d, Schematic of the multiome experiment in the primary developing brain. e, UMAP embedding of the cortical neuron trajectory from the gw 19 primary developing brain multiome data colored by cell state. f, UMAP embedding colored by pseudotime. g, UMAP embedding as is (b) colored by expression ( $\log(\text{normalized transcript counts} + 1)$ ) of different neuronal and progenitor marker genes (top) and gene activity based on chromatin accessibility ( $\log(\text{fragment counts per } 10k + 1)$  on gene body +2kb promoter region) (bottom). h, Barplots with smoothed lines showing detection rate of scATAC fragments and scRNA expression over pseudotime measured from the same cells of a primary human developing embryonic brain at gw 19. Neuronal genes (NEUROD2, STMN2, BCL11B) show priming of chromatin. i, Boxplot (median  $\pm$  Q1/Q3) showing the shift in pseudotime between chromatin accessibility and RNA expression in the fetal cortex on the same set of neuronal genes as in b ( $n=357$

genes). More details on the analysis can be found in the method section on Detection of inflection point for neuronal genes.

## Organoid developmental time course

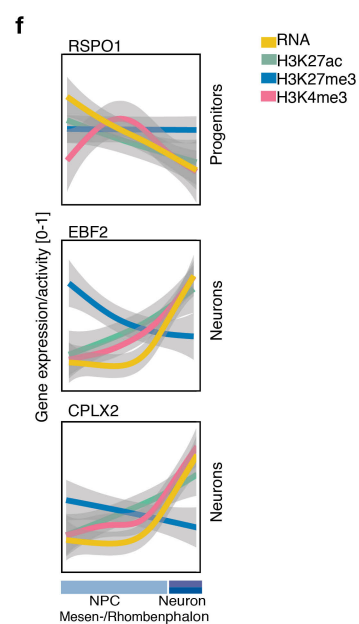
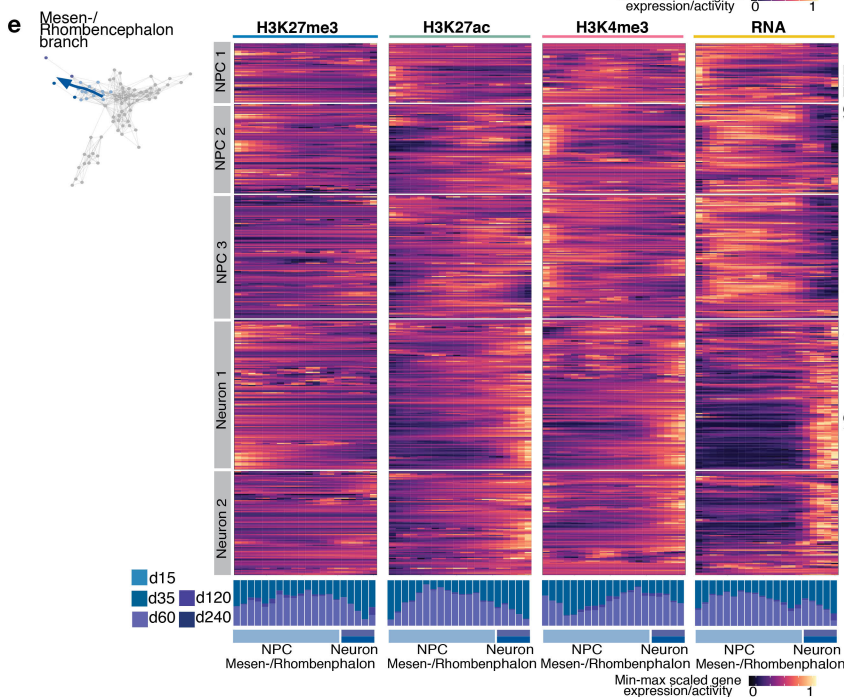
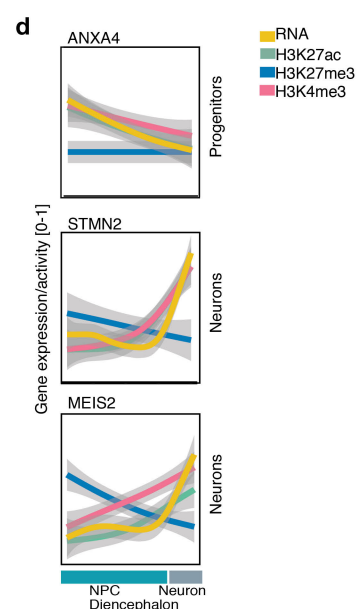
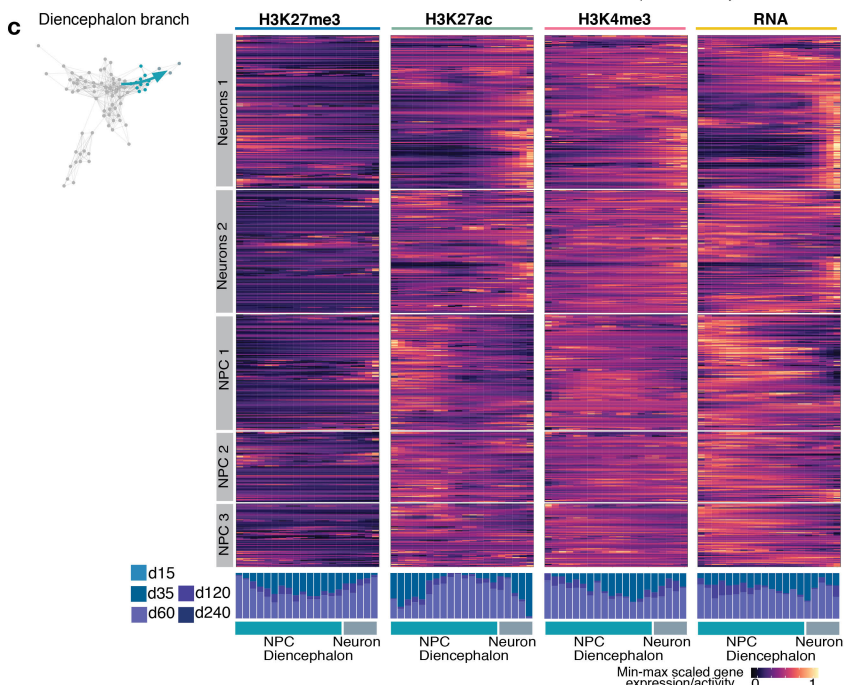
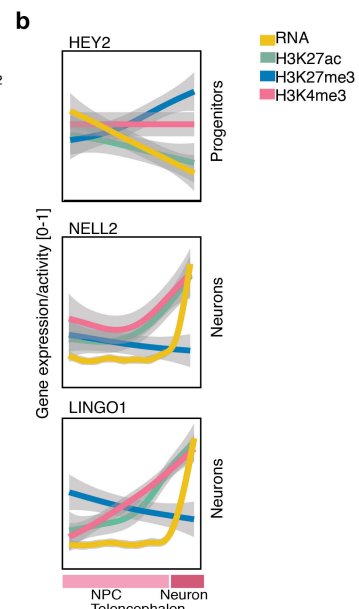
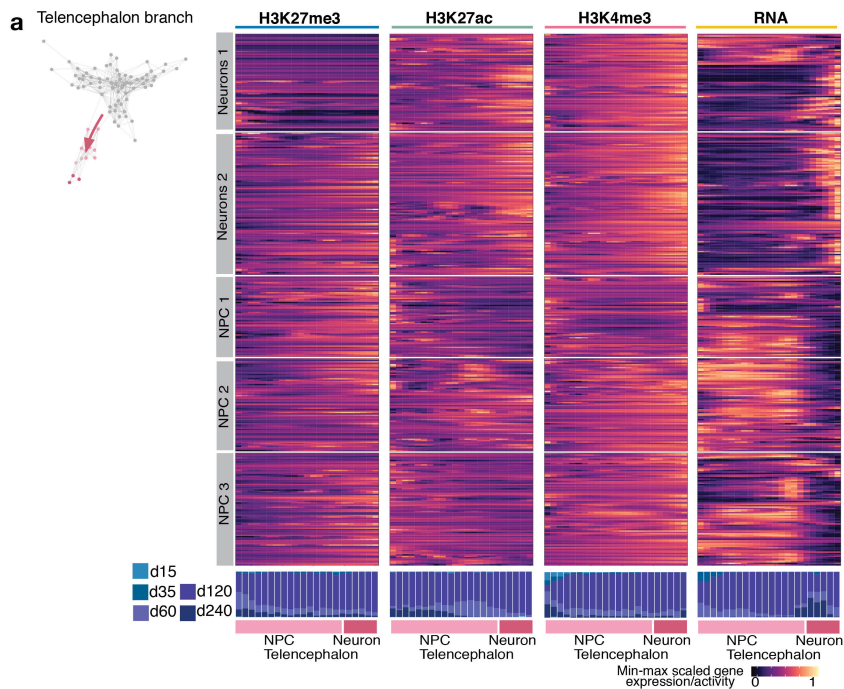


## Expression of bHLH transcription factors in organoid development



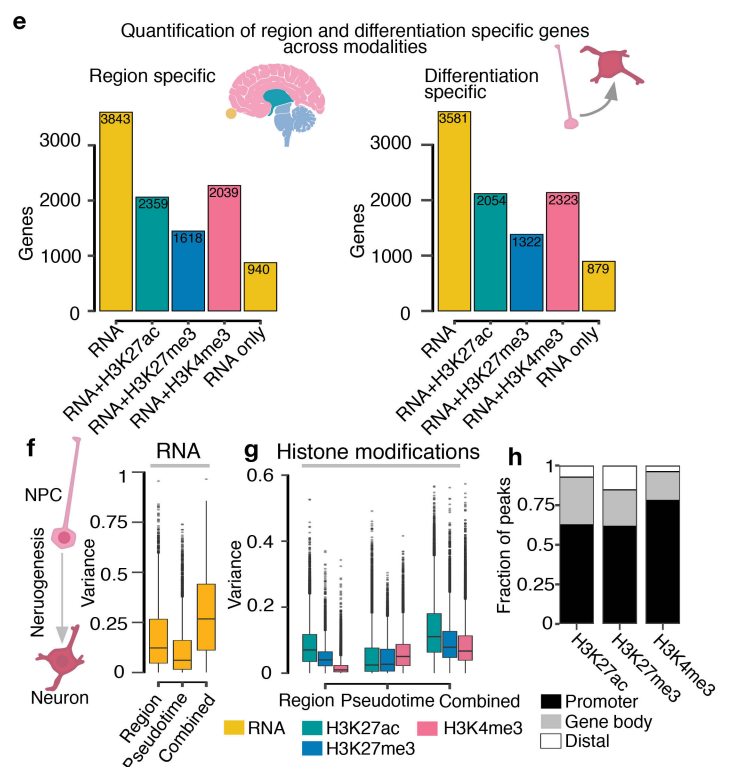
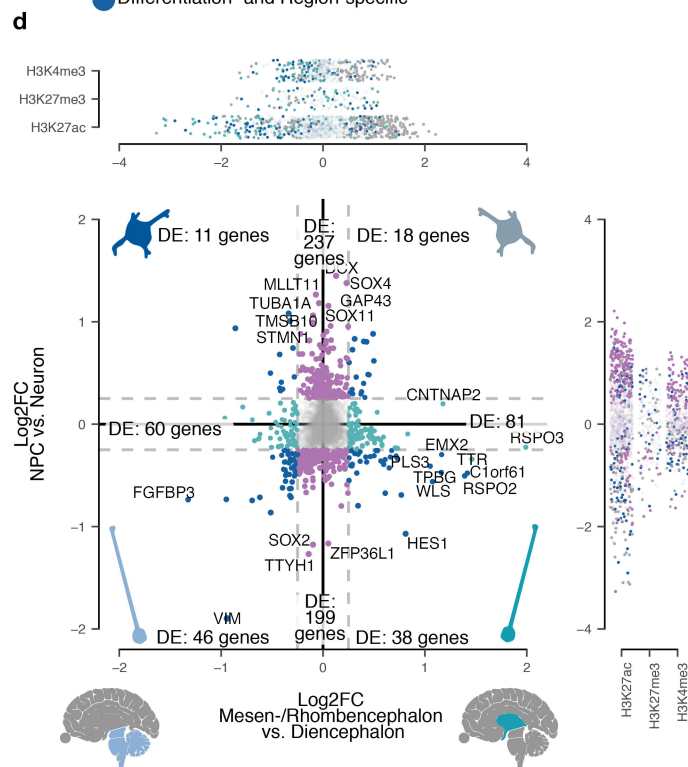
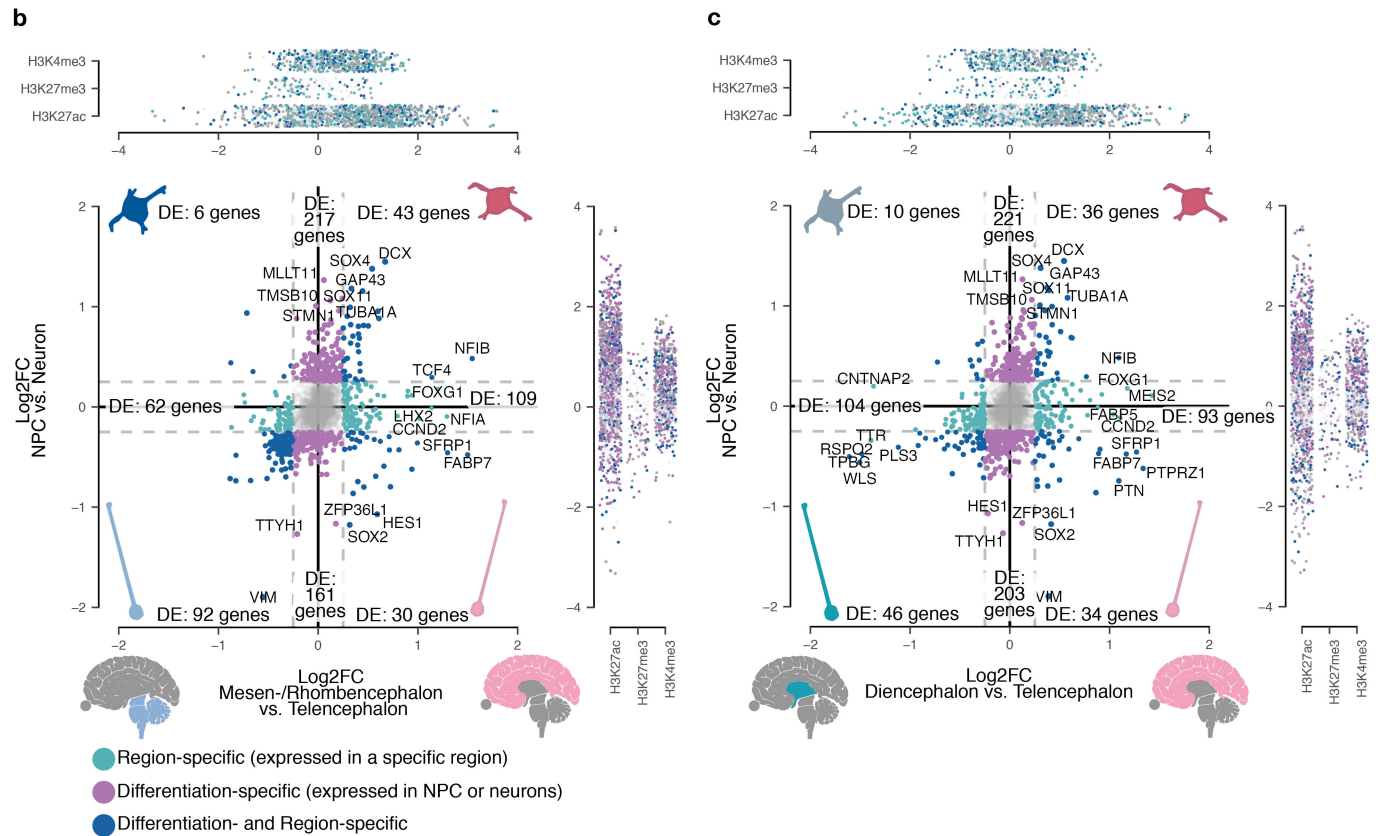
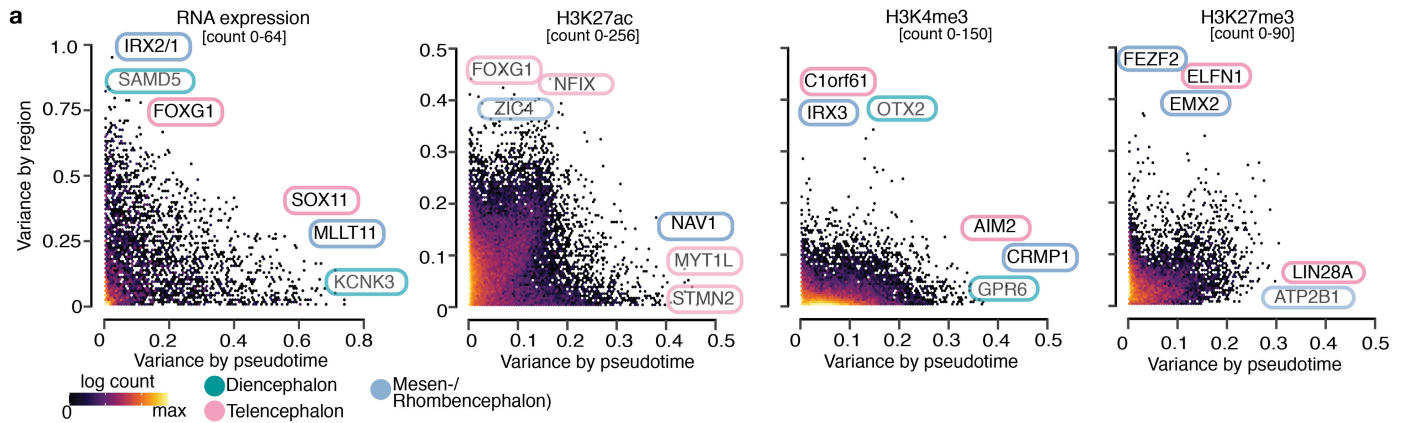
### **Supplementary Fig. 5 Expression analysis of bHLH transcription factors in the time course**

a, UMAP embedding of scRNA-seq data with cells colored and labeled by cell state (IP - intermediate progenitor, PSC - pluripotent stem cell, RGCs - retinal ganglion cells, RPs - retinal progenitors). b, UMAP embedding of the developmental time course as in (a) colored by expression of bHLH transcription factors ( $\log(\text{transcript counts per } 10k + 1)$ ) with expression in the telencephalon branch and their corresponding motif. In particular NEUROG2 and BHLHE22 show very similar motifs and expression profile compared to NEUROD2 and might also act in the GRN (Fig. 4h) c, bHLH transcription factors with expression in the central nervous system. d, bHLH transcription factors with expression in the mesenchyme. e, bHLH transcription factors not expressed in the time course.



## **Supplementary Fig. 6 Pseudotime reconstruction of dynamic changes of all modalities during neurogenesis in different developing organoid brain regions**

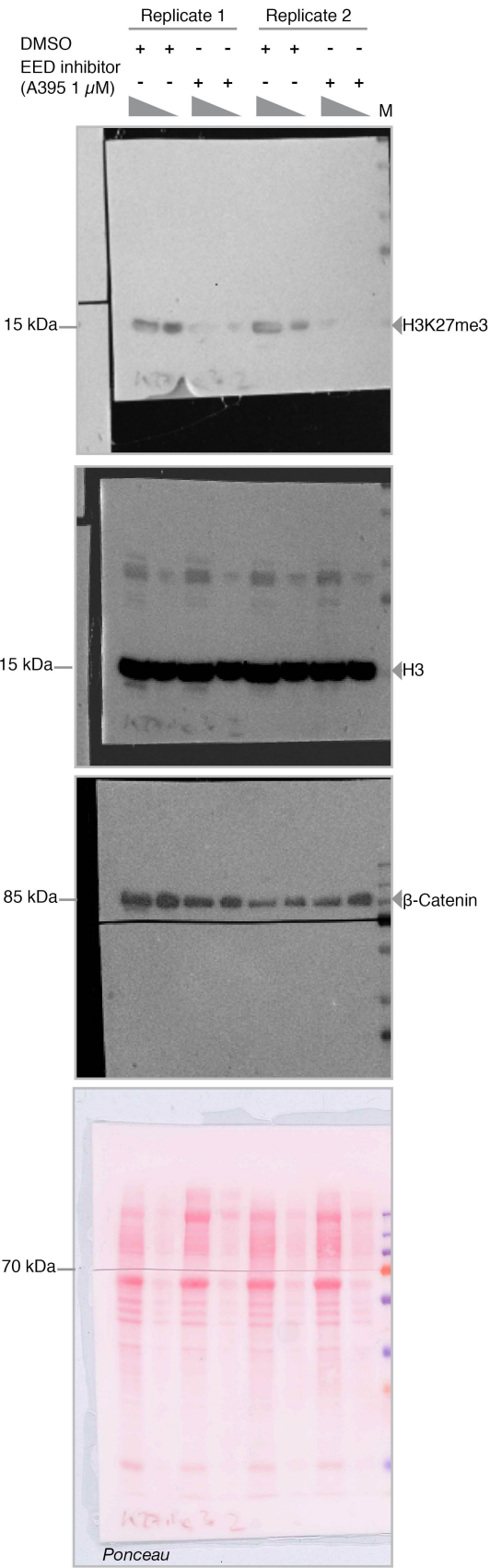
a, Heatmap showing scaled gene activity scores for H3K27me3, H3K27ac and H3K4me3 as well as RNA expression over the telencephalic neuron differentiation trajectory from NPCs. Pseudotime was binned and genes were K-means clustered based on average expression/activity of all marks and RNA in all bins. (The averaged pseudotemporal gene expression and activities of cluster 1 are shown as line plots in Fig. 4j, as well as their GO term enrichment) b, Line plots showing generalized additive model-based smoothing of pseudotemporal gene expression and gene activities of selected examples from multiple K-means clusters. HEY2 is expressed at the progenitor state, showing alignment of pseudotimes. NELL2 and LINGO1 are expressed in neurons and exhibit epigenetic priming with activating marks. c, Same as (a) for the diencephalic neuron differentiation trajectory from NPCs. (The averaged pseudotemporal gene expression and activities of cluster 1 are shown as line plots in Fig. 4j, as well as their GO term enrichment) d, Line plots showing smoothed pseudotemporal gene expression and gene activities of selected examples from multiple K-means clusters. ANXA4 is expressed at the progenitor state, showing alignment of pseudotimes. STMN2 and MEIS2 are expressed in neurons and exhibit epigenetic priming with activating marks. e, Same as (a) for the rhomben- and mesencephalic neuron differentiation trajectory from NPCs. (The averaged pseudotemporal gene expression and activities of cluster 4 are shown as line plots in Fig. 4j, as well as their GO term enrichment) f, Line plots showing smoothed pseudotemporal gene expression and gene activities of selected examples from multiple K-means clusters. RSPO1 is expressed at the progenitor state, showing alignment of pseudotimes. EBF2 and CPLX2 are expressed in neurons and exhibit epigenetic priming with activating marks.



## **Supplementary Fig. 7 Comparing differentiation- and region-specific gene expression between different brain regional branches**

a, Scatter plot showing explained variance over pseudotime (x-axis) and regional branches (y-axis) for gene expression for RNA (top) and gene activity for all histone modifications (see Supplementary Table 16 for details). b, Scatter plot showing the log<sub>2</sub> fold change of gene expression between neurons and progenitors (y-axis) and telencephalon and mesen/rhombencephalon (x-axis). The scatter-plots on the right and top show the same gene set comparing the log<sub>2</sub> fold change of the individual histone modifications. c, Scatter plot showing the log<sub>2</sub> fold change of gene expression between neurons and progenitors (y-axis) and diencephalon and telencephalon (x-axis). The scatter-plots on the right and top show the same gene set comparing the log<sub>2</sub> fold change of the individual histone modifications. d, Scatter plot showing the log<sub>2</sub> fold change of gene expression between neurons and progenitors (y-axis) and mesen/rhombencephalon and diencephalon (x-axis). The scatter-plots on the right and top show the same gene set comparing the log<sub>2</sub> fold change of the individual histone modifications. Using this representation genes that are involved in differentiation (change over pseudotime) and regionalization (change between different brain regions) can be identified. e, Barplot showing the overlap for region and differentiation specific genes detected with each modality (RNA and histone modifications). f, Boxplot (median +/- Q1/Q3) showing gene expression variance over regional branches, pseudotime and both measures combined (n=20641 genes). g, Boxplot (median +/- Q1/Q3) showing gene activity (log(fragment counts per 10k + 1) on gene body +2kb promoter region) variance over regional branches, pseudotime and combined (n=20641 genes). h, Barplot showing the genome-wide distribution of branch-specific peaks for each histone modification. (see Methods Analysis of pseudotemporal and regional variance for details)

Uncropped Western Blot signal shown in Extended Data Fig. 9b



### **Supplementary Fig. 8 Uncropped Western Blot**

Western Blot of cellular extracts of 15 day old organoids shows depletion of H3K27me3 upon treatment with 1  $\mu$ M A395. H3,  $\beta$ -Catenin and Ponceau serve as loading controls.

### **Supplementary Tables:**

Supp\_Data\_1\_all\_branches\_peaks.csv, contains the top 50 peaks of H3K27me3, H3K4me3 and H3K27ac from all regional branches and cell states for the different cell lines. P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_2\_lineage\_peaks\_GREAT\_enrich.tsv, contains GO terms shown in Extended Data Fig. 5a-c. P-values were derived from a hypergeometric test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_3\_TF\_motifs\_regional\_branches.tsv, contains enriched transcriptions factor motifs shown in Extended Data Fig. 5d-f. P-values were derived from a two-sided fisher exact test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_4\_H3K4me3\_bivalent.tsv, contains filtered regions bivalent for H3K27me3 and H3K4me3 between the neuroepithelium and the regional branches (Extended Data Fig. 6a-c). P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_5\_TF\_motifs\_bivalent.tsv, contains enriched transcriptions factor motifs shown in Extended Data Fig. 6d. P-values were derived from a two-sided fisher exact test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_6\_H3K27ac\_switches.tsv, contains filtered regions switching between H3K27me3 and H3K27ac in the regional branches (Fig. 2e). P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_7\_TF\_motifs\_switches.tsv, contains enriched transcriptions factor motifs shown in Extended Data Fig. 6e. P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_8\_all\_peaks\_GREAT\_enrich.tsv, contains GREAT annotation for the clusters shown in Extended Data Fig. 7f. P-values were derived from a hypergeometric test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_9\_all\_peaks\_TF\_motifs.tsv contains enriched transcriptions factor motifs for the clusters shown in Extended Data Fig. 7f. P-values were derived from a two-sided fisher exact test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_10\_ctx\_trajectory\_genes.tsv, all genes and cluster annotation of Fig. 4a.

Supp\_Data\_11\_ctx\_trajector\_4m\_genes.tsv, all genes and cluster annotation Supplementary Fig. 3b.

Supp\_Data\_12\_ctx\_primed\_GRN.tsv, regions and enriched histone modifications underlying Fig. 4h. P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_13\_ctx\_pt\_GO.tsv, contains the clusters and GO analysis of Fig. 4j and Supplementary Fig. 6a. P-values were derived from a hypergeometric test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_14\_dien\_pt\_GO.tsv, contains the clusters and GO analysis of Fig. 4j and Supplementary Fig. 6c. P-values were derived from a hypergeometric test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_15\_mrh\_pt\_GO.tsv, contains the clusters and GO analysis of Fig. 4j and Supplementary Fig. 6e. P-values were derived from a hypergeometric test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_16\_variance\_all.tsv, contains variance scores for all genes during differentiation and regionalization for all modalities Supplementary Fig. 7a

Supp\_Data\_17\_EED\_inhib\_K27me3\_peaks\_log2FC.tsv, contains the log2FC of all H3K27me3 peaks upon treatment including the closest gene to each peak (Fig. 5c).

Supp\_Data\_18\_EED\_inhib\_global\_DE.tsv, contains the global differential expression analysis of the DMSO control and EED inhibition. P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_19\_cluster\_marker\_EED\_inhib.tsv, contains all cluster markers of Fig. 5e. P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_20\_EED\_inhib\_NE\_DE.tsv, contains the differential expression analysis of the DMSO control and EED inhibition for the neuroepithelium clusters (Extended Data Fig. 10a). P-values were derived from a likelihood ratio test and were multiple testing corrected using the Benjamini-Hochberg method.

Supp\_Data\_21\_Antibodies, contains an overview of all antibodies used in the study.