



OPEN

# Temporally divergent regulatory mechanisms govern neuronal diversification and maturation in the mouse and marmoset neocortex

Wen Yuan<sup>1</sup>, Sai Ma<sup>1,2,3</sup>, Juliana R. Brown<sup>1</sup>, Kwanho Kim<sup>1,2,4</sup>, Vanessa Murek<sup>5</sup>, Lucia Trastulla<sup>5</sup>, Alexander Meissner<sup>1,6,7</sup>, Simona Lodato<sup>8,9</sup>, Ashwin S. Shetty<sup>1,4</sup>, Joshua Z. Levin<sup>2,4</sup>, Jason D. Buenrostro<sup>1,7</sup>, Michael J. Ziller<sup>5,10</sup> ✉ and Paola Arlotta<sup>1,4</sup> ✉

**Mammalian neocortical neurons span one of the most diverse cell type spectra of any tissue. Cortical neurons are born during embryonic development, and their maturation extends into postnatal life. The regulatory strategies underlying progressive neuronal development and maturation remain unclear. Here we present an integrated single-cell epigenomic and transcriptional analysis of individual mouse and marmoset cortical neuron classes, spanning both early postmitotic stages of identity acquisition and later stages of neuronal plasticity and circuit integration. We found that, in both species, the regulatory strategies controlling early and late stages of pan-neuronal development diverge. Early postmitotic neurons use more widely shared and evolutionarily conserved molecular regulatory programs. In contrast, programs active during later neuronal maturation are more brain- and neuron-specific and more evolutionarily divergent. Our work uncovers a temporal shift in regulatory choices during neuronal diversification and maturation in both mice and marmosets, which likely reflects unique evolutionary constraints on distinct events of neuronal development in the neocortex.**

The mammalian cerebral cortex contains a great diversity of neurons that differ in their connectivity, neurotransmitter usage, morphology, gene expression and electrophysiological properties<sup>1</sup>. Although recent work has uncovered the molecular states that define different classes of terminally differentiated neurons in the adult brain<sup>2–11</sup>, profiling of cortical neurons over developmental time courses has been limited and largely confined to transcriptional profiling of a few embryonic and neonatal stages<sup>12–15</sup> or of adult timepoints<sup>10,13,16–24</sup>. There is not yet a holistic picture of the molecular dynamics and regulatory landscapes of individual cortical neuron classes, or any other class of mammalian central neurons, over extended trajectories of cell identity acquisition and maturation during postnatal life. This has precluded in-depth understanding of the molecular strategies used by neurons to acquire their identity, to mature and to wire. Similarly, how such molecular strategies might vary across species has not been addressed.

Genomic approaches have been instrumental to studying the molecular logic of cellular differentiation and cell type identity acquisition in the brain<sup>10,25–29</sup> and other tissues<sup>30,31</sup>. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) methods have shown that the number of active enhancer regions varies across development in different cell types; for example, during differentiation of helper T cells, enhancer usage decreases during maturation<sup>32</sup>, whereas cardiomyocytes employ a constant number of enhancers across differentiation<sup>33</sup>. During B cell maturation, the number of active enhancers decreases as they mature from hematopoietic stem cells to terminal cell types<sup>34</sup>. Despite progress,

the regulatory logic that accompanies cell-type-specific differentiation across tissues remains understudied. This is particularly true for the central nervous system, where it is unclear whether common regulatory strategies controlling the development of the great diversity of neural cell types exist.

Here we provide a comprehensive single-cell dataset of defined transcriptional and epigenomic changes in different classes of postmitotic cortical neurons over a time course spanning perinatal ages to adulthood, in both mouse and marmoset. We uncover previously unappreciated divergence in pan-neuronal regulatory mechanisms governing early stages of neuronal development versus later stages of neuronal circuit formation. These distinct regulatory modes represent a common strategy across all cortical neuron types and are conserved between mice and non-human primates.

## Results

**Early and late stages of cortical pyramidal neuron postmitotic development use divergent regulatory programs.** Even after becoming postmitotic, cortical neurons undergo extensive development, from establishment of subtype identity to postnatal refinement of terminally differentiated features. It is largely not understood whether regulatory strategies at play during postmitotic development remain constant over postnatal life.

To understand these regulatory strategies, we first applied inducible Cre mouse lines to examine two major classes of neocortical pyramidal neurons: *Cux2*-lineage<sup>35</sup> layer 2/3 (L2/3) callosal projection neurons (CPNs)<sup>1</sup>, which are involved in associative functions

<sup>1</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. <sup>2</sup>Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>3</sup>Department of Biology and Koch Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>4</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>5</sup>Max Planck Institute of Psychiatry, Munich, Germany. <sup>6</sup>Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>7</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA.

<sup>8</sup>Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy. <sup>9</sup>IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.

<sup>10</sup>Department of Psychiatry, University of Münster, Münster, Germany. ✉e-mail: [ziller@uni-muenster.de](mailto:ziller@uni-muenster.de); [paola\\_arlotta@harvard.edu](mailto:paola_arlotta@harvard.edu)

and are the most recently evolved population of cortical neurons (henceforth, Cux2 CPNs), and *Tle4*-lineage<sup>36</sup> layer 6 (L6) corticothalamic projection neurons (CThPNs)<sup>1</sup>, which are responsible for integration of sensory and motor information (henceforth, Tle4 CThPNs). We isolated these neuronal subtypes across a time course spanning the acquisition of class-specific neuronal identity and early neuronal development (embryonic day (E) 18.5 and postnatal day (P) 1, P3 and P7), through periods of cortical plasticity, neuronal maturation and integration into cortical circuits (P21 and P48) (Fig. 1a, Extended Data Fig. 1 and Supplementary Fig. 1). Labeled neurons from dissected somatosensory and motor cortex (ten animals per library, five male and five female, from two litters) were isolated by fluorescence-activated cell sorting (FACS) and profiled in bulk for gene expression by RNA sequencing (RNA-seq), for DNA methylation (DNAm) by whole-genome bisulfite sequencing (WGBS) and for open chromatin by ATAC-seq. Two biological replicates were performed for each age and neuron type for each assay.

We identified features (transcribed genes, differentially accessible chromatin peaks and differentially methylated regions (DMRs)) that were dynamic over age or between cell types (examples in Supplementary Fig. 2) and applied k-means clustering to group features with similar patterns (Fig. 1c–e, Supplementary Fig. 3 and Supplementary Tables 1–3). For all datasets, 40–60% of dynamic features were assigned to clusters that were associated with developmental stage and independent of neuronal subtype (Fig. 1c–f). These shared, developmentally regulated clusters fell into two major categories: those predominantly active (transcriptionally upregulated, accessible or hypomethylated) at embryonic and/or early postnatal ages (E18.5 to P7; early developmental, yellow bars in Fig. 1c–e) and those predominantly active at weaning and older ages (P21 to P48; late developmental, light green bars in Fig. 1c–e).

A smaller proportion of dynamic features showed neuron class-specific patterns. As expected, these clusters included known molecular markers of CPNs and CThPNs<sup>37,38</sup> (Fig. 1c–e). Notably, although class-specific clusters accounted for only 23% of dynamic transcriptional features, they comprised 34% (DNAm) to 45% (ATAC) of dynamic epigenetic features (Fig. 1f), indicating higher neuronal subtype specificity of epigenetic changes, in agreement with findings that epigenetic signatures may be particularly powerful in discriminating neuronal subclasses<sup>16,39</sup>.

We applied multi-dimensional scaling (MDS) to visualize the relative distance between the high-dimensional transcriptional and epigenetic landscapes over time. Despite widespread changes over development, these two classes of neurons showed only limited changes in overall similarity with time in either gene expression or open chromatin profiles (Fig. 1g and Supplementary Fig. 3b). In contrast, the DNA methylation landscape became more divergent between neuronal subtypes over postnatal life ( $P < 0.03329$ , one-sided *t*-test; Fig. 1g and Supplementary Fig. 3b). This increase was not associated with global changes in methylation or expression of DNA methylases (Extended Data Fig. 2a,b) but, rather, with changes in distribution patterns across the genome. DNA methylation increased over time at genes and gene regulatory elements (GREs, inferred from open chromatin sites) characteristic of other cell types and earlier developmental stages (Extended Data Fig. 2c,d), consistent with its known role in stabilizing silencing of inappropriate transcriptional programs<sup>40</sup>. Similarly, analysis of CpA methylation (the dominant form of non-CpG methylation in mammals) identified 11,150 DMRs across cell types and developmental time, although with small effect sizes (Extended Data Fig. 2f). Notably, CpA methylation exhibited dynamic patterns that were similar to those found for CpG methylation but showed only minimal overlap with CpG DMRs (<10%).

Next, to identify strategies for genome regulation that are common to pyramidal neuron classes across developmental time, we examined clusters whose temporal dynamics were shared

between both cell types (shared-early and shared-late clusters; Fig. 1c–e). This uncovered a pronounced temporal transition in regulatory dynamics, across multiple modalities, between shared programs of early (E18.5 to P7) versus late (P21 to P48) postmitotic neuronal development.

The shared-early clusters contained genes typical of early differentiation events, such as *Sox4*, *Sox11* and *Apc2*, whereas the shared-late clusters contained genes controlling later processes such as synaptic function, including *Egr3*, *Syp* and *Nefm*. We validated temporal expression patterns in the cortex for selected example genes using the Allen Institute mouse brain in situ hybridization (ISH) database<sup>41</sup> (Extended Data Figs. 3a,b and 4a,b) and a Slide-seq spatial transcriptomics dataset of young and adult mouse cortex<sup>42</sup> (Extended Data Figs. 3c and 4c).

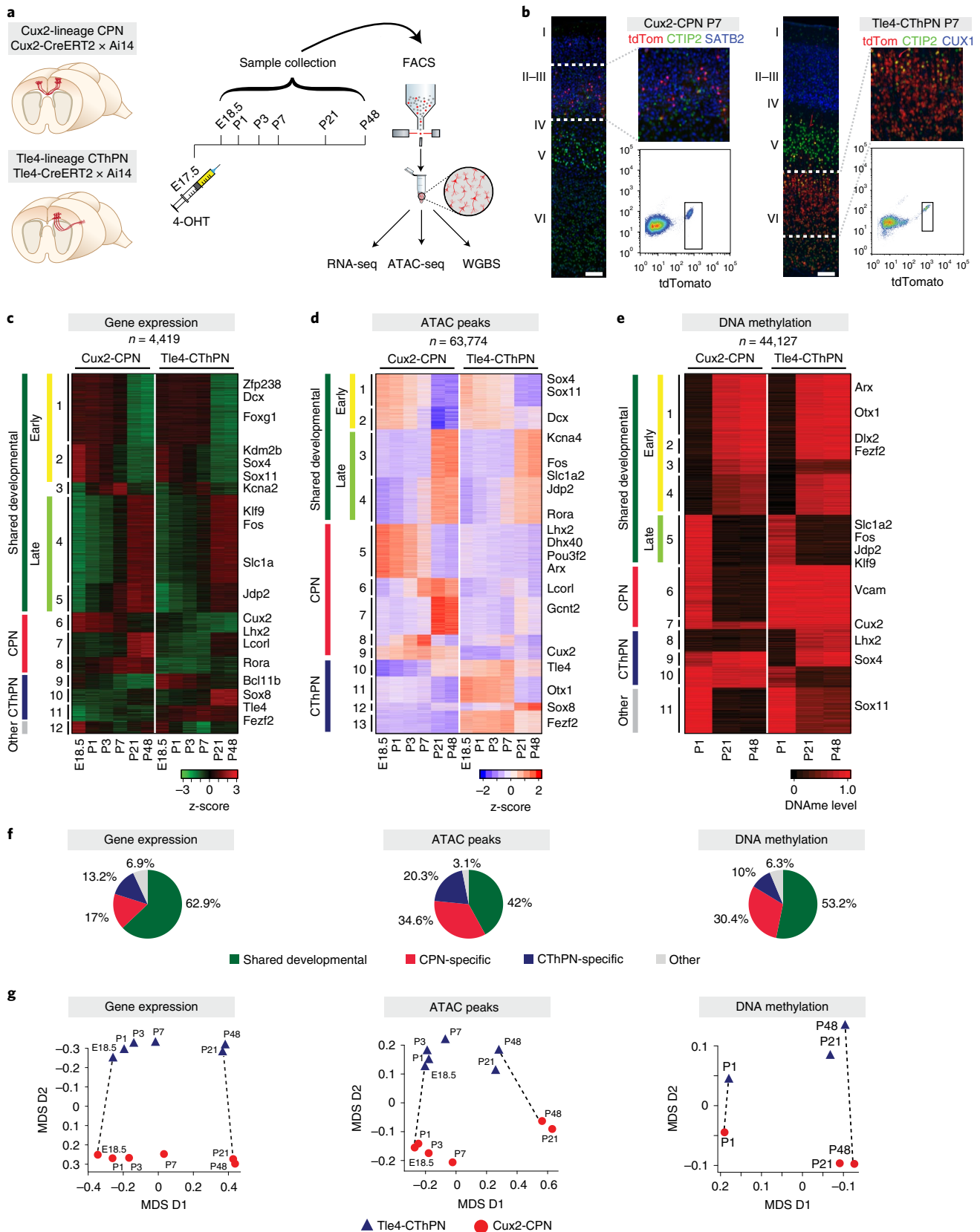
To examine whether these temporal patterns were conserved in humans, we leveraged human developing brain transcriptomic data from BrainSpan<sup>43</sup>. About 85% of the genes in both the shared-early (1,129/1,334) and shared-late (1,234/1,446) gene clusters had orthologs that were expressed in the human data, and most of these showed similar temporal dynamics as in mouse (Supplementary Fig. 4).

We then systematically characterized each of these distinct classes of genes and GREs (Fig. 2a). We first assessed the temporal, cell type and tissue specificity of gene expression across the transcriptional clusters (Fig. 2b), for both all genes in each cluster (Fig. 2c) and for transcription factors (TFs) only. We examined expression across a wide range of tissues and cell types, using 77 mouse brain regions and developmental timepoints from the Allen Brain Atlas<sup>41</sup> and 294 mouse cell types and tissues from the FANTOM5 project<sup>44</sup> (Fig. 2b,c and Extended Data Fig. 5a). For both all genes and for TFs only, the shared-late genes showed significantly higher temporal and/or cell type specificity in both tissue panels compared to the shared-early genes ( $P \leq 2.576 \times 10^{-7}$  for all comparisons, one-sided Mann–Whitney test). This suggests that the pan-neuronal genes and TF programs that are active in perinatal postmitotic neurons reflect more broadly used developmental processes compared to those active at later ages.

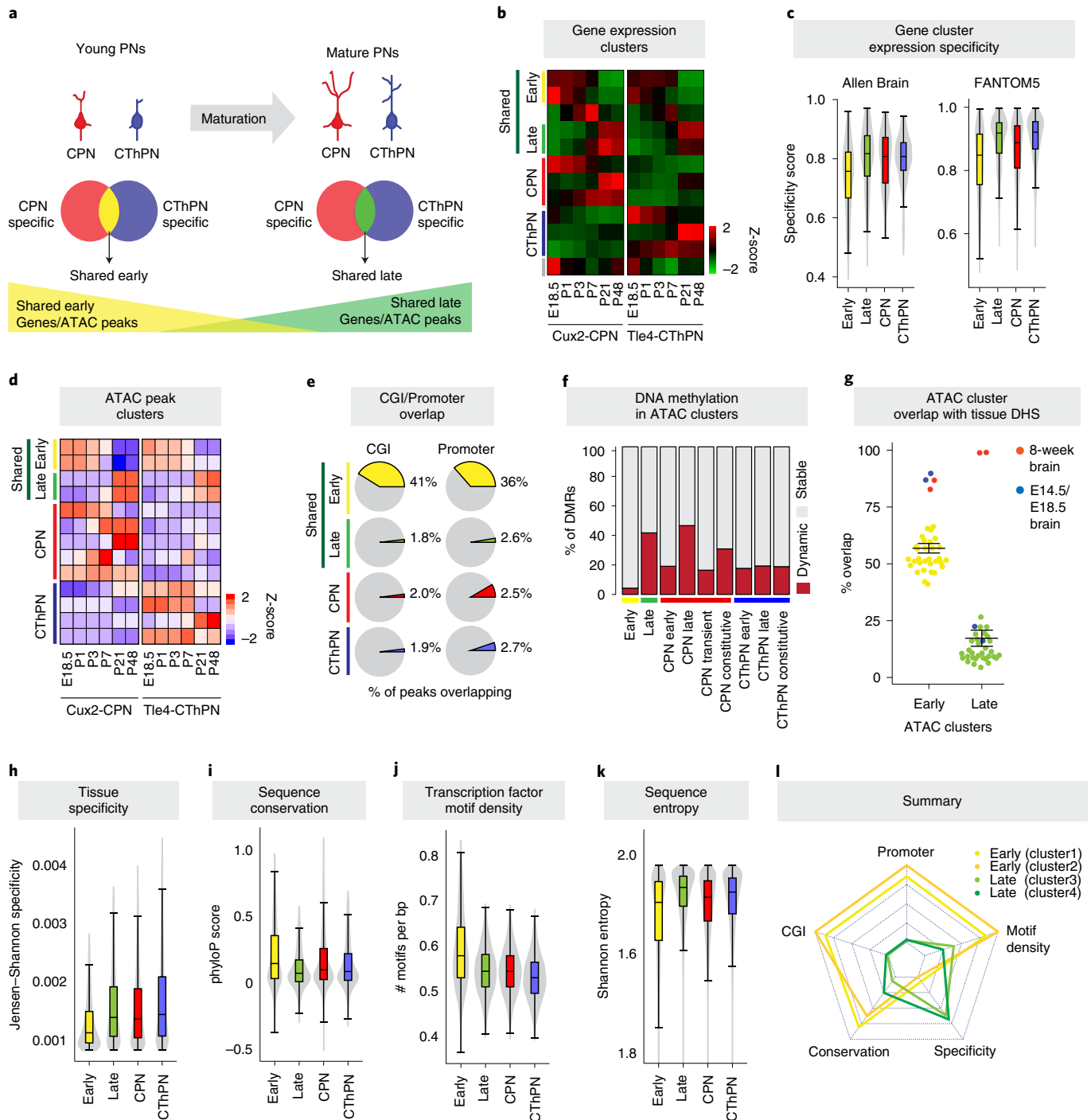
Interestingly, although some cell-type-specific clusters also showed temporal correlation, early-active and late-active cell-type-specific clusters did not show significant differences in tissue specificity (Fig. 2b,c), indicating that the pan-neuronal and the neuron-type-specific gene programs use different regulatory logic during neuron maturation.

To investigate the relationship between gene expression and ATAC peak cluster dynamics, we calculated the overlap between genes in the transcriptomic clusters and genes putatively regulated by GREs in the ATAC clusters (defined as peaks within 100 kilobases (kb) of the transcriptional start site (TSS)). We found that the shared-early gene expression clusters were most strongly associated with the shared-early ATAC clusters (Benjamini–Hochberg-corrected  $P = 5.748203 \times 10^{-36}$ , Fisher's exact test) and similarly for the shared-late gene expression and ATAC clusters (Extended Data Fig. 5c), indicating that the shared developmental transcriptional and epigenetic programs affect similar sets of genes.

Next, we more finely characterized the putative GREs identified from our ATAC-seq profiles. We found that GREs in the shared-early ATAC clusters were more than ten-fold enriched for annotated promoter regions (within 1 kb upstream or downstream of the TSSs) and CpG islands (CGIs) compared to the shared-late clusters ( $P < 2.2 \times 10^{-16}$ , Fisher's exact test; Fig. 2d,e and Extended Data Fig. 5b). CGIs are strongly associated with TSSs and play an important role in gene regulation<sup>45</sup>. Conversely, members of the shared-late ATAC clusters were largely TSS distal and showed lower frequency of overlap with CGIs and annotated promoters (2.6% and 1.8%, respectively; Fig. 2d,e and Extended Data Fig. 5b), suggesting that they disproportionately function as distal regulatory



**Fig. 1 | Profiling of genetically defined cortical pyramidal neuron classes.** **a**, Schematic of experimental design. **b**, Representative coronal sections ( $n \geq 3$  biological replicates) showing correct laminar location of tdTomato<sup>+</sup> cells in the somatosensory cortex and FACS plots of Tle4-CThPNs and Cux2-CPNs at P7. Scale bar, 100  $\mu$ m. Also see Extended Data Fig. 1. **c–e**, Developmental dynamics of differentially expressed genes (**c**), differentially enriched ATAC-seq peaks (**d**) and DMRs (**e**). **f**, Fraction of dynamic features classified into each overall category. **g**, 2D MDS plots for each modality.



**Fig. 2 | Divergent regulatory principles of early and late stages of neuronal development.** **a**, Schematic of analyses. **b**, Summary of gene expression clusters from Fig. 1c. **c**, TF expression specificity within the brain for TFs in different categories of cluster, from expression data from the Allen Brain Atlas (left). Higher values indicate more specific expression. TF expression specificity across 397 mouse cell types from the FANTOM5 project (right). Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5x interquartile range from box. **d**, Summary of ATAC peak clusters from Fig. 1d. **e**, Fraction of open chromatin regions (ATAC-seq peaks) in different categories of clusters that overlap CGIs or annotated promoter regions. **f**, Fraction of DMRs in ATAC clusters of different categories that are static or dynamic over the time course. **g**, Fraction of ATAC-seq peaks from different classes of ATAC clusters that overlap DNase hypersensitivity sites (DHSs) across 35 cell types from the mouse ENCODE project. Adult (8-week) whole-brain and telencephalon samples highlighted in red; embryonic (E14.5 and E18.5) brain samples are highlighted in blue. Also see Extended Data Fig. 5f. Error bars: mean  $\pm$  s.e.m. **h**, Open chromatin specificity across ATAC peak clusters ( $n_{\text{Early}} = 10,488$ ,  $n_{\text{Late}} = 17,570$ ), as the concordance of ATAC peaks with an scATAC-seq panel of 85 mouse primary tissues and cell types<sup>17</sup>. Box plot as in **c**. **i**, Distribution of phyloP sequence conservation across all placental mammals for early developmental and late developmental ATAC clusters. Box plot as in **c**. **j**, Density of known TF binding motifs within ATAC clusters of different types. Box plot as in **c**. **k**, Average sequence entropy within ATAC clusters of different types. Box plot as in **c**. **l**, Summary of characteristics from **c** and **g-j** for each of the shared developmental ATAC peak clusters individually. Each arm of the plot represents an individual metric (center = low, edge = high). Also see Extended Data Fig. 5b.

elements, such as enhancers or insulators. To confirm the enhancer identity of the candidate GREs in our datasets, we examined putative activated enhancers (H3K27ac-positive regions) in mouse brain samples from the mouse ENCODE project<sup>46</sup> (Extended Data Fig. 5g). We observed enriched overlap of H3K27ac-positive regions from E14.5 whole-brain samples with candidate GREs from the ATAC shared-early clusters and between H3K27ac-positive regions from adult cortex samples with candidate GREs from the ATAC shared-late clusters, supporting the identity as enhancers of the GREs in these clusters.

In agreement with our findings in the transcriptional dataset, the cell-type-specific ATAC clusters did not show a similar transition from promoter to enhancer/insulator usage during neuron maturation. Additionally, the cell-type-specific ATAC clusters showed low overlap with CGIs and annotated promoters regardless of developmental dynamics and, thus, likely act as enhancer regions (Fig. 2d,e and Extended Data Fig. 5b,g). This further supports the suggestion that, during neuronal maturation, neurons use different mechanisms to regulate shared versus cell-type-specific maturation processes.

Given the close relationship between chromatin accessibility and DNase state, we examined the concordance of these states across cell types and developmental time. We found significant overlap between clusters with similar dynamics, where regions losing open chromatin gained DNase and vice versa, for shared-early, shared-late and cell-type-specific clusters (Extended Data Figs. 2g,h and 5d). However, the fraction of open chromatin regions exhibiting temporally dynamic DNase patterns was significantly higher in the shared-late ATAC cluster compared to the shared-early cluster (41.72% versus 4.29%; Fig. 2f and Extended Data Fig. 5e), consistent with the elevated frequency of promoter and CGI regions found in the shared-early clusters and the much higher frequency of distal putative regulatory elements in the shared-late clusters (Fig. 2d,e and Extended Data Fig. 5b,c). Promoter regions are well-known to be much less prone to DNase changes over development, whereas distal differentially accessible sites frequently coincide with enhancer regions and TF binding sites that exhibit highly dynamic DNase patterns<sup>47</sup>. This observation suggests that DNA methylation plays a greater role in regulating the shared-late developmental programs by mediating stable silencing of distal putative regulatory elements.

To confirm the biological activity of the predicted GREs, we selected three open chromatin regions each for the *Sox4* and *Sox11* genes and silenced them using the enCRISPRi system<sup>48</sup> in an in vitro differentiated neuroectodermal cell line (NE-4C; Methods). Inactivation of four of the six predicted GREs resulted in downregulation of their respective gene, indicating that these genomic regions have properties of enhancers and are able to regulate gene expression (Extended Data Fig. 6).

Previous work has suggested that cell-type-specific accessible chromatin sites are preferentially localized in putative enhancer regions compared to promoter regions<sup>4,49–51</sup>. We, therefore, evaluated the tissue specificity of early-active versus late-active GREs across a panel of DNase hypersensitivity sites in 35 adult and embryonic mouse primary tissues and cell types from the ENCODE database<sup>46</sup>. GREs in shared-late developmental epigenetic clusters were, on average, found in an open state in significantly fewer tissues than GREs in shared-early developmental clusters (ATAC clusters early versus late:  $P < 2.2 \times 10^{-16}$ , Mann–Whitney test; Fig. 2g and Extended Data Fig. 5f). Notably, late-active ATAC clusters were highly enriched for chromatin regions that are accessible in adult (8-week) mouse brain tissues but not for those open in embryonic brain, adult cerebellum, adult and perinatal retina or any non-central nervous system tissue. In contrast, GREs in early-active ATAC clusters showed high overlap with adult and embryonic brain but also showed broad associations with open chromatin regions

in many other cell types and tissues (Fig. 2g and Extended Data Fig. 5f). The ratio of accessible sites suggests that the shared-late active regions are highly enriched for GREs that are specific to the adult brain, consistent with the enrichment of neuron-specific Gene Ontology (GO) processes in these clusters (Supplementary Table 4).

We then compared the average tissue specificity of accessible sites in the shared-early and shared-late ATAC clusters across a single-cell ATAC-seq (scATAC-seq) panel of 85 mouse primary tissues and cell types<sup>17</sup> (Methods). The shared-early active sites showed lower average specificity ( $P < 2.2 \times 10^{-16}$ , Mann–Whitney test; Fig. 2h), again indicating that these sites are more widely used across tissues.

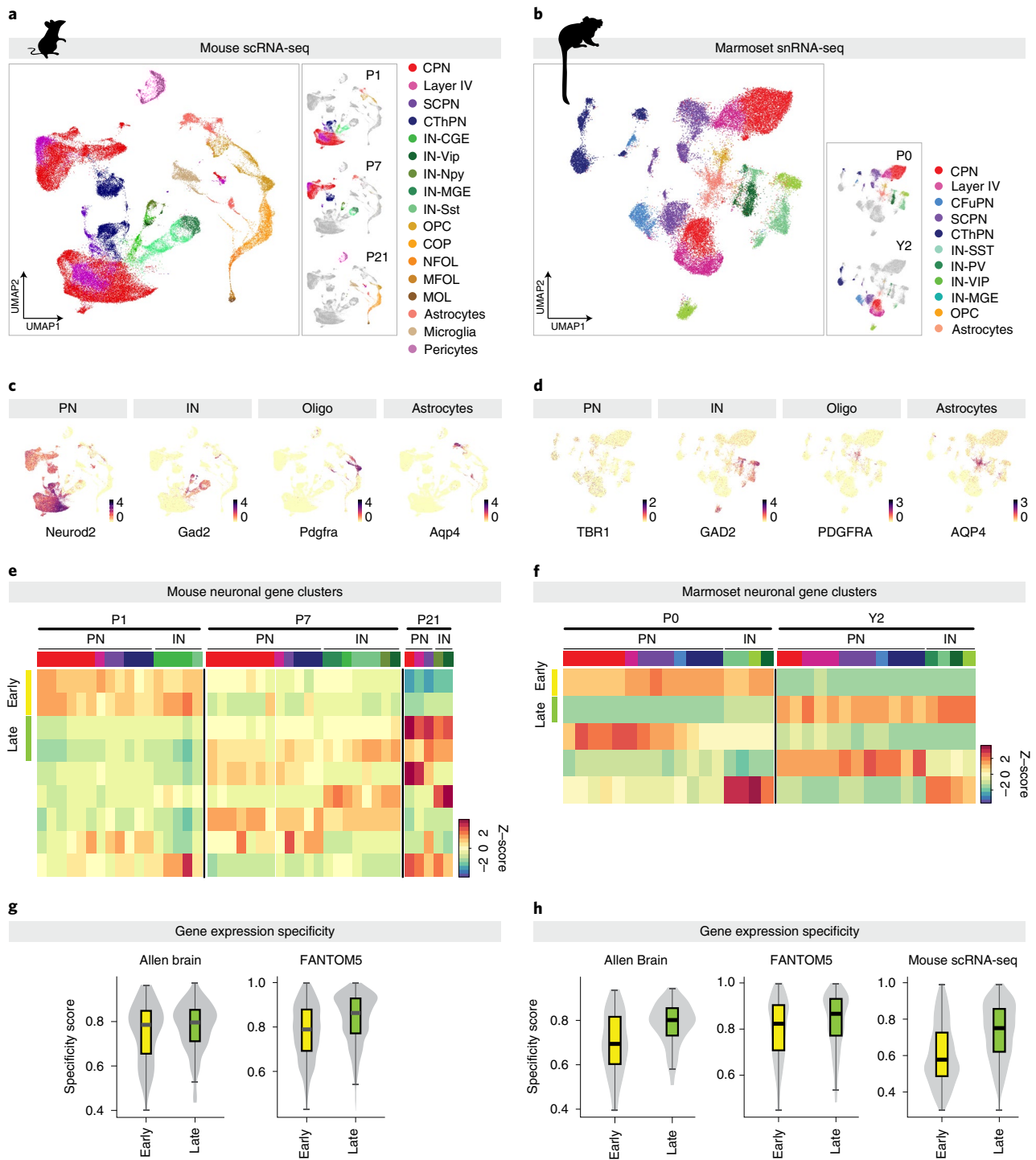
To validate that the shared-early and shared-late accessible chromatin sites from our ATAC-seq data confer different degrees of restriction in expression, we examined the activities of these sites in the VISTA enhancer dataset, in which enhancer activity is visualized by a LacZ reporter assay driven by non-coding DNA fragments in transgenic mice<sup>52</sup>. We examined 16 representative chromatin sites selected from our ATAC-seq data in E11.5 mouse embryos in the VISTA dataset. We found that the early ATAC-seq open regions predominantly drove broad LacZ expression in multiple tissues throughout the organism, whereas the late ATAC-seq open regions drove more selective expression, with LacZ signal visible in fewer tissues (Extended Data Fig. 7), consistent with our bioinformatic analysis (Fig. 2h).

Previous studies of DNA methylation have reported that sequence conservation varies for DMRs characteristic of different tissue types<sup>53</sup> and that DMRs specific to excitatory cortical neurons are less conserved than those specific to interneurons<sup>16</sup>. Given these observations and the more ubiquitous activity of genes and GREs in the shared-early developmental clusters, we hypothesized that these regions might be under different degrees of evolutionary constraint. Quantifying sequence conservation across placental mammals of the ATAC peak regions in the shared developmentally regulated ATAC clusters found significantly higher conservation of shared-early elements ( $P < 2.2 \times 10^{-16}$ , Mann–Whitney test; Fig. 2i and Extended Data Fig. 5b). Shared-early active GREs also showed a higher density of TF binding motifs ( $P < 2.2 \times 10^{-16}$ , Mann–Whitney test; Fig. 2j) and lower sequence entropy (a metric of sequence constraint;  $P < 2.2 \times 10^{-16}$ , Mann–Whitney test; Fig. 2k), both broadly associated with increased CpG density. Interestingly, across all of these metrics, the neuron subtype-specific gene and GRE clusters largely resembled the shared-late developmental clusters (Fig. 2b–e,h–i and Extended Data Fig. 5b).

Together, these findings (summarized in Fig. 2l) indicate that the shared-early GREs use more widely shared regulatory mechanisms that are consequently under greater evolutionary constraint. In contrast, the shared-late GREs, which have more restricted tissue usage, may be more amenable to variation and may employ more species-specific or evolutionarily recent mechanisms. Notably, we did not find a comparable difference between the cell-type-specific early and late gene and GRE clusters, suggesting that this change in regulatory mechanisms applies specifically to genes and GREs involved in processes shared by all neurons.

### Temporal divergence in global regulatory strategies is a conserved principle of cortical neuron maturation across species.

Next, we sought to investigate whether these regulatory principles are common to all neuronal classes in the cortex and, furthermore, whether cortical neurons in other species follow the same principles. Marmosets (*Callithrix jacchus*) are an attractive non-human primate model for neurobiology, offering more human-like brain anatomy and circuitry, cognitive capacities and behavioral repertoires<sup>54,55</sup>. We performed single-cell (mouse) or single-nucleus (marmoset) RNA sequencing (scRNA-seq and snRNA-seq) and single-cell ATAC sequencing (scATAC-seq) on unfractionated cortical tissue from



**Fig. 3 | scRNA-seq demonstrates a developmental shift in specificity of shared gene expression programs across multiple neuronal subclasses in both mouse and marmoset.** **a**, UMAP representation of gene expression profiles from 60,989 single cells from mouse cortex at P1, P7 and P21, color-coded by major cell type. Left: UMAP plots showing cell distribution by age. **b**, UMAP representation of 36,592 single nuclei from marmoset cortex at P0 and Y2, color-coded by major cell type. Left: UMAP plots showing cell distribution by age. **c**, Representative marker genes for major cell types in the mouse data. Also see Supplementary Fig. 6a,c. **d**, Representative marker genes for major cell types in the marmoset data. Also see Supplementary Fig. 8a,c. **e**, Developmental dynamics of clusters of differentially expressed genes across the mouse excitatory and inhibitory neuronal populations (cell type indicated by color-coded bar at top, corresponding to colors in **a**). **f**, Developmental dynamics of clusters of differentially expressed genes across the marmoset excitatory and inhibitory neuronal populations (cell type indicated by color-coded bar at top, corresponding to colors in **b**). **g**, Mouse gene expression specificity in the shared-early and shared-late gene clusters, within the mouse brain (from expression data from the Allen Brain Atlas; left), and across 397 mouse cell types (from the FANTOM5 project; right). Higher values indicate more specific expression. Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5 $\times$  interquartile range from box. **h**, Marmoset gene expression specificity in the shared-early and shared-late gene clusters, within the mouse brain (from expression data from the Allen Brain Atlas; left), across 397 mouse cell types (from the FANTOM5 project; center) and across cell populations from our mouse single-cell dataset (right). Higher values indicate more specific expression. Box plot as in **g**.

these two species across early and late stages of postnatal development. For mice, we profiled combined somatosensory and motor cortex at three selected ages: P1, when early active gene and GRE clusters predominate; P7, an intermediate stage; and P21, when late active gene and GRE clusters predominate. For marmoset, we profiled somatosensory cortex at neonatal (P1 to P2) and adult ages (2 years (Y2) for scRNA-seq and Y7 to Y8 for scATAC-seq, based on limited tissue availability for this species). We expect that the Y2 and Y7/8 marmoset samples represent comparable adult neurons, given the 12–16-year life span of this species<sup>56</sup>.

We first examined programs of gene expression across the two species. After quality control and filtering, the final dataset contained a total of 60,989 mouse and 36,592 marmoset cells across the different timepoints (Fig. 3a,b and Supplementary Figs. 5–8). We performed cell clustering and assigned cell identity based on expression of known canonical marker genes (Fig. 3a–d and Supplementary Figs. 6 and 8). This identified major neuronal and glial cell populations, including different populations of pyramidal neurons (CPNs, subcerebral projection neurons (SCPNs) and CThPNs) and interneurons, astrocytes and oligodendrocyte-lineage populations, which expressed classical cell type marker genes, including *Neurod2* and *Tbr1* (glutamatergic neurons), *Gad2* (GABAergic interneurons), *Pdgfra* (oligodendrocyte lineage) and *Aqp4* (astrocytes) (Fig. 3c,d). For both mice and marmoset, clusters showed separation by both age and cell type, with projection neuron populations being predominantly separated by age (Fig. 3a,b and Supplementary Figs. 5c and 7c), consistent with our findings that most changes in the mouse bulk sequencing datasets were developmentally related rather than neuron subtype-specific (Fig. 1).

We then identified shared developmental and cell-type-specific gene clusters using unsupervised k-means clustering (Methods), following the same strategy used for the bulk datasets. For each species, we selected all pyramidal neuron and interneuron populations, identified genes that showed differential expression across the sc/snRNA-seq datasets and clustered them by their expression pattern across age and cell type (Fig. 3e,f). For mice, we identified two shared-early and two shared-late developmentally regulated pan-neuronal gene clusters (that is, shared across all pyramidal and interneuron populations), and, for marmoset, we identified one early and one late pan-neuronal gene cluster (Fig. 3e,f).

Next, we analyzed tissue specificity and cell type specificity for TFs in these clusters as we did for the bulk RNA-seq and found that, for both species, the shared-late developmental clusters showed greater TF specificity across both the Allen (brain) and FANTOM5 (tissue) datasets than the shared-early clusters (Fig. 3g–h and Extended Data Fig. 8b,c,e,f). In addition, the marmoset shared-late gene cluster showed greater cell type specificity across our mouse scRNA-seq dataset compared to the shared-early clusters (Fig. 3h). Notably, although the pyramidal and interneuron cell types contained multiple subclusters (both separating different subclasses and further dividing some subclasses), we did not observe striking differences

within the subclusters of individual types, indicating that heterogeneity within neuron populations does not explain the gene clustering results. Collectively, these data indicate that the shift in global regulatory principles observed in CPN and CThPN applies to all cortical neuron subtypes and is conserved in the non-human primate cortex.

To examine regulatory characteristics of GREs across cell types and species, we performed scATAC-seq at neonatal and juvenile/adult ages for both species (Fig. 4a,b and Methods). After quality control and filtering, the final dataset included 19,145 mouse and 15,919 marmoset cells (Fig. 4a,b and Supplementary Figs. 9 and 10). Cell type identities were assigned to cell clusters based on inferred expression of the same panel of known marker genes used in the sc/snRNA-seq analyses (Supplementary Fig. 11). Similarly to the scRNA-seq data, neuron populations were predominantly separated by age in both species (Supplementary Figs. 9b and 10b).

We then collapsed differentially accessible ATAC peak regions by cell cluster and clustered these pseudo-bulk profiles. For mice, we identified one pan-neuronal early-peak cluster and one pan-neuronal late-peak cluster, as well as two late-peak clusters shared across pyramidal neuron populations but not interneurons (Fig. 4c). To confirm the biological relevance of these regions, we compared them to histone acetylation chromatin immunoprecipitation followed by sequencing (ChIP-seq) enrichment data from the ENCODE project<sup>46</sup>. The mouse shared-early and shared-late clusters showed high overlap with H3K27ac-enriched regions from embryonic mouse whole brain and adult mouse cortex, respectively, supporting their status as active enhancer elements (Extended Data Fig. 9g,h).

Notably, the dynamics of the scATAC peak clusters were correlated with DNAm dynamics (from the mouse bulk dataset): early pan-neuronal scATAC peaks remained mostly unmethylated, whereas late pan-neuronal peaks lost methylation over developmental time (Extended Data Fig. 10e). Interestingly, glia-specific open chromatin regions (such as clusters 2 and 4) showed relatively low DNAm levels in P1 CPNs and CThPNs and only gained DNAm in these neurons during maturation (Extended Data Fig. 10f), consistent with the progressive silencing of alternative lineage programs by DNAm<sup>57</sup>. Interneuron-specific open chromatin regions (cluster 7), however, showed consistently high DNAm levels in CPNs and CThPNs across all ages, suggesting that these programs may need to be silenced at an earlier stage of pyramidal neuron development.

For marmoset, we likewise identified one pan-neuronal early-peak cluster and one late-peak cluster shared across pyramidal neurons but not interneurons. However, we did not identify a pan-neuronal late-peak cluster (Fig. 4e). This may suggest that, in marmoset, adult regulatory programs may diverge more between interneurons and projection neurons than they do in mice, which may reflect greater neuronal specialization in more evolutionarily advanced cortices.

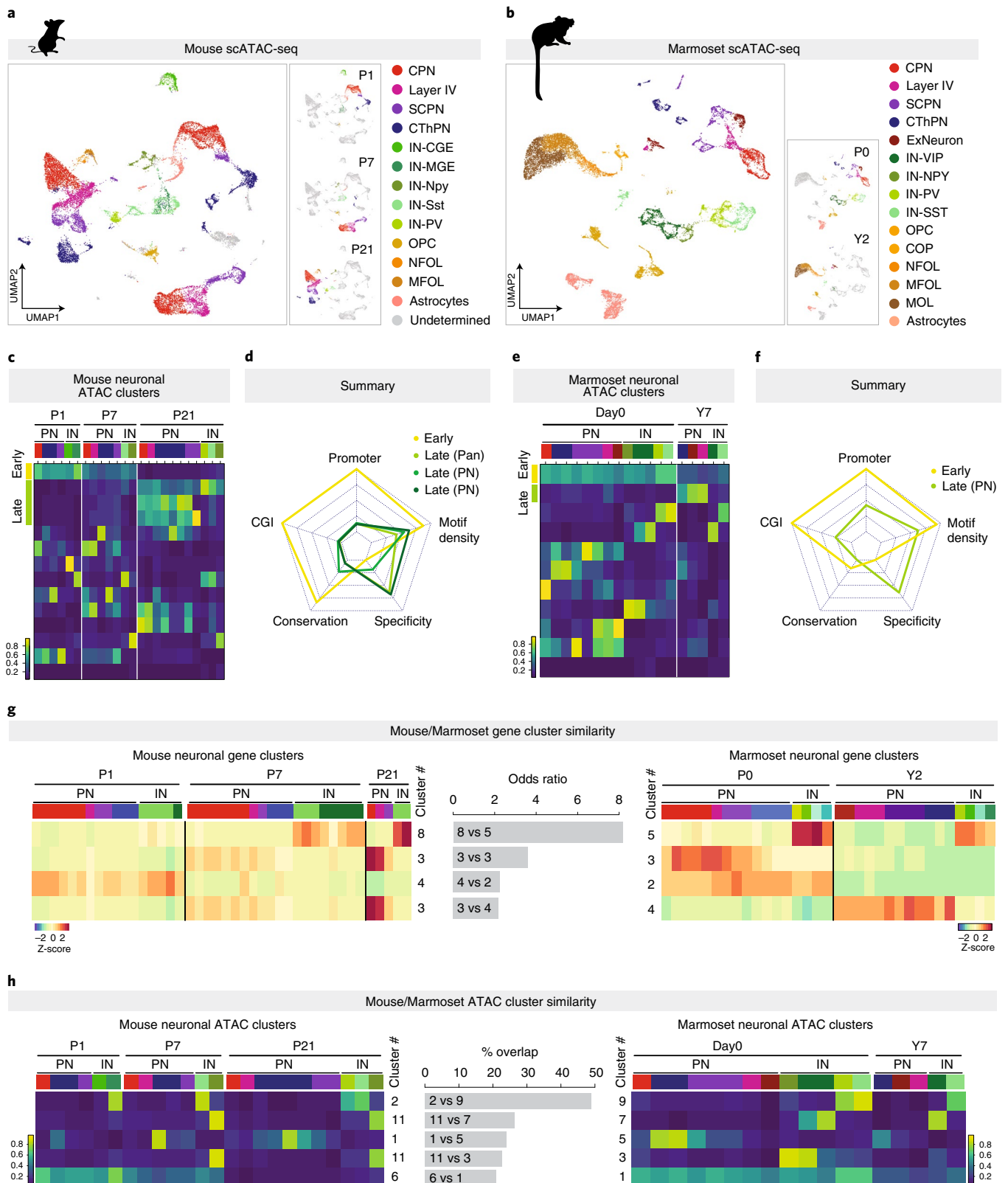
We repeated the ATAC peak characterization previously performed on the mouse Cux2-CPN/Tle4-CThPN bulk sequencing dataset on the scATAC-seq datasets. Consistent with the bulk data,

**Fig. 4 | Developmental shift in gene regulatory principles is common to multiple neuronal subclasses and is conserved across mouse and marmoset.**

**a**, UMAP representation of ATAC chromatin accessibility profiles from 19,145 single cells from mouse cortex at P1, P7 and P21, color-coded by major cell type. Left: UMAP plots with cell distribution by age. **b**, UMAP representation of ATAC chromatin accessibility profiles from 15,919 single cells from marmoset cortex at P0 and Y2, color-coded by major cell type. Left: UMAP plots with cell distribution by age. **c**, Developmental dynamics of clusters of differentially accessible ATAC peaks across the mouse excitatory and inhibitory neuronal populations (cell type indicated by color-coded bar at top, corresponding to colors in **a**). Color scale: average peak normalized read count. **d**, Summary of CGI overlap, promoter overlap, TF motif density, tissue specificity and sequence conservation for each of the mouse shared developmentally regulated clusters, as in Fig. 2k. Also see Extended Data Fig. 9a–c. **e**, Developmental dynamics of clusters of differentially accessible ATAC peaks across the marmoset excitatory and inhibitory neuronal populations (cell type indicated by color-coded bar at top, corresponding to colors in **b**). Color scale: average peak normalized read count. **f**, Summary of CGI overlap, promoter overlap, TF motif density, tissue specificity and sequence conservation for each of the marmoset shared developmentally regulated clusters. Also see Extended Data Fig. 9d–f. **g**, Overlap between genes in the mouse and marmoset single-cell gene expression clusters, showing the four most similar pairs. Also see Extended Data Fig. 10b. **h**, Overlap between accessible regions in the mouse and marmoset single-cell ATAC chromatin accessibility clusters, showing the five most similar pairs. Also see Extended Data Fig. 10d. Color scale: average peak normalized read count.

we found that, for both species, the pan-neuronal/pan-projection neuron (pan-PN) early active peak clusters showed greater enrichment for promoters and CGI; increased density of binding motifs (mouse not significant, marmoset early versus late:  $P < 2.2 \times 10^{-16}$ ); lower tissue specificity (mouse early versus pan-neuronal late:  $P = 0.04574$ , mouse early versus pan-PN late:  $P < 2.2 \times 10^{-16}$ ,

marmoset early versus late:  $P = 2.282 \times 10^{-8}$ ); and higher sequence conservation (mouse early versus pan-neuronal late:  $P < 2.2 \times 10^{-16}$ , mouse early versus pan-PN late:  $P < 2.2 \times 10^{-16}$ , marmoset early versus late:  $P = 0.0008774$ ) (Fig. 4d,f and Extended Data Fig. 9b,c,e,f). Similarly to the sc/snRNA-seq data, we did not observe significant differences between more finely subclustered neuronal subtypes.





Together, these data show that the transition in pan-neuronal developmental programs between broadly used, more highly conserved regulatory elements at earlier stages of postmitotic development versus more cell-type-specific and tissue-specific regulatory elements at later stages holds true across different neuronal classes and is conserved in the non-human primate cortex. The data points at these regulatory principles as broadly generalized properties of neuronal maturation across species.

To compare the similarity of the shared developmental programs between species, we examined the overlap between mouse and marmoset gene and ATAC peak clusters (Fig. 4g,h). We found that the shared-early gene clusters were among the most highly similar pairs of clusters between species; similarly, the shared-early ATAC peak clusters were among the most similar pairs between species (Fig. 4g,h). These findings indicate that general programs of early pan-neuronal development are more frequently shared between species compared to later pan-neuronal programs, which show greater species specificity. This observation is consistent with the broader sequence-level conservation found in shared-early clusters identified from both the bulk pyramidal neuron and scATAC-seq datasets. Interestingly, interneuron-specific clusters were the most highly similar between species for both the gene and ATAC-seq datasets, suggesting that interneuron-specific developmental programs are more highly conserved (Fig. 4g,h).

Lastly, we also compared the mouse bulk ATAC-seq and snATAC-seq clusters to an available snATAC-seq dataset of human fetal cerebrum<sup>58</sup>. This analysis revealed that the mouse shared-early clusters had significant overlap with the human fetal open chromatin regions (Supplementary Fig. 12), suggesting evolutionary conservation of chromatin dynamics at these sites.

In sum, our analysis uncovered a temporal shift in regulatory principles—across multiple modalities and between generalized programs of early (perinatal) and late (juvenile/adult) postmitotic neuronal development—that is found across neuronal subtypes and evolutionarily conserved in rodents and non-human primates.

## Discussion

The neocortex contains a great diversity of neuronal classes that are born during embryogenesis but undergo substantial postnatal maturation to acquire their adult features. Here we have defined a global outline of the regulatory principles underlying key steps of late-embryonic and postnatal development of postmitotic cortical neurons in both rodents and non-human primates. In both species, we uncovered a striking developmental shift between two distinct strategies of epigenomic and transcriptional regulation active at early (perinatal) and late (juvenile/adult) stages of pan-neuronal development. Notably, early and late non-shared (cell-type-specific) programs did not show similar differences. These rules also apply across species, including non-human primates, suggesting that this temporal change in regulatory programs represents a broadly applied, core strategy for cortical neuron development.

Our finding of greater evolutionary conservation of shared, early-active regulatory elements is consistent with previous findings of higher conservation of regulatory elements at earlier stages in bulk forebrain tissue<sup>25</sup>. We now show that this differential conservation applies specifically to shared pan-neuronal regulatory programs but not to cell-type-specific programs active at the same ages and, furthermore, that this strategy is shared by multiple neuron types and conserved across evolution.

A previous study comparing multiple developing and adult human cell types indicated that, whereas programs specific to progenitor stages are often shared between multiple progenitor types, regulatory elements that become active in differentiated cell types are mostly cell type specific<sup>59</sup>. We show here that differentiated, post-mitotic neurons nonetheless use a high proportion of GREs with broad tissue and cell type distribution to regulate early pan-neuronal

programs. Taken together, these findings suggest that the regulatory programs necessary to produce a baseline cortical neuronal identity are under different developmental and evolutionary constraints from the programs required to confer the distinct, subtype-specific features of each neuronal class. Notably, these results could not have been predicted from analysis of other tissues. Indeed, the greater enhancer usage that we observed in the shared-late GRE clusters contrasts with observations in hematopoietic lineages, where the differentiation and maturation of hematopoietic stem cells into terminal cell types is associated with decreased enhancer usage<sup>32,34,60</sup>.

The data support a conceptual framework in which fundamental events of general, pan-cortical neuron development that occur during perinatal stages, such as establishment of neuronal identity and the acquisition of basic aspects of neuronal architecture, use molecular programs that are shared with other tissues and are, thus, mediated by more generic regulatory programs, requiring a more constrained degree of variation. It is tempting to speculate that this reflects the need for the nervous system to build its basic cell types in a reproducible and invariant manner. In contrast, as neurons transition to phases of neuronal, circuit and synaptic plasticity and function, they employ more specialized transcriptional and epigenetic programs that may allow for more flexibility. The greater variation in cell and circuit behavior at these later stages of cortical maturation may benefit from increased customization, reflected by more rapid species divergence in late developmental regulatory programs, similar to that found in the subtype-specific programs.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-022-01123-4>.

Received: 16 March 2021; Accepted: 16 June 2022;

Published online: 1 August 2022

## References

- Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
- Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
- Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
- Gray, L. T. et al. Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *eLife* **6**, e21883 (2017).
- Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
- Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
- Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
- Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
- Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
- Liu, H. et al. DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021).
- Zuccaro, E. et al. Human-specific enrichment of schizophrenia risk-genes in callosal neurons of the developing neocortex. Preprint at <https://www.biorxiv.org/content/10.1101/2021.09.10.459747v1.full> (2021).
- Molyneaux, B. J. et al. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85**, 275–288 (2015).
- Frazer, S. et al. Transcriptomic and anatomic parcellation of 5-HT<sub>3A</sub>R-expressing cortical interneuron subtypes revealed by single-cell RNA sequencing. *Nat. Commun.* **8**, 14219 (2017).
- Loo, L. et al. Single-cell transcriptomic analysis of mouse neocortical development. *Nat. Commun.* **10**, 134 (2019).

15. Noack, F. et al. Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. *Nat. Neurosci.* **25**, 154–167 (2022).
16. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
17. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
18. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
19. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
20. Nott, A. et al. Brain cell type-specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
21. Li, Y. E. et al. An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021).
22. Zhang, Z. et al. Epigenomic diversity of cortical projection neurons in the mouse brain. *Nature* **598**, 167–173 (2021).
23. Fang, R. et al. Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **12**, 1337 (2021).
24. Bhattacharjee, A. et al. Cell type-specific transcriptional programs in mouse prefrontal cortex during adolescence and addiction. *Nat. Commun.* **10**, 4169 (2019).
25. Nord, A. S., Pattabiraman, K., Visel, A. & Rubenstein, J. L. R. Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* **85**, 27–47 (2015).
26. Spiess, K. & Won, H. Regulatory landscape in brain development and disease. *Curr. Opin. Genet. Dev.* **65**, 53–60 (2020).
27. Ziffra, R. S. et al. Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).
28. Kozareva, V. et al. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. *Nature* **598**, 214–219 (2021).
29. Stroud, H. et al. An activity-mediated transition in transcription in early postnatal neurons. *Neuron* **107**, 874–890.e878 (2020).
30. Ardini-Poleske, M. E. et al. LungMAP: the molecular atlas of lung development program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313**, L733–L740 (2017).
31. Waardenberg, A. J., Ramalison, M., Bouveret, R. & Harvey, R. P. Genetic networks governing heart development. *Cold Spring Harb. Perspect. Med.* **4**, a013839 (2014).
32. Maqbool, M. A. et al. Alternative enhancer usage and targeted polycomb marking hallmark promoter choice during T cell differentiation. *Cell Rep.* **32**, 108048 (2020).
33. Wamstad, J. A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
34. Choukallah, M. A., Song, S., Rolink, A. G., Burger, L. & Matthias, P. Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. *Nat. Commun.* **6**, 8324 (2015).
35. Franco, S. J. et al. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science* **337**, 746–749 (2012).
36. Matho, K. S. et al. Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598**, 182–187 (2021).
37. Arlotta, P. et al. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron* **45**, 207–221 (2005).
38. Galazo, M. J., Emsley, J. G. & Macklis, J. D. Corticothalamic projection neuron development beyond subtype specification: *Fog2* and intersectional controls regulate intraclass neuronal diversity. *Neuron* **91**, 90–106 (2016).
39. McKinley, K. L., Castillo-Azofeifa, D. & Klein, O. D. Tools and concepts for interrogating and defining cellular identity. *Cell Stem Cell* **26**, 632–656 (2020).
40. Guibert, S. & Weber, M. Functions of DNA methylation and hydroxymethylation in mammalian development. *Curr. Top. Dev. Biol.* **104**, 47–83 (2013).
41. Thompson, C. L. et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* **83**, 309–323 (2014).
42. Di Bella, D. J. et al. Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).
43. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
44. FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
45. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
46. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
47. Ziller, M. J. et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).
48. Li, K. et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* **11**, 485 (2020).
49. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
50. Song, L. et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757–1767 (2011).
51. Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E. & Ohler, U. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* **22**, 1711–1722 (2012).
52. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
53. Hon, G. C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
54. Miller, C. T. et al. Marmosets: a neuroscientific model of human social behavior. *Neuron* **90**, 219–233 (2016).
55. Feng, G. et al. Opportunities and limitations of genetically modified nonhuman primate models for neuroscience research. *Proc. Natl Acad. Sci. USA* **117**, 24022–24031 (2020).
56. Sawiak, S. J. et al. Trajectories and milestones of cortical and subcortical development of the marmoset brain from infancy to adulthood. *Cereb. Cortex* **28**, 4440–4453 (2018).
57. Ehrlich, M. & Lacey, M. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. *Epigenomics* **5**, 553–568 (2013).
58. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
59. Stergachis, A. B. et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* **154**, 888–903 (2013).
60. Ye, M. et al. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc. Natl Acad. Sci. USA* **117**, 7905–7916 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

## Methods

**Ethics approvals.** All animal housing and procedures were conducted in accordance with the US National Institutes of Health *Guide for the Care and Use of Laboratory Animals*. All mouse experiments were approved by the Institutional Animal Care and Use Committee of Harvard University. All marmoset experiments were approved by the Institutional Animal Care and Use Committee of the Massachusetts Institute of Technology.

**Experimental design.** No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications<sup>61</sup>. For mice, animals were selected according to age, sex and genotype and were randomized within these criteria where possible. For marmosets, sample inclusion was constrained by tissue availability; all tissue was collected from healthy, un-manipulated individuals. Data collection and analysis were not performed blinded to the conditions of the experiments, as the identity of the samples was central to the analysis. The assumptions of the statistical tests used is detailed in the sections for each experimental method, where applicable. No animals or data points were excluded from the analysis.

**Mice.** Mice were group-housed in standardized, individually ventilated cages with a 12-hour light/dark cycle, food and water ad libitum, 30–70% humidity and a temperature of 22 °C ± 1 °C. Male and female mice were used for each experiment.

*Cux2*-lineage CPNs were labeled using a *Cux2*-CreERT2 knock-in line<sup>35,62</sup> (MMRRC 032779-MU). *Tle4*-lineage CTHPNs were labeled with a *Tle4*-2A-CreERT2 mouse line<sup>36</sup> (JAX 036298). The CRE-inducible tdTomato reporter line Ai14 (ref. <sup>63</sup>) (JAX 007914) was used to detect recombined cells. Animals were maintained on a mixed C57BL/6J background. *Cux2*-CreERT2;Ai14 or *Tle4*-2A-CreERT2;Ai14 double homozygous male mice were crossed with wild-type C57BL/6J females. Cre recombination was induced at E17.5, after the major wave of CPN neurogenesis (E15.5 to E17.5) was complete<sup>64</sup>. 4-Hydroxytamoxifen (4-OHT; Sigma-Aldrich) dissolved in corn oil was administered to pregnant mice at 1 mg of 4-OHT per 10 g of body weight.

For FACS isolation, CRE recombination was induced at E17.5, and the somatosensory cortex and a portion of the motor cortex from transgenic animals was dissected and dissociated at E18.5, P1, P3, P7, P21 and P48. We primarily aimed to analyze somatosensory cortex; however, our dissection strategy was informed by the limited recombination frequency typical of these experiments and the need to collect sufficient tissue to obtain enough labeled cells for bulk RNA-seq, ATAC-seq and WGBS. Tissue dissociation was performed as described<sup>37</sup>, and td-Tomato<sup>+</sup> cortical pyramidal neurons were isolated by FACS. In brief, cortex was enzymatically digested at 37 °C for 30 min with 10 U ml<sup>-1</sup> of papain (Worthington Biochemical, LS003126) in dissociation medium (20 mM glucose, 0.8 mM kynurenic acid (Sigma-Aldrich, K3375), 0.05 mM DL-2-amino-5-phosphopentanoic acid (APV; Sigma-Aldrich, A5282), 50 μl ml<sup>-1</sup> of penicillin-streptomycin solution (Gibco, 15140122), 0.09 M Na<sub>2</sub>SO<sub>4</sub>, 0.03 M K<sub>2</sub>SO<sub>4</sub> and 0.014 M MgCl<sub>2</sub>) supplemented with 0.016 μg ml<sup>-1</sup> of L-cysteine HCl (Sigma-Aldrich, C7477). For P21 and P48 animals, papain concentration was increased to 20 U ml<sup>-1</sup>. Papain digestion was stopped with room temperature dissociation medium supplemented with 10 mg ml<sup>-1</sup> each of ovomucoid protease inhibitor and BSA (Worthington Biochemical, LK003182), and tissue was mechanically dissociated by gentle trituration in ice-cold Opti-MEM (Gibco, 31985070) supplemented with 20 mM glucose, 0.4 mM kynurenic acid and 0.025 mM APV. We performed RNA-seq and ATAC-seq at E18.5, P1, P3, P7, P21 and P48 and WGBS at P1, P21 and P48. Each library represents a pool of tissue from ten animals (five male and five female) from two litters. Two biological replicates were performed for each age and neuron type for each assay.

**Marmoset.** Marmoset tissue was obtained from the laboratory of Guoping Feng at the Massachusetts Institute of Technology. For tissue collection, adult marmosets were deeply sedated with ketamine (20–40 mg kg<sup>-1</sup>, intramuscular) and/or alfaxalone (5–10 mg kg<sup>-1</sup>, intramuscular), followed by intravenous injection of sodium pentobarbital (10–30 mg kg<sup>-1</sup>). Because venous access was not possible in neonates, infant marmosets were sedated with intraperitoneal injection of sodium pentobarbital (10–30 mg kg<sup>-1</sup>). When pedal withdrawal reflex was eliminated and/or respiratory rate was diminished, animals were transcardially perfused with ice-cold PBS or sucrose-HEPES buffer. Whole brains were rapidly extracted into fresh buffer on ice. A series of 2-mm coronal blocking cuts were rapidly made using a custom-designed marmoset brain matrix. Slabs were transferred to a dish with ice-cold buffer, and regions of interest were dissected using a marmoset atlas as reference. Samples of somatosensory cortex were flash-frozen in RNAlater (Invitrogen) or immediately processed for cell dissociation.

**Evaluation of interneuron representation in FACS-purified *Cux2*-CreERT2 cells.** It was reported that the *Cux2*-CreERT2 line used here also labels a subset of cortical interneurons<sup>35,65</sup>. To evaluate the effects of this on our analysis, we performed single-cell sequencing of a total of 14,792 cells for the *Cux2* CPN and *Tle4* CTHPN populations at three ages, labeled using the same induction strategy used for our primary analysis (Supplementary Fig. 1), using the same approaches as described below for library preparation and analysis. Of these, 239 cells were

positive for expression of the interneuron marker Gad1 (1.61%); 339 cells were positive for the interneuron marker Gad2 (2.29%); and 118 cells were positive for both (0.798%), for a total of <5% cells positive for either marker. We conclude that the effect of interneuron contamination on our analysis was minimal.

**RNA-seq of genetically identified projection neuron populations.** Pools of 5,000–10,000 cortical pyramidal neurons were sorted directly into TRIzol-LN buffer (Invitrogen), and RNA was extracted according to the manufacturer's protocol. Ten nanograms of total RNA was used for library preparation using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech) and the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's protocols. All libraries were sequenced according to the manufacturer's protocols on the HiSeq 2500 system (Illumina), using 125-bp paired-end reads at a depth of >20 million reads per library.

**ATAC-seq of genetically identified projection neuron populations.** Pools of 3,000–5,000 cortical pyramidal neurons were sorted into Opti-MEM media (Life Technologies). Nuclei were extracted and libraries prepared following a previously published protocol<sup>66</sup>. All libraries were sequenced according to the manufacturer's protocols on the HiSeq 2500 system, using 50-bp single-end reads at a depth of ~50 million reads per library.

**WGBS of genetically identified projection neuron populations.** Pools of 5,000–10,000 cortical pyramidal neurons were sorted into PBS. The EZ DNA Methylation-Direct Kit (Zymo Research) was used to perform bisulfite conversion, and libraries were prepared with the EpiGnome Methyl-Seq Kit (Illumina). All libraries were sequenced according to the manufacturer's protocols on the HiSeq 2500 system, using 125-bp paired-end reads at a depth of >200 million reads per library.

**Comparison to human BrainSpan data.** We first identified human orthologs of the genes in the mouse shared-early and shared-late gene clusters using the Ensembl database. We then examined expression of these orthologs in human developing brain transcriptomic data from BrainSpan<sup>43</sup>; 1,129 out of 1,334 genes in the shared-early gene cluster and 1,234 out of 1,446 genes in the shared-late gene cluster had orthologs that were expressed in the human data. We then examined expression of these genes within four brain regions (primary motor cortex, primary somatosensory cortex, amygdaloid complex and striatum) over a time course spanning eight (amygdaloid complex) or 12 (all others) post-conceptual weeks to 40 years. Z-scores of the bulk RPKM values were plotted as heat maps using the 'pheatmap' package<sup>67</sup> version 1.0.12 in R version 3.6.0 (Supplementary Fig. 4).

**Comparison to the VISTA enhancer database.** We downloaded all enhancers in the VISTA enhancer dataset<sup>52</sup> from their website (<https://enhancer.lbl.gov/>) and compared their genomic coordinates to our ATAC data to identify tested regions that spanned ATAC peaks in our shared-early and shared-late clusters. We selected eight representative regions for each cluster and show the ATAC peak signal over each region in our data, as well as the expression pattern it drives in E11.5 mouse embryos in the VISTA dataset (Extended Data Fig. 7).

**scRNA-seq. Mouse.** Somatosensory and motor cortex from wild-type animals was dissected and dissociated at P1, P7 and P21. For consistency, we collected the same cortical regions as described above for the bulk profiling of transgenically labeled neurons. Each library was made from tissue pooled from at least eight animals, and a balanced sex ratio was used. Tissue dissociation was performed as described above, and live cells were isolated by FACS sorting as DAPI-negative, Vybrant DyeCycle Ruby (Thermo Fisher Scientific)-positive events. Libraries were prepared using the 10x Genomics Chromium Single Cell 3' kit v2 according to the manufacturer's protocol.

**Marmoset.** Somatosensory cortex from wild-type animals was dissected and flash-frozen in RNAlater (Invitrogen). Each library was made from tissue from an independent individual. Nuclei were extracted by a previously published protocol<sup>68</sup>, and debris was removed by FACS isolation of DAPI-positive nuclei. In brief, cortical tissue was placed in 1 ml of cold nuclei extraction buffer (0.32 M sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(Ac), 0.1 mM EDTA, 10 mM Tris-HCl and 0.1% Triton X-100) with 10 μl of protease inhibitor cocktail (Sigma-Aldrich, P8340), 1 μl of 100 mM phenylmethylsulfonyl fluoride (PMSF; Sigma-Aldrich, 78830), 1 μl of 1 M 1,4-dithiothreitol (DTT; Sigma-Aldrich, D9779) and 3 μl of 40 U μl<sup>-1</sup> mRNase inhibitor (Promega, N2611). Nuclei were liberated by dounce homogenization (Sigma-Aldrich, D9063), using 15 strokes with the loose pestle and 25 strokes with the tight pestle. The crude suspension was filtered through a 40-μm nylon mesh cell strainer (Thermo Fisher Scientific, 22363547), transferred to a 15-ml conical tube and centrifuged at 1,000g for 10 minutes. The supernatant was removed, and the cell pellet was gently resuspended in 1 ml of cold nuclei extraction buffer with 10 μl of protease inhibitor cocktail, 1 μl of 100 mM PMSF, 1 μl of 1 M DTT and 1.5 μl of 40 U μl<sup>-1</sup> RNase inhibitor. The suspension was divided between two 1.5-ml microcentrifuge tubes, and each 500-μl sample was gently mixed with 0.75 ml of 50% iodixanol, for a final concentration of 30% iodixanol. The 50% iodixanol

solution was prepared by adding 0.4 ml of diluent (150 mM KCl, 30 mM MgCl<sub>2</sub> and 120 mM Tris-HCl, pH 7.8) to 2 ml of 60% iodixanol (Sigma-Aldrich, D1556). The samples were then centrifuged at 10,000g for 20 minutes, and the pellet was resuspended in PBS with 60 U ml<sup>-1</sup> of RNase inhibitor for FACS isolation. Nuclei were sequenced using the 10x Genomics Chromium Single Cell 3' kit v2 according to the manufacturer's suggested protocol for nuclei.

**scATAC-seq. Mouse.** Somatosensory and motor cortex from wild-type animals was dissected at P1, P7 and P21. For consistency, we collected the same cortical regions as described above for the bulk profiling of transgenically labeled neurons. Libraries were prepared as described in LaFave et al.<sup>69</sup>, a modification of Cusanovich et al.<sup>70</sup>. In brief, cells were fixed with 0.1% formaldehyde and incubated at room temperature for 5 minutes. The fixation was stopped by adding glycine to the final concentration of 125 mM. The sample was incubated at room temperature for 5 minutes and washed in PBS. The cell concentration was counted, and approximately 1,600–2,000 cells per well were distributed into each well of a 96-well plate. Cells were transposed with 96 uniquely barcoded Tn5 at 37°C for 30 minutes with shaking at 300 r.p.m. The reaction was stopped by adding 0.5 M EDTA and incubated at 37°C for 15 minutes. All the cells were then pooled, and MgCl<sub>2</sub> was added to the pooled sample to quench EDTA. The sample was re-distributed onto another 96-well plate with 20 cells in each well by FACS sorting. Reverse crosslinking buffer and barcode PCR primers were added to each sample. The plate was incubated at 55°C for 16 hours for reverse crosslinking. Tween 20 was then added to quench SDS before PCR amplification.

The PCR reaction was carried out at the following conditions: 72°C for 5 minutes (extension), 98°C for 5 minutes and then thermocycling at 98°C for 10 seconds, 70°C for 30 seconds and 72°C for 1 minute for 12–15 cycles. Libraries were pooled and purified using Qiagen MinElute PCR purification column. The libraries were quantified using KAPA library quantification kit. Libraries were sequenced on the Next-seq platform (Illumina) using a 150-cycle kit (Read 1: 47 cycles, Index 1: 36 cycles, Index 2: 36 cycles, Read 2: 47 cycles).

**Marmoset.** Somatosensory and motor cortex from wild-type animals was dissected and dissociated with the Worthington Papain Dissociation System (Worthington Biochemical), and live cells were isolated by FACS sorting as DAPI-negative, Vybrant DyeCycle Ruby (Thermo Fisher Scientific)-positive events. Libraries were prepared using the 10x Genomics Chromium Single Cell ATAC kit.

**Immunohistochemistry.** To confirm class specificity of labeling, CRE recombination was induced at E17.5 by tamoxifen administration, and mice were sacrificed at E18.5, P1, P3, P7, P21 and P48 for co-immunolabelling with antibodies against the canonical layer markers CUX1, CTIP2 and SATB2.

Mice were deeply anesthetized with tribromoethanol and perfused transcardially with PBS, followed by 4% paraformaldehyde (PFA) in PBS. Brains were post-fixed in 4% PFA overnight, washed in PBS, embedded in low-melting-point agar and sectioned at 20 μm using a Leica VT1000 S vibrating microtome. Sections were transferred to six-well plates with Netwell Inserts (Corning, 3479), washed twice with PBST (1× PBS with 0.2% Triton X-100) and then incubated in blocking buffer consisting of PBST with 8% (v/v) normal goat serum (Invitrogen, 16210-072) or normal donkey serum (Sigma-Aldrich, D9663). Sections were incubated overnight at 4°C with primary antibodies diluted in blocking buffer, washed in PBST and then incubated with Alexa Fluor-conjugated secondary antibodies diluted in blocking buffer for 2 hours at room temperature. Finally, sections were washed in PBST and mounted with DAPI Fluoromount-G (Southern Biotech, 0100-20). Primary antibodies and dilutions were as follows: mouse anti-Satb2, 1:50 (Abcam, ab51502); rat anti-Ctip2, 1:100 (Abcam, ab18465); and rabbit anti-Cux1 (CDP M-222), 1:300 (Santa Cruz Biotechnology, sc-13024). Secondary antibodies were: goat anti-rat Alexa Fluor 488 (Thermo Fisher Scientific, A48262), donkey anti-mouse Alexa Fluor 647 (Thermo Fisher Scientific, A-31571) and donkey anti-rabbit Alexa Fluor 647 (Thermo Fisher Scientific, A-31573). All secondary antibodies were used at 1:1,000 dilution. Imaging was performed using a Nikon 90i fluorescence microscope equipped with a Retiga EXi camera (Qimaging). Analysis was done with Volocity image analysis software version 4.0.1 (Improvision).

**Enhancer silencing by enCRISPRi.** We performed the CRISPRi experiment in the mouse neuroepithelial cell line NE4C (American Type Culture Collection (ATCC)). We selected three regions predicted to be open around each of Sox4 and Sox11 in our mouse bulk ATAC dataset. For each of these predicted enhancer regions, we designed three single guide RNAs (sgRNAs) using the Benchling sgRNA design tool (<https://www.benchling.com/crispr/>).

NE4C cells were cultured following the supplier's protocol (ATCC). After transfection with the CRISPRi constructs<sup>48</sup>, neuronal differentiation was induced with retinoic acid. Starting 5–7 days after induction of differentiation, doxycycline was added to the plates to induce the dCas9 expression and silence the candidate regions for 5 days. Cells were then collected, and Sox4 and Sox11 expression was quantified by qPCR. Expression was normalized against cells transfected with an irrelevant control sgRNA. Each experiment included three biological replicates, each with two technical replicates.

sgRNA sequences used:

Sox4:  
sgRNA1 AGTTAACTGTTGAGAAAAGATG  
sgRNA2 CTAAGGTCTTGAGATAAACACG  
sgRNA3 TTAATATAACATGACAGCCACG  
Sox11:  
sgRNA1 GTCCAACAGCCAGATCTTATAG  
sgRNA2: AGTCCTTGCCCATAGCTCCAG  
sgRNA3: GATTGCCTTGATTCCTAAAACG

**Bioinformatics analysis. Data processing.** Bulk RNA-seq. Raw reads were trimmed using Trimmomatic<sup>71</sup> version 0.33, removing 8 bp from the 5' end and 25 bp from the 3' end. Subsequently, reads were aligned to the Ensembl NCBI37 (mm9) genome build (downloaded from the Illumina iGenomes file collection), using TopHat2 (ref.<sup>72</sup>) version 2.0.13 with default parameters. Subsequently, differential gene expression analysis and FPKM quantification was performed using Cuffdiff<sup>3</sup> version 2.2.1 for all pairwise comparisons. Differentially expressed genes were defined as all genes that showed a significant (false discovery rate (FDR) ≤ 0.05) change in gene expression with ≥ 1.5 log<sub>2</sub> fold change in at least one comparison and were expressed at levels greater than 10 FPKM in at least one condition within that comparison, resulting in *n* = 4,419 differentially expressed genes across the entire dataset.

Next, genes were clustered using k-means clustering on the log<sub>2</sub>-transformed and Z-scored FPKM values of all differentially expressed genes using 100 random starts. To determine the number of clusters, we used the gap statistics as implemented in the R package cluster<sup>74</sup> version 2.0.7 in combination with the Tibshirani 2001 method<sup>75</sup> based on the standard deviation evaluating *k* = 2–20, identifying 12 clusters in total. Finally, we classified each cluster manually based on its expression dynamic into one of five categories: shared developmental clusters (1–5), Cux2 CPN-specific (6–8), Tle4 CThPN-specific (9–11) and other (12), as shown in Fig. 1c.

**Bulk ATAC-seq.** ATAC-seq raw reads were aligned to the genome build NCBI37 downloaded from the Illumina iGenomes collection using Bowtie2 (ref.<sup>76</sup>) with default parameters. Subsequently, aligned reads were filtered for duplicates using MarkDuplicates from the Picard software toolbox version 2.7.1 (<http://broadinstitute.github.io/picard/>)<sup>77</sup>. Next, we performed peak calling for each sample group using the irreproducible discovery rate (IDR) framework<sup>78</sup> in combination with the macs2 peak caller (MACS2 version 2.1.1), with two independent biological replicates in each group. All peaks detected at IDR ≤ 0.1 in each group were retained for further analysis. We then performed differential peak enrichment analysis across all pairwise group comparisons using the diffBind package<sup>79,80</sup> in combination with DESeq2 (ref.<sup>81</sup>). To that end, we employed the DBA\_SCORE\_TMM\_READS\_EFFECTIVE score for normalization and subsequent differential enrichment analysis. We defined all peaks exhibiting significant (FDR ≤ 0.01) differential enrichment above a log<sub>2</sub> fold change of 1.5 and a minimum enrichment ≥ 1 trimmed mean of M values (TMM) normalized reads in at least one condition as differential, resulting in *n* = 66,784 differentially enriched ATAC-seq peaks across the entire dataset. Subsequently, we averaged over all replicates for each group and transformed the resulting TMM value to log<sub>2</sub> space. Next, we conducted k-means clustering on Z-scored and log<sub>2</sub>-transformed TMM values on all differentially enriched regions using 100 random starts. We again used the gap statistics in combination with the Tibshirani SE method, which identified 13 clusters. After inspection of the cluster dynamics, we annotated each cluster as shared developmental (split into shared-early and shared-late), CPN-specific, CThPN-specific or other.

Finally, we associated each differentially active ATAC-seq peak with the closest Ensembl gene TSS using the ChIPpeakAnno package<sup>82</sup>. Selected gene names based on this association are shown in Fig. 1d. We then classified each cluster as developmental, neuron class-specific or other according to its dynamic enrichment patterns (Fig. 1d).

**Bulk WGBS.** Raw sequencing reads were trimmed 8 bp from the 5' end and 40–60 bp from the 3' end, depending on library quality. Next, reads were aligned to the genome build NCBI37 using bsmapp<sup>83</sup> version 2.9 with parameter settings -v 0.1 -s 16 -q 20 -w 100 -S 1 -u -R. Aligned data were then filtered for PCR duplicates using the MarkDuplicates function implemented in the Picard toolbox. Next, CpG methylation calling was performed on the duplicate filtered data using the mcall function implemented in the MOABS suite<sup>84</sup> version 1.3.2 with default parameters.

**scATAC.** Base calls were converted to FASTQ format using bcl2fastq (Illumina). Raw sequencing reads were trimmed using custom Python scripts to remove adapter sequences. The data were demultiplexed tolerating one mismatched base within barcodes. Mitochondrial, unpaired and low-quality reads were removed using SAMtools<sup>85</sup> version 1.5 (samtools view -b -q 30 -f 0x2). Duplicate sequences were removed using the Picard toolkit<sup>77</sup>. The reads were aligned to the mm10 or CalJac3 genomes using Bowtie2 (ref.<sup>76</sup>) version 2.3.2 with maximum fragment length set to 2 kb and all other default settings (bowtie2 -X2000 -rg-id).

**Data analysis. Identification of DMRs.** DMRs were identified using the R package DSS<sup>86</sup>. To that end, we performed all pairwise comparisons across sample groups. For each of these pairwise comparisons, we applied the following three functions from the DSS package to the appropriate biological replicates. First, we used the `dm1Test` 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. We set the parameters `delta=0.3`, `p.threshold=0.01`, `minCG=3` and `dis.merge=400`; 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 employed the `methylKit` package<sup>87</sup> version 0.9.5. In particular, we computed the methylation level and coverage of each DMR in each sample, defined as the weighted average of CpG methylation levels weighted by coverage. We then retained only those DMRs that were covered by more than five reads in at least three samples. Next, we averaged the methylation levels of each DMR across replicates and assigned the DMRs to different groups based on their methylation differences between the replicate-averaged, DMR-level methylation values. DMRs that exhibited an absolute methylation difference  $\geq 0.3$  between any pair of samples and exceeded a size of 100 bp were defined as differentiation DMRs.

We then performed clustering using *k*-means, initially identifying 12 clusters that we collapsed upon further inspection into ten distinct clusters. We again annotated each cluster according to its dynamic enrichment patterns (Fig. 1e).

We determined global CpG and non-CpG methylation levels as the fraction of methylated CpGs over the total number of detected CpGs (and correspondingly for non-CpGs) for all ATAC-seq regions overlapping with DMRs and covered by at least ten reads in 80% of the samples using the function in the `regionCounts` function in the `methylKit` package. We report the corresponding feature methylation values in Extended Data Fig. 2h.

**MDS analysis.** We performed MDS using the `cmdscale` R function with 1 minus the absolute Person correlation coefficient as metric, reducing the dimensionality to two dimensions. We then computed the distances of individual samples shown in Fig. 1g and corresponding text as the two-dimensional (2D) Euclidean distance. As input, we used the  $\log_2 + 1$  transformed FPKM values of all expressed genes ( $\geq 10$  FPKM in at least one condition) (RNA-seq), the  $\log_2 + 1$  and quantile-normalized ATAC-seq TMM values (ATAC-seq) and the methylation level of all 1-kb tiles of the mouse genome covered by at least five reads in more than two samples.

**Specificity analysis for bulk RNA-seq clusters.** We report specificity analysis for differentially expressed TFs as well as gene expression clusters. These specificity analyses were conducted using three distinct datasets that were processed in the following manner:

**FANTOM5:** Similarly, we computed the expression specificity of TFs in each expression cluster, all TFs and all genes across 294 mouse cell types and tissues based on CAGE data from the FANTOM5 consortium<sup>44</sup>. To that end, we downloaded the CAGE-tag data for promoter regions from the FANTOM5 cell and tissue collection from <http://fantom.gsc.riken.jp/5/>. We then collapsed all CAGE-tag peaks for each gene by summing up the tag counts, including only primary cell types. Subsequently, we collapsed replicates for each cell type or tissue by averaging.

**Allen Brain in situ<sup>41,88</sup>:** The Allen brain data were downloaded from the Allen Brain Atlas Developing Mouse Brain website (<https://developingmouse.brain-map.org/>). We obtained ISH counts for the developing mouse brain at seven distinct fetal timepoints and 11 different brain substructures. We then intersected the resulting list of genes with the list of TFs/genes in each expression cluster and determined the expression specificity of each TF/gene across the 77 conditions (see below for further details) from this atlas. We then plotted the distribution of specificities for each cluster in Fig. 2c for control purposes.

**Mouse single-cell RNA-seq data:** Here, we used the  $\log_2$ -normalized expression values averaged over all the cells in a particular cell cluster using the `AverageExpression()` function in Seurat. We then used these pseudo-bulk expression values for each gene across all identified cell types as input for the specificity analysis.

Similarly, we computed the expression specificity of TFs in each expression cluster, all TFs and all genes across 294 mouse cell types and tissues based on CAGE data from the FANTOM5 consortium<sup>44</sup>. To that end, we downloaded the CAGE-tag data for promoter regions from the FANTOM5 cell and tissue collection from <http://fantom.gsc.riken.jp/5/>. We then collapsed all CAGE-tag peaks for each gene by summing up the tag counts.

We then computed expression specificity for each TF/gene following previous approaches using the tau specificity measure<sup>89</sup> according to<sup>90</sup>:

$$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1}; \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} x_i}$$

with  $n$  being the number of samples/tissues and  $x_i$  being the expression of the gene in tissue  $i$ . We report the result in Fig. 2c.

**Specificity analysis for scATAC clusters.** To assess the specificity of each ATAC peak in the bulk or scATAC dataset, we downloaded pre-computed ATAC peak specificity scores computed over more than 80 distinct cell types of an entire mouse using scATAC-seq data from Cusanovich et al.<sup>17</sup>.

We then intersected our peak library with the Cusanovich et al. dataset and report the Cusanovich et al. specificity values for all peaks that overlap with at least one peak in the Cusanovich et al. dataset.

**Bulk ATAC peak overlap with DNase data.** To create a catalog of gene regulatory elements in the mouse genome, we downloaded a set of DNase HS I peak tracks from the mouse ENCODE consortium<sup>46</sup> (Supplementary Table 5). Subsequently, we collapsed replicates for each condition and required that each peak was present in at least two replicates. This step resulted in DNase I tracks for 35 distinct primary mouse cell types and tissues. Next, we merged all DNase I tracks into a union peak set using the `reduce()` function in the `IRanges` R package<sup>91</sup>. Subsequently, we size-standardized the resulting union peak set to 350 bp by extending 175 bp from the center of each peak. Next, we assigned a binary value for each peak in each of the 35 cell types, depending on whether or not the peak was present in the individual cell-type-level peak set. We then used this union peak set and overlapped all ATAC peaks from the mouse bulk ATAC-seq dataset with this library. We report the percent of peaks in each cluster that overlap DNase I HS sites in each cell type.

**Definition of genomic features.** CGIs were defined as previously described<sup>92</sup> and are listed in Supplementary Table 6. Promoters were defined as all NCBI m37 Ensembl version 67 TSSs, extended by 1 kb upstream and downstream.

**TF binding site density analysis.** For each set of regions of interest (DMRs and ATAC peaks), we performed motif detection analysis using FIMO<sup>93</sup> with a  $P$  value filter of less than  $10^{-4}$  and a joint motif database comprising the TRANSFAC Professional library (version 2011)<sup>94</sup> and a set of previously published motifs by Jolma et al.<sup>95</sup>. All genomic regions were size-standardized (if not already) before motif analysis.

**Entropy analysis.** To compute the Shannon entropy of the size-standardized ATAC-seq peak sequence, we used the `entropy()` function in the `sequtils`<sup>96</sup> Python package.

**Integrative analysis.** To evaluate the concordance of changes in the transcriptome, open chromatin and DNA methylation landscape, we associated each ATAC-peak or DMR with its nearest gene within 100 kb upstream or downstream. Peaks/DMRs without any assigned gene were not considered. Subsequently, we performed a hypergeometric test between all pairs of ATAC-peak/DMR clusters and bulk expression clusters in gene space to assess the significance of overlap. After multiple-testing correction using the Benjamini–Hochberg method<sup>97</sup>, we report the odds ratio of associations significant below a  $q$  value of 0.001 in Extended Data Fig. 5c,d, capping the odds ratio at 10.

**Phylogenetic conservation analysis.** We performed phylogenetic conservation analysis for ATAC-seq cluster groups by computing the average placental mammal phyloP scores<sup>98</sup> for each region. We then plotted the distribution of these mean scores.

**Signature gene set analysis.** First, we associated each of our consensus DMRs with the nearest mouse Ensembl TSS within 100 kb using the `ChIPPeakAnno` R package<sup>82,99</sup>. Subsequently, we performed gene-set-level analysis and determined the mean methylation level of all DMRs associated with a member gene of each signature gene set using the aforementioned DMR–gene associations. We used signature gene sets for CThPNs, CPNs and SCPNs obtained from the DeCoN database<sup>12</sup> as well as manually curated gene sets for glial and interneuron cell types from published transcriptomic data<sup>100,101</sup>. We then report the mean methylation level of all DMRs associated with each gene set in Extended Data Fig. 2c for each timepoint and cell type.

**Global methylation level analysis.** Global CpG methylation level was defined as the total number of detected methylated Cs in CpG context over the total number of CpGs sequenced. Similarly, we defined the global non-CpG methylation level comprising all other dinucleotide contexts.

**scRNA-seq analysis.** Mouse scRNA-seq data were processed using Cell Ranger version 3.0.1 (10x Genomics) using standard parameters and genome assembly GRChm38 downloaded from the 10x Genomics website. After initial alignment and processing by `cellranger count`, all replicates across all timepoints were aggregated using the `cellranger aggr` function, downsampling the individual libraries to a similar overall coverage by cell.

All following analyses were conducted in R using the Seurat package (version 2.3.4). We first filtered the dataset to retain only cells with at least 1,000 genes, a mitochondrial read fraction below 10% and not more than 6,000 genes or 15,000 unique molecular identifiers (UMIs).

Subsequently, we initially performed cell clustering and subtype identification separately for each timepoint P1, P7 and P21. For that purpose, we subsetted the data for each timepoint and then applied the following workflow. Expression data were normalized using the LogNormalize method, and variable genes were identified based on the mean/dispersion relation (initially using the following parameters:  $x_{\text{low.cutoff}} = 0.05$ ,  $x_{\text{high.cutoff}} = 3$ ,  $y_{\text{cutoff}} = 0.05$ ).

Next, the data were scaled using the ScaleData function, regressing out percentage of mitochondrial reads and UMI count per cell. We then computed the first 100 principal components (PCs) and examined the resulting elbow plot for variance explained by each PC. Based on that, we selected the first 35–40 PCs for subsequent dimensionality reduction by uniform manifold approximation and projection (UMAP) ( $\text{min\_dist} = 0.2$  to  $0.5$ ) and cluster identification using the Seurat FindClusters (resolution =  $0.4$  to  $0.6$ ,  $\text{nn.eps} = 0.5$ ) function.

We conducted several rounds of principal component analysis (PCA), clustering and UMAP embedding, successively removing clear outlier cell clusters based on the UMAP (four for P1, two for P7 and two for P21).

After cleanup and cluster identification by timepoint, we analyzed all timepoints jointly, assigning the cell original cluster definitions based on the individual timepoint analysis. For that purpose, we used as cutoffs in the variable gene feature analysis:  $x_{\text{low.cutoff}} = 0.025$ ,  $x_{\text{high.cutoff}} = 3$ ,  $y_{\text{cutoff}} = 0.05$ , 50 PCs, a resolution of  $0.9$  in the clustering analysis and a  $\text{min\_dist} = 0.4$  in the UMAP.

Next, we performed cell type identification for each cluster using a set of manually curated marker genes for each cell type, identified from previous literature, the DeCoN database<sup>12</sup> (for projection neuron subtypes) and published transcriptomic data for glial and interneuron cell types<sup>100,101</sup>. To assign identities, we examined the expression of all marker genes individually using the UMAP. In addition, we computed the AverageExpression for each cluster and examined the pseudo-bulk profiles. In particular, we computed joint cell type scores for each cluster and potential identity by normalizing the average cell type scores for each cluster to the maximum observed score for each cell type separately. Moreover, we performed hierarchical clustering, correlation analysis, PCA and MDS evaluation of the pseudo-bulk profiles to identify outlier clusters and investigate the relationship between cell clusters in more detail. Based on these analyses, we were able to assign identities to most cell clusters. For a few clusters that appeared to have mixed identities (such as a cluster containing multiple subtypes of interneurons), we performed subclustering to refine the identities of these cells. We then collapsed subclusters with the same subtype identity into a single subtype cluster to reduce the complexity for subsequent analysis (final cluster  $n = 32$ ).

**Differential expression analysis.** To identify differentially expressed genes between timepoints within distinct cell type classes, we assigned each cell cluster to seven sets of general cell classes that we analyzed separately: excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, neurons, glia and all cell types combined.

We then performed differential expression analysis in pairwise fashion between P1 and P7, P1 and P21 and P7 and P21 for each of the aforementioned cell type classes using the FindMarker function in Seurat and the MAST method for differential expression testing.

In this manner, we obtained seven distinct lists of genes differentially expressed across development. Next, we averaged the expression of all cells within one cluster using the AverageExpression function in Seurat. To avoid biases driven by cluster complexity and cell number, we downsampled each cluster to a maximum of 500 cells per cluster before averaging. Next, we performed clustering on  $Z$ -score-transformed expression values of each of the seven differentially expressed gene sets separately. For that purpose, we used the clusGap function in the R package cluster<sup>74</sup> using  $k$ means with a maximum of 20 clusters and the B parameter set to 60. We then select the final number of clusters based on the Tibshirani 2001 criterion<sup>75</sup> for the standard deviation as implemented in the cluster package. Finally, we performed  $k$ -means clustering with 100 random starts using the identified number of clusters and average over the  $Z$ -scored genes in each cluster (Fig. 3 and Extended Data Fig. 8).

Lastly, we performed expression specificity analysis for all gene clusters in a similar manner as for the bulk data, using the Allen Brain in situ dataset, the FANTOM5 and the mouse scRNA-seq low-resolution dataset for comparison for all differentially expressed genes as well as for TFs only.

**Marmoset snRNA-seq analysis.** Marmoset snRNA-seq data were processed using Cell Ranger version 2.1.0 and aligned to a pre-mRNA custom build transcriptome of assembly version ASM275486v1.93. All replicates were first processed independently using cellranger count and then aggregated, downsampling all libraries to the same complexity per cell. Subsequently, the data were processed using Seurat following a similar workflow as for the mouse scRNA-seq, retaining only cells with more than 1,000 genes and genes detected in more than ten cells. Initially, variable genes were again identified using the mean/dispersion relation with the following parameters:  $x_{\text{low.cutoff}} = 0.05$ ,  $x_{\text{high.cutoff}} = 3$ ,  $y_{\text{cutoff}} = 0.05$ . During initial quality control of the five individual marmoset libraries, we observed a separation by experimental batch, where all three libraries (D0 and Y2) from experimental batch 1 and the two libraries from batch 2 (D0 and Y2) grouped together. Given that developmental timepoint and experimental batch are

not confounded in our experimental design, we performed batch correction using CCA as implemented in the Seurat package. Here, we discarded all cells where the variance explained by CCA is  $< 2$ -fold. After correction, a clear separation by timepoint and cell type became apparent, as originally observed when processing each experimental batch separately. We thus proceeded with the CCA-corrected data following the same strategy as for the mouse scRNA-seq data, performing PCA and using 20 dimensions in subsequent analyses, identifying clusters with a resolution of  $1.8$  and creating a UMAP embedding with a  $\text{min\_dist}$  of  $0.3$ .

Subsequently, data analysis was conducted similarly to the mouse scRNA-seq data, assigning cell types and cell classes, performing differential expression analysis between timepoints within each of seven cell classes, collapsing clusters with similar subtype identity and performing gene clustering within each cell group to identify gene clusters.

**scATAC analysis. Read alignment and pre-processing.** Base calls were converted to FASTQ format using bcl2fastq (Illumina). Raw sequencing reads were trimmed using custom Python scripts to remove adapter sequences. The reads were aligned to the mm10 or CalJac3 genome using Bowtie2 (ref.<sup>76</sup>) with maximum fragment length set to 2 kb and all other default settings (bowtie2 -X2000 -rg-id). The data were demultiplexed tolerating one mismatched base within barcodes. Mitochondrial, unpaired and low-quality reads were removed using SAMtools<sup>85</sup> (samtools view -b -q 30 -f 0x2). Duplicate sequences were removed using the Picard toolkit<sup>77</sup>. To counteract differential complexity across the libraries, each library was sampled to a similar fragment depth per cell.

Mouse and marmoset libraries were first downsampled to similar overall complexity in terms of reads per cell across all conditions. Cleaned data were processed with the Scasat<sup>102</sup> pipeline using macs2 as peak caller with parameters set to -q 0.2 -nomodel -nolambda, giving rise to a peak  $\times$  cell matrix count matrix. For genome size, we used -g mm for mouse and -g 2.1e+9 for marmoset. We then size-standardized all peaks to 500 bp and recomputed the peak  $\times$  cell count matrix, considering only reads overlapping the size-reduced peaks.

Only cells with at least 1,000 reads and 300 peaks from the master list overlapping with at least one read, but not more than 15,000 reads and 8,000 peaks, were retained for analysis. In addition, only peaks present in at least 20 cells were retained.

Subsequently, this matrix was processed using R implementing a custom processing pipeline based on the strategy outlined in Cusanovich et al.<sup>17</sup> and refined by Hill<sup>103</sup>, following the log-latent semantic indexing (log-LSI) workflow. In brief, the count matrix was first binarized and then transformed using the TF-IDF method<sup>70</sup>, log-scaling the results. Next, PCA was performed on the transformed matrix using 50 dimensions, after cell cluster identification based on Seurat's FindCluster() function with resolution set to  $0.5$  and UMAP embedding with a  $\text{min\_dist}$  of  $0.3$ . Notably, we split each cluster containing cells from different timepoints into separate clusters for each timepoint and then filtered out all clusters with fewer than 100 cells.

**Cell type identification in scATAC data.** To reliably identify distinct cell types, we next computed gene activity scores for all genes based on the presence of ATAC peaks. To that end, we used the R package Cicero<sup>104</sup>, because it implements an approach considering not only peaks located within the promoter/gene body of each gene but also weights the contribution of each peak to the overall gene score based on the correlation of the gene peaks with each other. To that end, we imported the pre-processed data into a cicero atac\_cds object providing the cluster ID and UMAP coordinates defined based on the aforementioned LSI analysis. Next, we estimated library size factors and reliable peaks using default parameters, followed by running Cicero's main function with default parameters to compute the connectivity graph of all peaks. Next, we annotated all peaks with the annotate\_cds\_by\_site function using the transcription start coordinates for all Ensembl genes retrieved from biomaRt, defining the region  $\pm 5$  kb of each TSS as the promoter region. With this annotation in place, we constructed the raw gene activity matrix using the build\_gene\_activity\_matrix function, followed by normalizing the resulting scores using the function normalize\_gene\_activities.

Next, we averaged the computed gene scores for each cell cluster and normalized the aggregated score for each gene to the maximum across all pseudo-bulk cell clusters. We then again computed cell type scores for each scATAC cell cluster by averaging a set of known marker genes (the same as used for scRNA-seq) for each cluster and plotting results along with hierarchical clustering, PCA and MDS analysis of the pseudo-bulk gene scores. Based on these analyses, we removed outlier clusters (for example, very low complexity) and assigned a final cell type annotation. Based on this cleaned dataset, we performed differential accessibility analysis, aggregating the signal of all cells in each cluster using Cicero's aggregate\_by\_cell\_bin function, followed by a negative binomial-based differential accessibility test implemented in the differentialGeneTest function using the cluster ID as a factor to test. In addition to the cell-cluster-based test, we also performed a second test on the timepoint variable to identify specifically those peaks variable across distinct timepoints. We then selected all peaks below a  $q$  value of  $0.5$  in either of the two analyses as a dynamic peak set for subsequent analysis. The results are not particularly sensitive to this threshold, as we tested  $q$  value thresholds between  $0.1$  and  $0.5$ . However, given the limited power for each peak, we decided

to include all peaks with some evidence for differential accessibility in further clustering analysis. We opted to use a 0.5 threshold, as this gave us a similar fraction of variable peaks as the bulk ATAC analysis conducted earlier.

All subsequent analyses were conducted on this dynamic peak set, after excluding all cells of unassigned identity. Next, we normalized the read counts for each cell by the respective size factor and averaged the resulting values across all cells within one cluster, giving rise to pseudo-bulk profiles. Next, we normalized these profiles for each peak across all pseudo-bulk clusters by dividing each row (corresponding to peaks) by the 95th quantile across all cell clusters, capping all values at 1. This gave rise to a normalized accessibility score in the unit interval. These values were then used for clustering analysis, following the same strategy as for the scRNA-seq, again assigning each cell cluster to one of seven categories and then performing k-means with 100 random starts and clustering on the respective subset of cell clusters for all peaks in the dynamic peak set. Again, the cluster number was determined using the `clusGap` function in the R package `cluster` and Tibshirani 2001 SE criterion. The resulting peak clusters were then subjected to the same characterization as the peak clusters from the bulk ATAC analysis. For the specificity analysis, the cluster coordinates were lifted from mm10 to mm9 using the UCSC liftover tool (<http://genome.ucsc.edu/>)<sup>105</sup>.

**Defining early, late and cell-type-specific clusters.** To assign clusters to a particular activity pattern to individual region clusters, we predefined a set of patterns according to possible dynamics of interest. These included temporal neurons (for example, specific to P1, P7, P21 or P1\_P7 or Y0 and Y2), excitatory neurons and inhibitory neurons as well as specific to a particular cell cluster. These patterns were summarized in a prototype binary indicator matrix of dimensions number of cell clusters times number of patterns. For each pattern and cell cluster, the indicator matrix cell was set to 1 if that cell cluster was associated with that pattern—for example, setting the indicator variable for P1\_CPN\_1 and P1\_CPN\_3 both to 1 in the P1 pattern definition.

We then computed the cosine similarity between each of these pattern vectors with the normalized accessibility score vector for each region cluster, taking the average of accessibility score for all regions within one region cluster for each cell cluster separately. This analysis gave rise to a similarity measure of each cluster's dynamic to a predefined set of temporal/cell type dynamics. Based on the maximum observed similarity to a predefined pattern, we then assigned the pattern label to the corresponding cluster.

**Mouse single-cell ATAC peak methylation analysis.** Before DNA methylation analysis, we performed a liftover of the size-standardized (500-bp) scATAC peaks from mm10 to mm9 using the UCSC liftover tool. Next, we computed DNAm levels in a similar manner as for the bulk data, using the scATAC peaks instead of the DMR regions as input, considering only regions that were covered by at least five reads in at least eight of the 12 WGBS samples. We report the distribution of methylation levels for each of the scATAC region clusters across CPN and CThPN neurons at P1, P21 and P48.

**Similarity in cell types across ages.** To compute the similarities among cell types within each age group and compare them, we first merged single-cell data into pseudo-bulk by computing the average expression for each cell type at each timepoint (using the `AverageExpression` function from Seurat package version 3.1.0 in R version 3.6.3). We then performed PCA (using the `prcomp` function included in R as its base package), taking the top ten PCs (ordered by the fraction of total variance explained), to project the data onto the two dimensions using the UMAP algorithm (`umap` package version 0.2.4.0 in R).

**GO enrichment analysis.** We prepared two input gene lists for each organism (shared-early and shared-late in the neuronal cell population) based on our gene module clustering. To query marmoset genes, we first converted them into human (GRCh38.p13) orthologous genes (using the Ensembl database). Enrichment analysis was done using the Panther overrepresentation test available through the Gene Ontology Consortium<sup>106–108</sup> web interface. Test type was set to Fisher's exact test, and FDR was computed for each term. Selected GO terms for the shared developmental clusters are presented in Supplementary Table 4.

**Computation of mouse/marmoset gene and open chromatin cluster similarity.** To determine overlap of mouse and marmoset gene clusters with each other, mouse Ensembl gene IDs were mapped to marmoset gene IDs using the Ensembl homolog database, and only mappable genes with known homologs were retained for further interspecies analysis. With this mapping in hand, we determined the overlap of all mouse and marmoset gene clusters, determining significance and odds ratio using Fisher's exact test with the union of all mappable and clustered genes (containing only genes differentially expressed in each dataset) as background set. These results are shown in Fig. 4g.

Open chromatin cluster overlap based on scATAC-seq between mouse and marmoset clusters was determined by first mapping size-standardized (500-bp) clustered marmoset open chromatin peaks to the mouse genome using `bnMapper` from the `bx-python` software suite (<https://github.com/bxlab/bx-python>) and mapping file `calJac3ToMm10.over.chain.pkl` downloaded from the UCSC genome

browser website. Subsequently, the resulting mapped regions were filtered to be at least 100 bp in size and showing a ratio of mapped region size to original region size of less than 1.2. All regions fulfilling these criteria were used for further analysis. We then compared the pairwise overlap in terms of genomic regions for each mouse and marmoset open chromatin cluster and report the percentage of overlap in Fig. 4f.

Similarly, we computed the overlap with open chromatin regions in human fetal cerebrum from ref.<sup>58</sup> (downloaded from <https://descartes.brotmanbaty.org/>) for mouse and marmoset scATAC-based clusters, mapping them to hg19 analogous to the strategy described above. Although this dataset is not restricted to the cortex, at the time of this study comparably large single-cell ATAC datasets for the human fetal cortex were not available.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The datasets generated in the current study are available through the Gene Expression Omnibus (GEO SuperSeries GSE204851).

The following publicly available datasets were used:

*Mus musculus* Ensembl genome build NCBIM37 (mm9); *Callithrix jacchus* (common marmoset) genome assembly version ASM275486v1.93; human developing brain transcriptomic data from BrainSpan<sup>43</sup> (<http://brainspan.org/rnaseq/search/index.html>); mouse enhancer data from the VISTA enhancer database<sup>52</sup> (<http://enhancer.lbl.gov/>); mouse CAGE-tag data from the FANTOM5 database<sup>44</sup> (<https://fantom.gsc.riken.jp/5/>); mouse ISH expression data from the Allen Brain Atlas Developing Mouse Brain database (<http://developingmouse.brain-map.org/>); mouse single-cell ATAC-seq data from Cusanovich et al.<sup>17</sup> (GEO: GSE111586); DNase HS I peak tracks from the mouse ENCODE consortium<sup>46</sup> ([www.encodeproject.org](http://www.encodeproject.org/)); and human fetal cerebrum single-cell ATAC-seq data from Domcke et al.<sup>58</sup> (<https://descartes.brotmanbaty.org/>).

## Code availability

All custom scripts are available from the corresponding authors upon reasonable request.

## References

- Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
- NIH Neuroscience Blueprint Cre Driver Network. Cre recombinase-expressing mice generated for the NIH Neuroscience Blueprint Cre Driver Network. MGI Direct Data Submission <http://www.credrivermice.org/database> (2009).
- Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
- Fame, R. M., MacDonald, J. L. & Macklis, J. D. Development, specification, and diversity of callosal projection neurons. *Trends Neurosci.* **34**, 41–50 (2011).
- Gil-Sanz, C. et al. Lineage tracing using *Cux2-Cre* and *Cux2-CreERT2* mice. *Neuron* **86**, 1091–1099 (2015).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Kolde, R. `heatmap: Pretty Heatmaps v. 1.0.12`. <https://cran.r-project.org/web/packages/heatmap/index.html> (2019).
- Ma, S. et al. Cell-type-specific brain methylomes profiled via ultralow-input microfluidics. *Nat. Biomed. Eng.* **2**, 183–194 (2018).
- LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228 (2020).
- Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Kim, D. et al. TopHat2: accurate alignment of transcriptsomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
- Maechler, M., Rousseau, P., Struyf, A., Hubert, M., Hornik, K. `cluster: Cluster Analysis Basics and Extensions`. R package version 2.1.3. <https://CRAN.R-project.org/package=cluster> (2022).
- Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **63**, 411–423 (2001).

76. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
77. Broad Institute of MIT and Harvard. Picard. <http://broadinstitute.github.io/picard/>
78. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
79. Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP-Seq peak data. <http://bioconductor.org/packages/release/bioc/html/DiffBind.html> (2011).
80. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
82. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
83. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* **10**, 232 (2009).
84. Sun, D. et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* **15**, R38 (2014).
85. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
86. Park, Y. & Wu, H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* **32**, 1446–1453 (2016).
87. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
88. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
89. Ravasi, T. et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752 (2010).
90. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* **18**, 205–214 (2017).
91. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
92. Bock, C. et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell* **47**, 633–647 (2012).
93. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
94. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **24**, 238–241 (1996).
95. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
96. Cock, P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
97. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
98. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
99. Zhu, L. J. in *Tiling Arrays: Methods and Protocols* (eds Lee T.-L. & Luk, A. C. S.) 105–124 (Humana Press, 2013).
100. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
101. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
102. Baker, S. M., Rogerson, C., Hayes, A., Sharrocks, A. D. & Rattray, M. Classifying cells with Scasat, a single-cell ATAC-seq analysis tool. *Nucleic Acids Res.* **47**, e10 (2019).
103. Hill, A. Dimensionality Reduction for scATAC Data. <http://andrewjohnhill.com/blog/2019/05/06/dimensionality-reduction-for-scatac-data/> (2019).
104. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 (2018).
105. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
106. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
107. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–d338 (2019).
108. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

## Acknowledgements

We thank the laboratory of G. Feng for providing marmoset tissue samples. We thank H.-H. Chen for advice and assistance in neuron isolation techniques. We thank B. Paulsen for assistance with preparing figures. We thank members of the Arlotta laboratory for insightful comments and suggestions. This work was supported by grants from the Stanley Center for Psychiatric Research at Broad Institute, the Harvard Stem Cell Institute and the US National Institutes of Health (NIH) (R01MH101268, R01NS078164 and U19MH114821) to P.A., by BMBF eMed program grant 01ZX1504 to M.J.Z. and by Cariplo Grant 2019-1785 to S.L. During the course of this work, P.A. was a NYSCF Robertson Investigator. J.D.B. and the Buenrostro laboratory acknowledge support from the Allen Distinguished Investigator Program through the Paul G. Allen Frontiers Group, the Chan Zuckerberg Initiative and the NIH New Innovator Award (DP2 mechanism). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

W.Y. and P.A. conceived all experiments. W.Y. performed all experiments. M.J.Z., S.M., V.M., L.T. and K.K. performed data analysis. S.L. and A.S.S. contributed to cell type identification and marker gene selection for single-cell/single-nucleus experiments. J.R.B. assisted with preparation of graphs and figures. The work was supervised by P.A., M.J.Z., J.D.B., J.Z.L. and A.M. The manuscript was written by W.Y., J.R.B., M.J.Z. and P.A., with contributions from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41593-022-01123-4>.

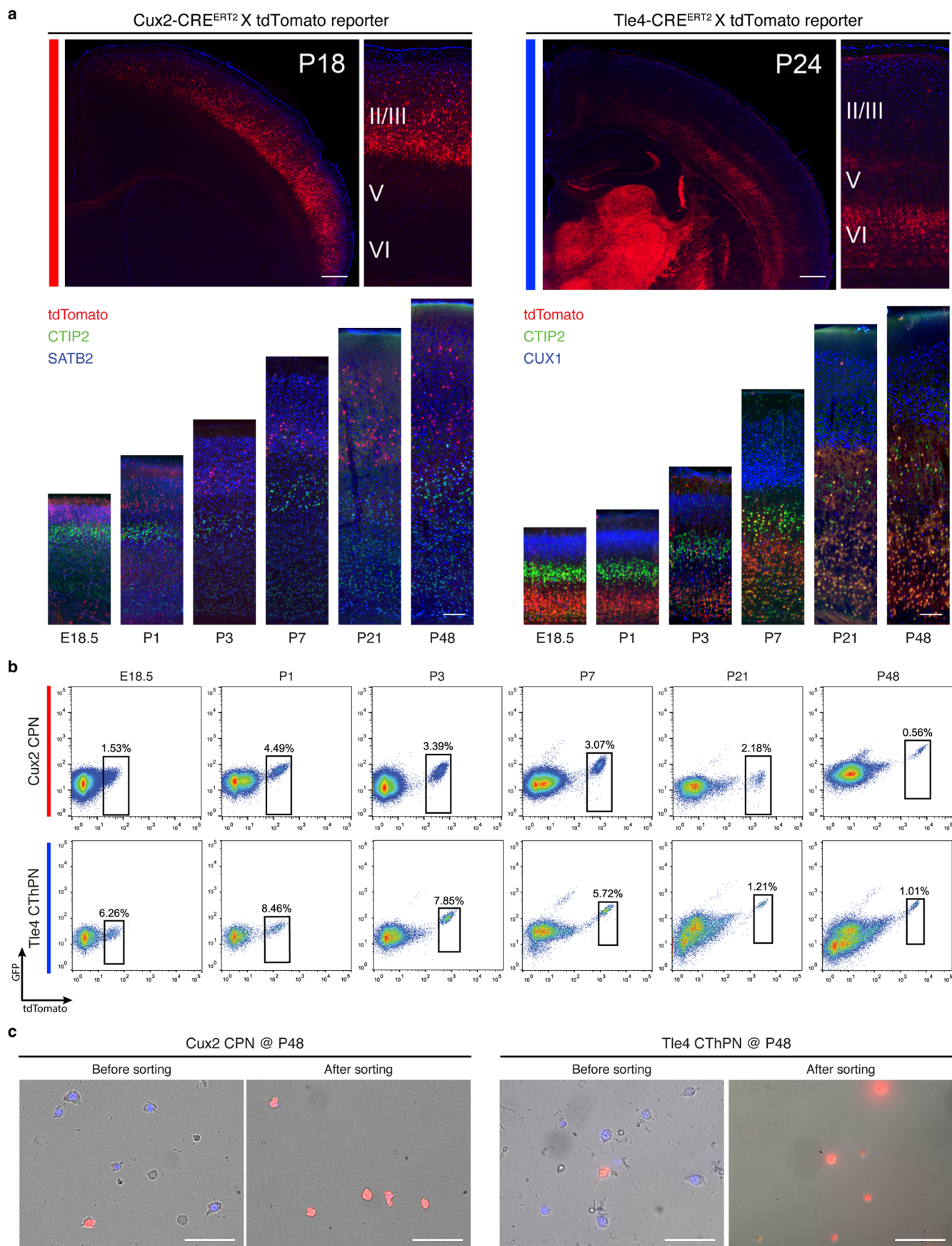
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41593-022-01123-4>.

**Correspondence and requests for materials** should be addressed to Michael J. Ziller or Paola Arlotta.

**Peer review information** *Nature Neuroscience* thanks Andrew Adey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

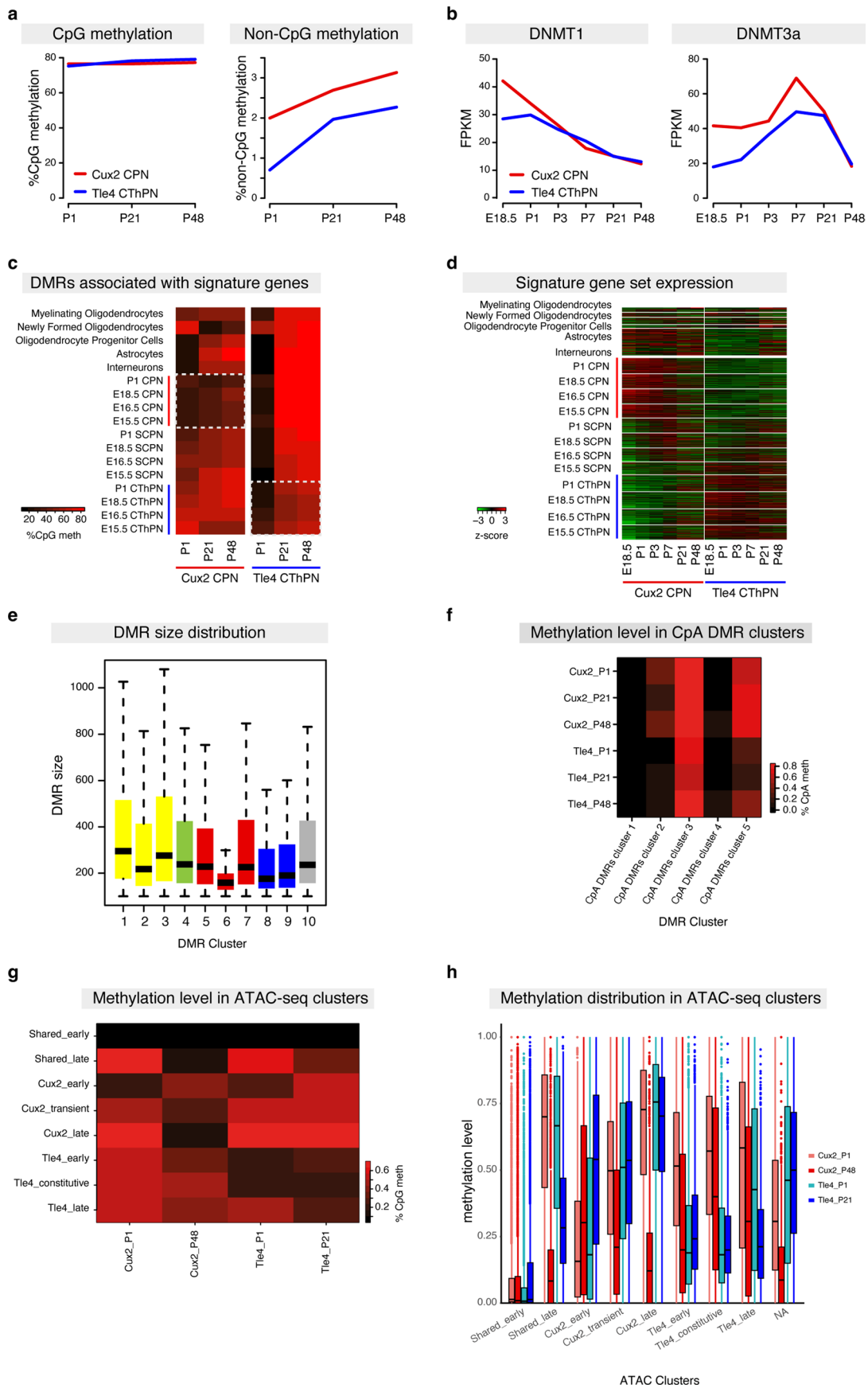
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).





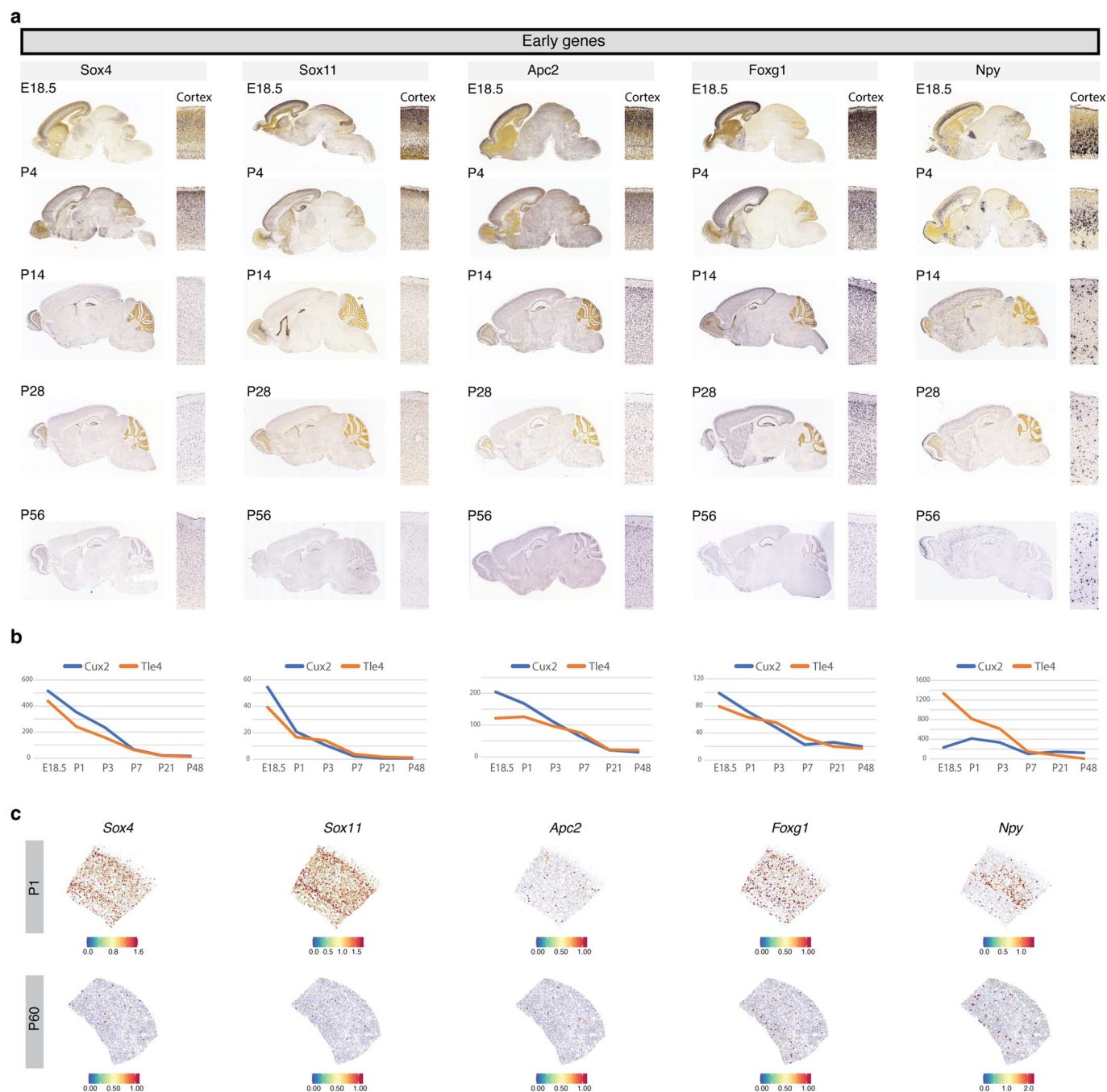
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Isolation of genetically-labelled CPNs and CThPNs.** **a**, Representative coronal sections ( $n \geq 3$  biological replicates) showing correct laminar location of tdTomato<sup>+</sup> cells in the somatosensory cortex at the different developmental stages. Upper panels, scale bar 500  $\mu\text{m}$ . Lower panels, scale bar 100  $\mu\text{m}$ . **b**, Representative FACS results for purification of Cux2 CPN and Tle4 CThPN populations at the different developmental stages. Boxed areas indicate tdTomato<sup>+</sup> cells collected for profiling. The GFP channel (empty) is used for evaluation of background signal. **c**, Representative images of pre- and post-sorting Tle4 and Cux2 neuronal populations at P48, showing good purity and viability even at this older age ( $n \geq 3$  biological replicates). Scale bar, 50  $\mu\text{m}$ .

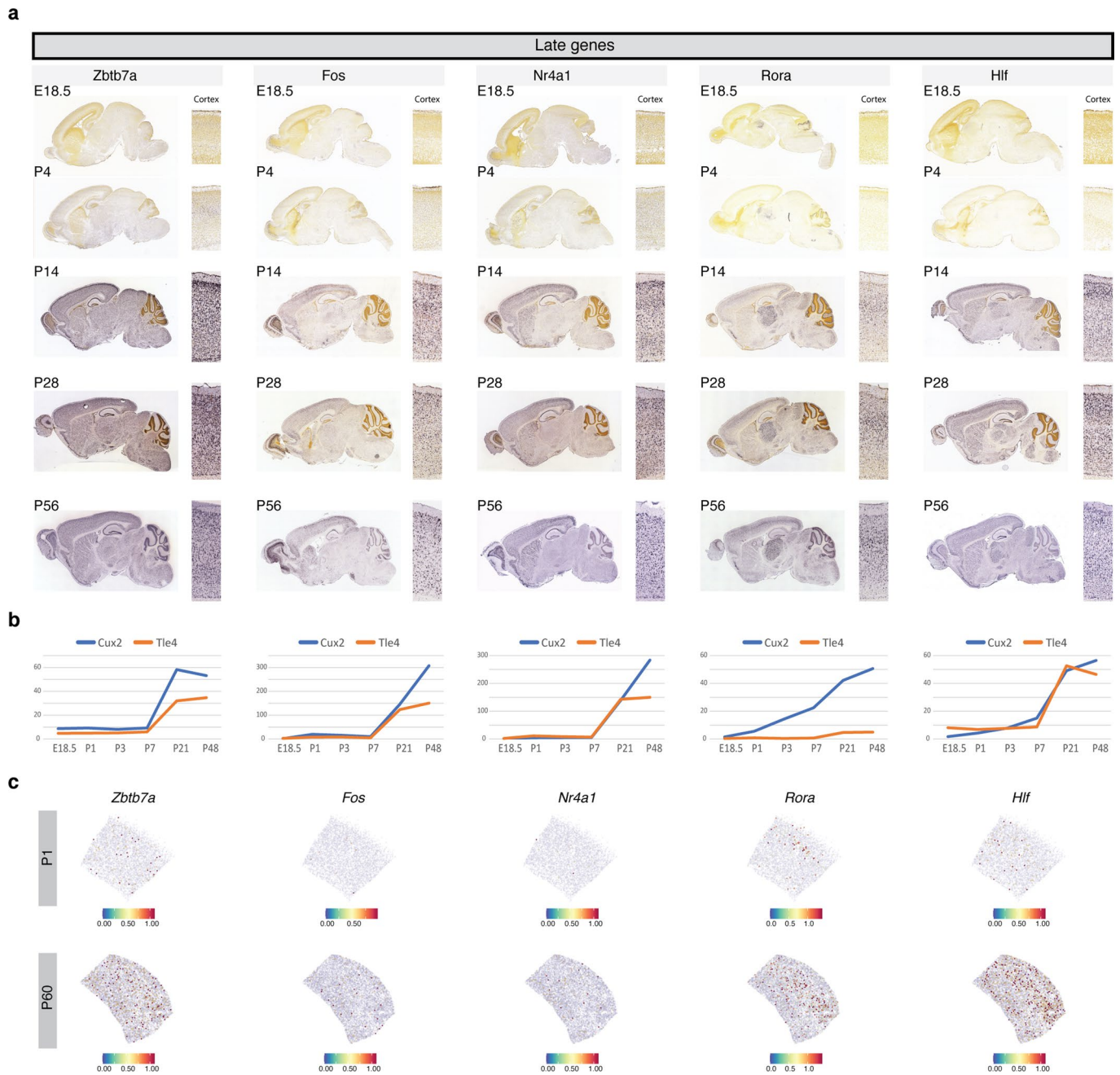


Extended Data Fig. 2 | See next page for caption.

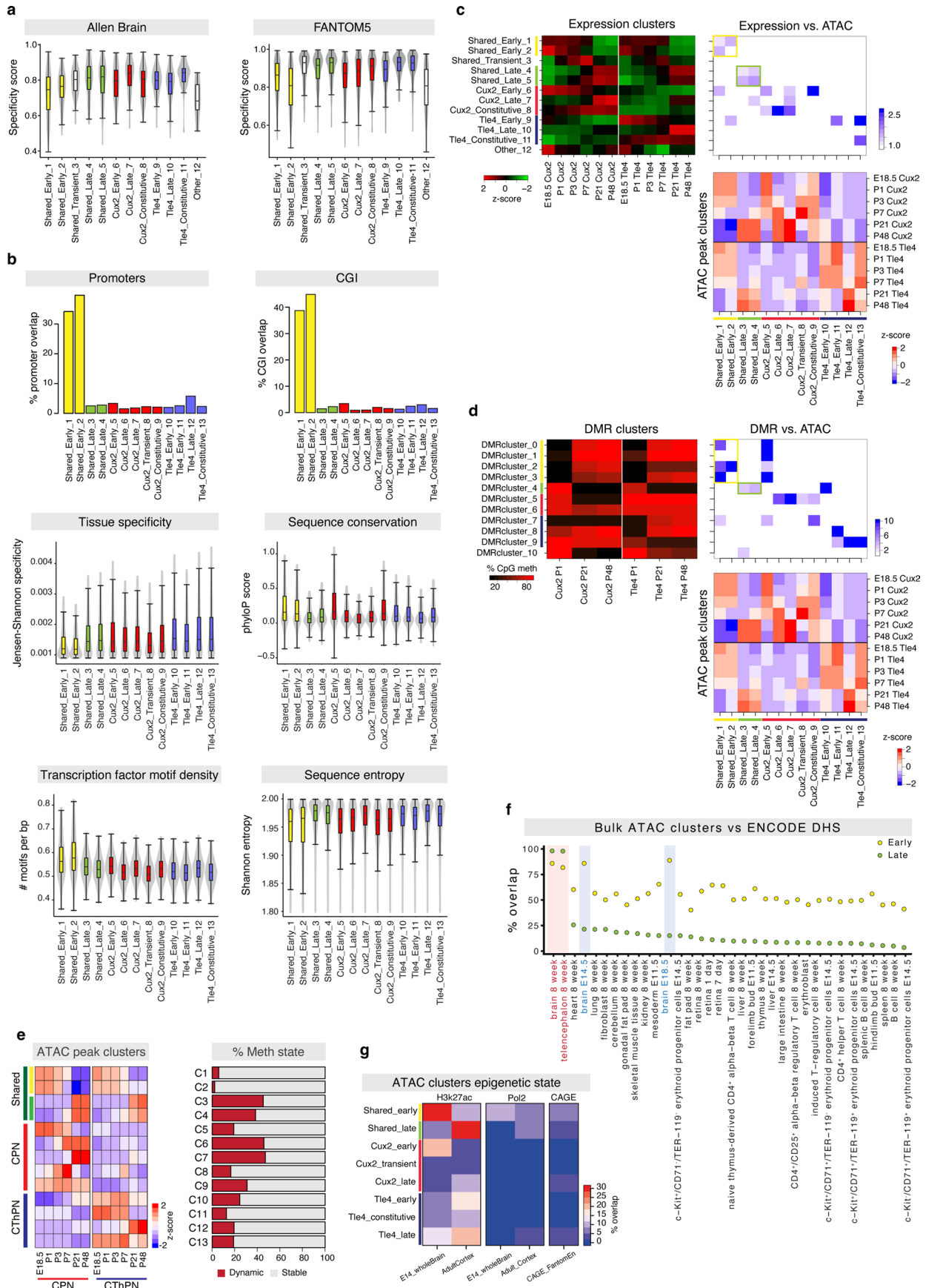
**Extended Data Fig. 2 | DNA methylation at dynamic loci.** **a**, Global DNA methylation levels at CpG (left) and non-CpG sites (right) across ages. **b**, Average expression of differentially expressed DNA methyltransferases across ages. **c**, Median methylation levels of all DMRs within 100 kb of signature genes for the specified cell types at each age. Boxes indicate gene sets corresponding to earlier ages of the same cell fate. **d**, Normalized gene expression patterns, shown as Z-scores, for the sets of signature genes used in panel **c**. **e**, Size distribution of DMRs within each cluster. Bar: median; box: 25–75th quantile; whisker: 1.5 x inter-quartile range. For size of clusters, see Supplementary Fig 3c. **f**, Average CpA methylation levels across 11,150 CpA DMRs clustered into 5 distinct clusters across developmental time and PN subtypes. **g**, Average methylation levels for ATAC-seq regions with overlapping DNAm data for each ATAC-seq cluster (y axis) across developmental time and neuronal subtype (x-axis) based on bulk ATAC-seq data. **h**, Methylation level distribution (y axis) for ATAC-seq regions with overlapping DNAm data for each ATAC-seq cluster (x axis) across developmental time and neuronal subtype (colored boxplots) based on bulk ATAC-seq data. Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5 interquartile range from box. Dots indicate data points outside this range.



**Extended Data Fig. 3 | Validation of temporal patterns of early gene expression using external datasets. a**, Expression of selected genes in the shared-early cluster at (from top to bottom) E18.5, P4, P14, P28 and P56, in the mouse brain in situ hybridization (ISH) database from the Allen Institute, showing the whole brain and an enlargement of the cortex. **b**, Gene expression values for the corresponding genes in the Cux2-CPN (red) and Tle4-CThPN (blue) RNA-seq data. Units: fragments per kilobase of exon per million reads mapped (FPKM). **c**, Gene expression level in P1 and P60 mouse cortex by Slide-seq<sup>42</sup>.



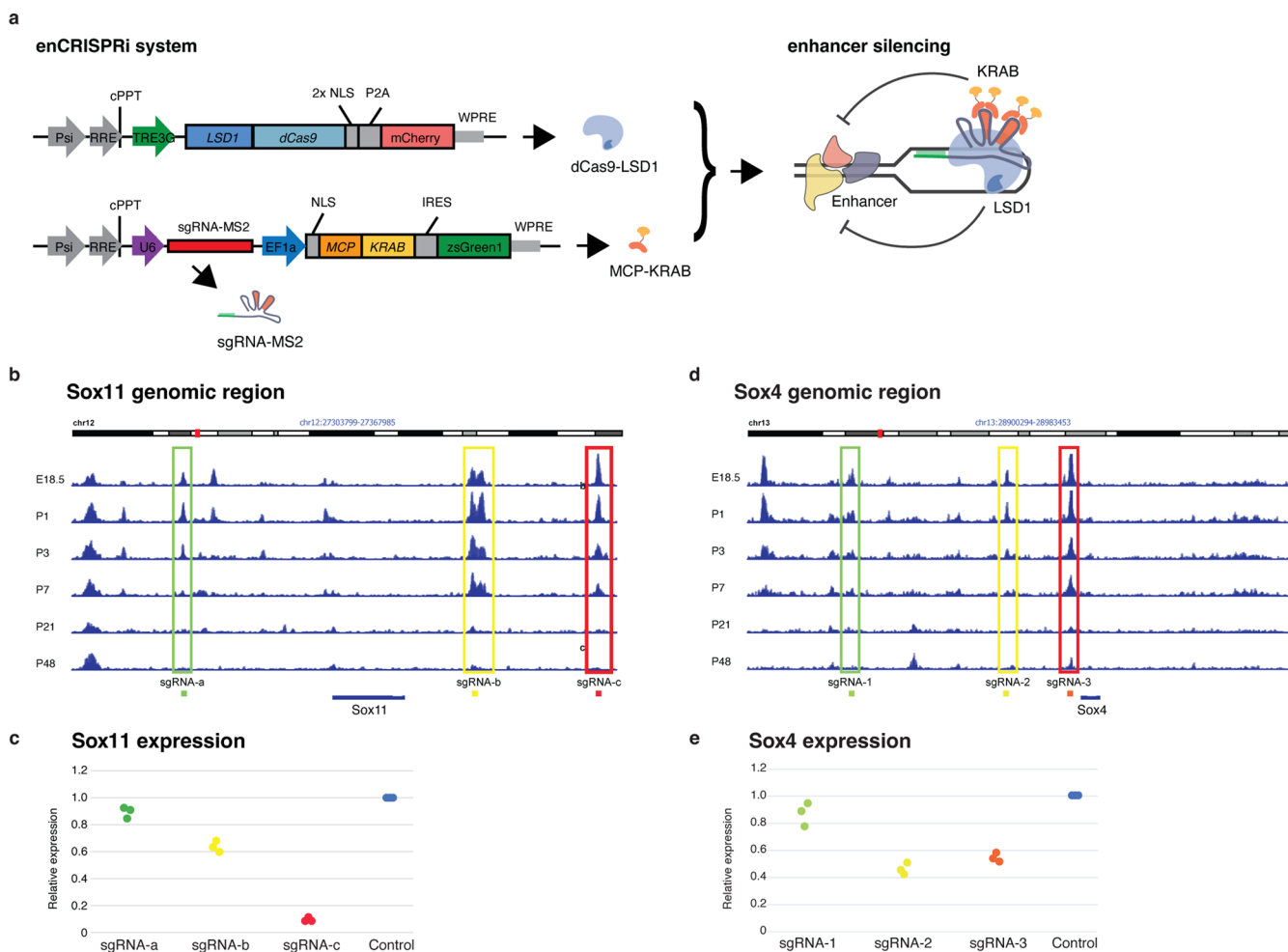
**Extended Data Fig. 4 | Validation of temporal patterns of late gene expression using external datasets. a**, Expression of selected genes in the shared-late cluster at (from top to bottom) E18.5, P4, P14, P28 and P56, in the mouse brain in situ hybridization (ISH) database from the Allen Institute, showing the whole brain and an enlargement of the cortex. **b**, Gene expression values for the corresponding genes in the Cux2-CPN (red) and Tle4-CThPN (blue) RNA-seq data. Units: fragments per kilobase of exon per million reads mapped (FPKM). **c**, Gene expression level in P1 and P60 mouse cortex by Slide-seq<sup>42</sup>.



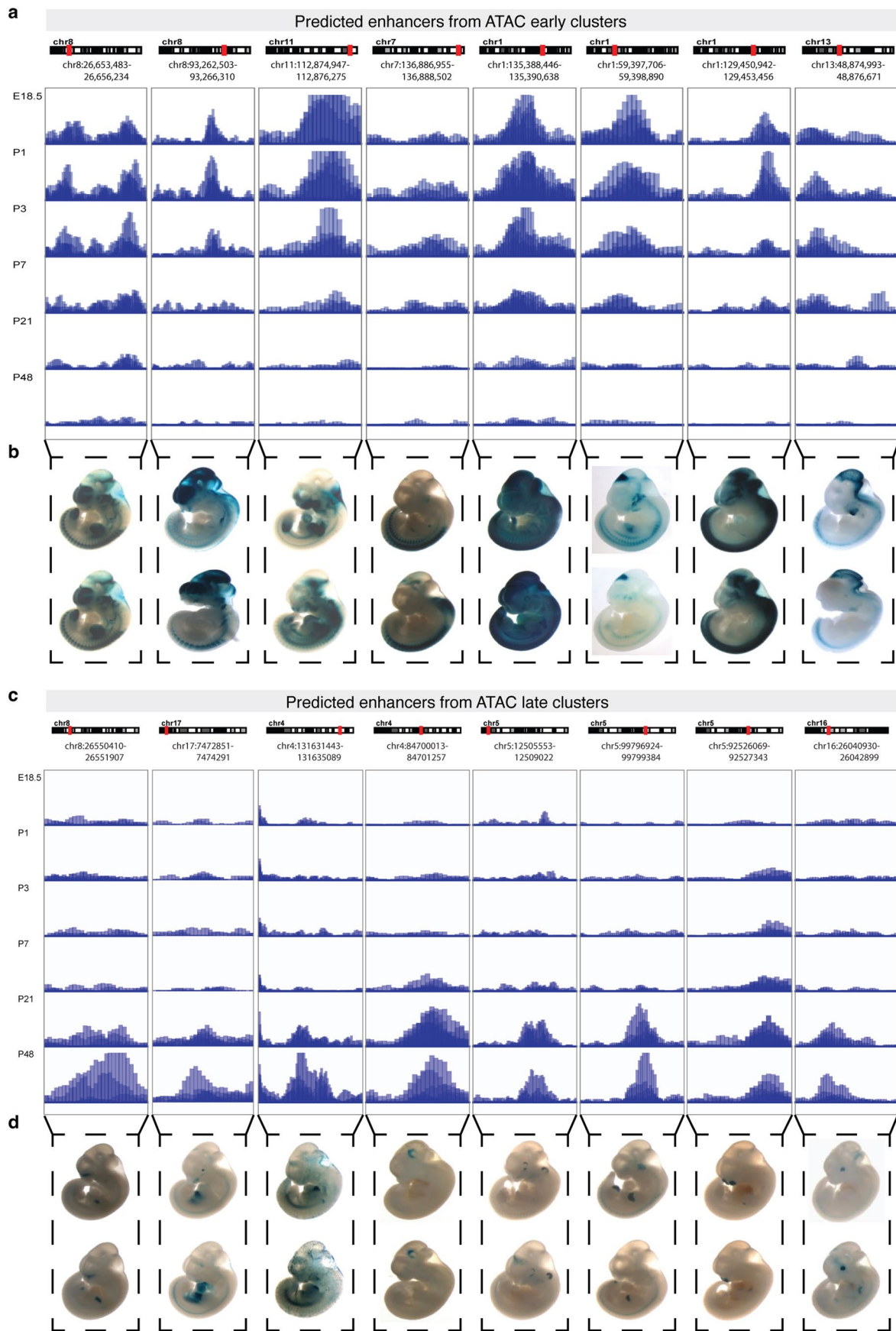
Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Characterization of RNA-seq and ATAC-seq clusters by cluster.** **a**, Expression specificity of genes in each RNA-seq cluster as in Fig. 2c, for each individual cluster. Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5 interquartile range from box. **b**, Analyses of ATAC peak clusters as in Fig. 2e and H-K, for each individual cluster. Boxplot as in **a**. **c**, Overlap between individual RNA-seq and ATAC clusters. Color scale for top-right panel: odds ratio, Fisher's exact test, only results with odds ratio >1, BH corrected  $P$  value  $\leq 0.001$ , and more than 50 overlapping regions are shown. **d**, Overlap between individual DNase I and ATAC clusters. Color scale for top-right panel: odds ratio, Fisher's exact test. **e**, Fraction of DMRs in the ATAC clusters that are dynamic versus static over time as in as in Fig. 2f, for each ATAC cluster. **f**, Overlap of early and late shared ATAC clusters with DNase I hypersensitivity sites in a panel of tissues and cell types from the ENCODE database, by tissue. **g**, Percentage of overlapping ATAC-seq regions for each bulk ATAC-seq cluster for CPN and CThPN (y axis) with genomic features derived from E14 whole mouse brain and adult cortex H3K27ac and polymerase 2 ChIP-Seq from ENCODE as well as enhancers defined by CAGE-Seq from the FANTOM5 consortium (x axis), excluding all ATAC-Seq regions overlapping with annotated Ensembl promoters.



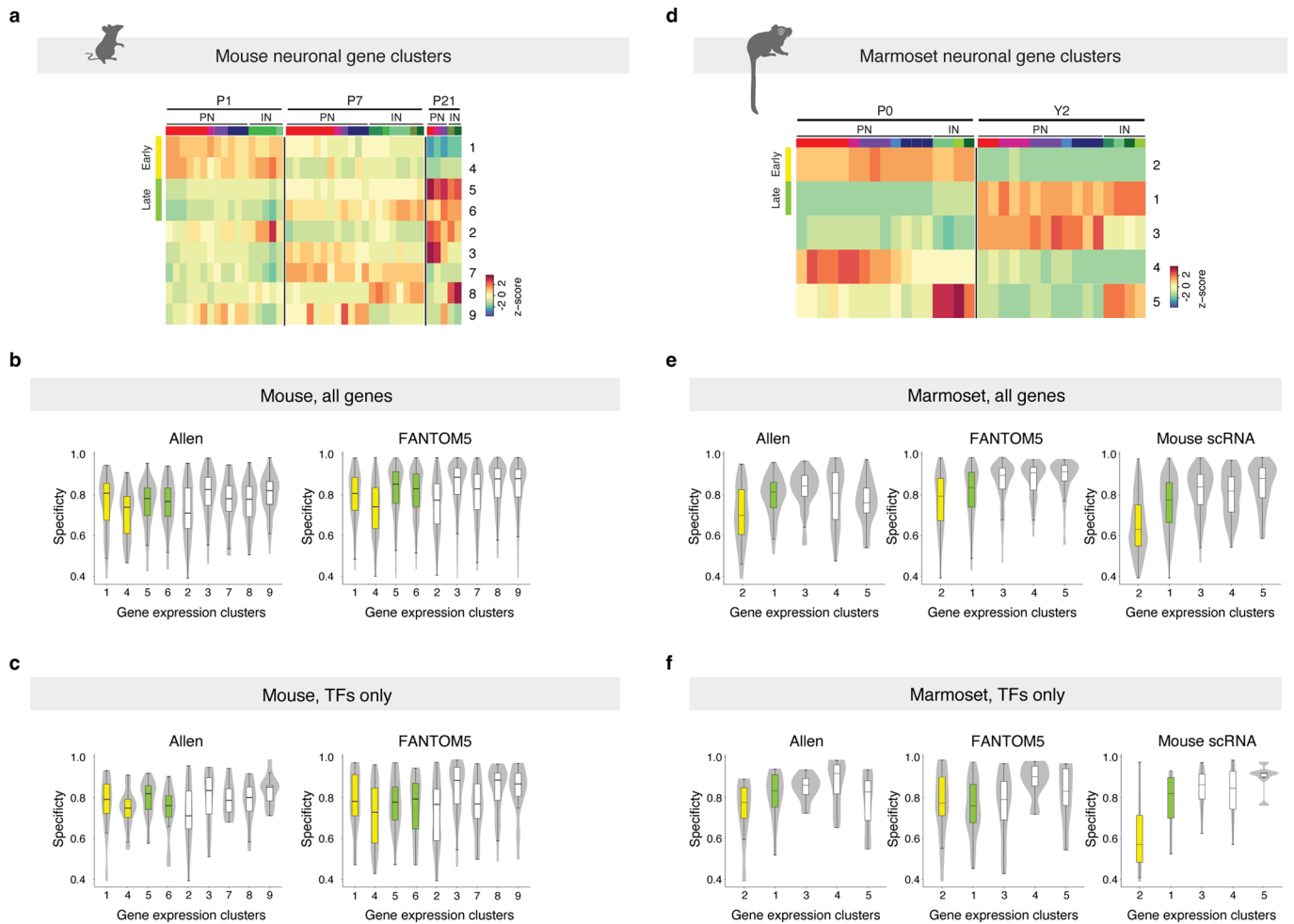


**Extended Data Fig. 6 | Validation of Sox4 and Sox11 enhancers.** **a**, Schematic of enCRISPRi system<sup>48</sup>. **b**, Predicted novel enhancer regions for Sox11 in the mouse bulk ATAC-seq dataset, and location of selected sgRNAs. **c**, Relative Sox11 expression by qPCR after silencing the indicated putative enhancer regions for 5 days ( $n=3$  biological replicates, each performed in duplicate, normalized to an irrelevant control sgRNA). **d**, Predicted novel enhancer regions for Sox4 in the mouse bulk ATAC-seq dataset, and location of selected sgRNAs. **e**, Relative Sox4 expression after silencing the indicated putative enhancer regions for 5 days.

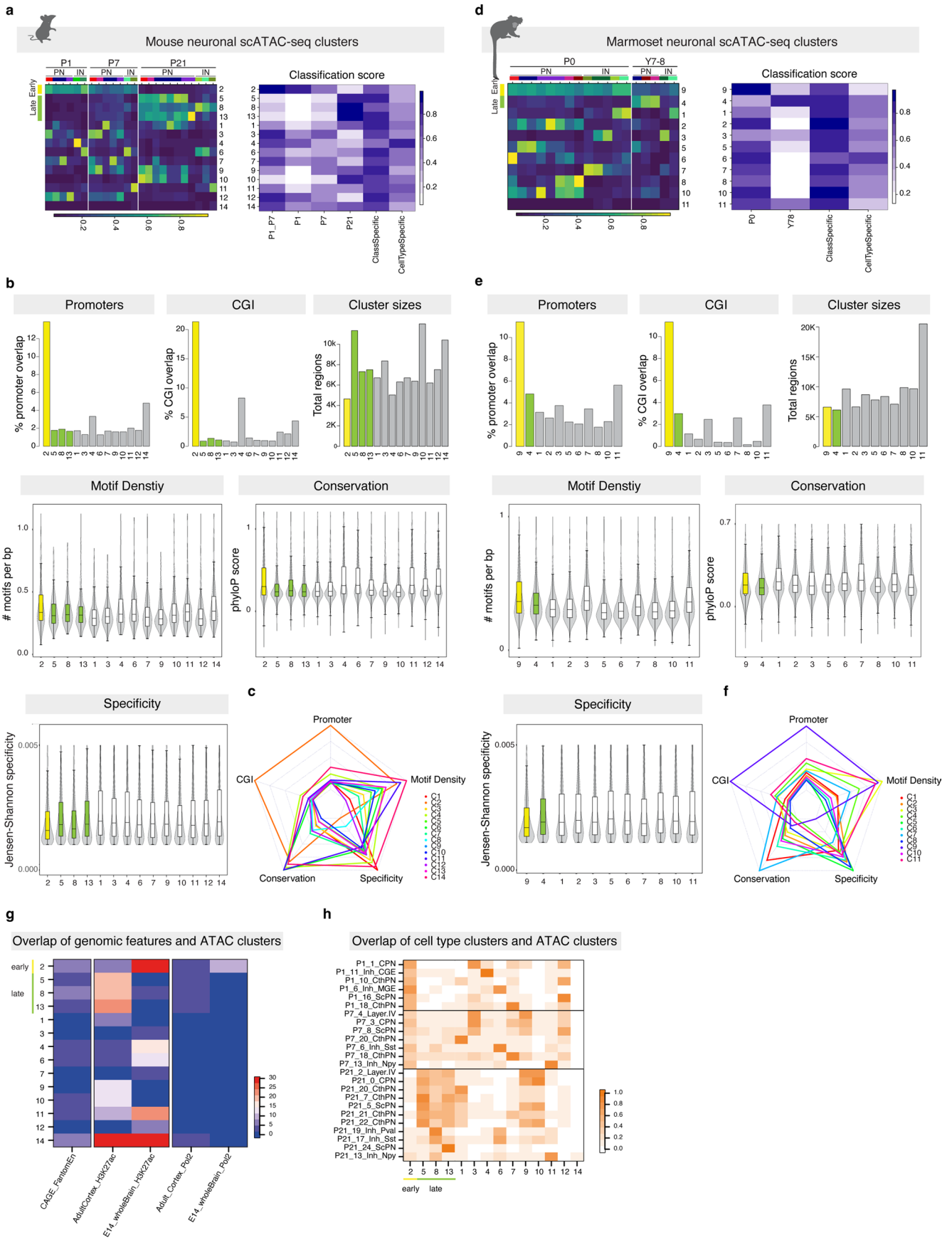


Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Validating specificity of shared-early and shared-late accessible chromatin sites in the VISTA enhancer dataset.** **a**, Example enhancer regions in the VISTA database that span ATAC peaks in the shared-early ATAC cluster, overlaid with ATAC peak accessibility in our data at each age. Top row shows the position of the region on the chromosome. **b**, Expression patterns driven by that DNA segment in E11.5 mouse embryos, from the VISTA dataset. Two examples are shown for each. **c**, Example enhancer regions in the VISTA database that span ATAC peaks in the shared-late ATAC cluster, overlaid with ATAC peak accessibility in our data at each age. Top row shows the position of the region on the chromosome. **d**, Expression patterns driven by that DNA segment in E11.5 mouse embryos, from the VISTA dataset. Two examples are shown for each.

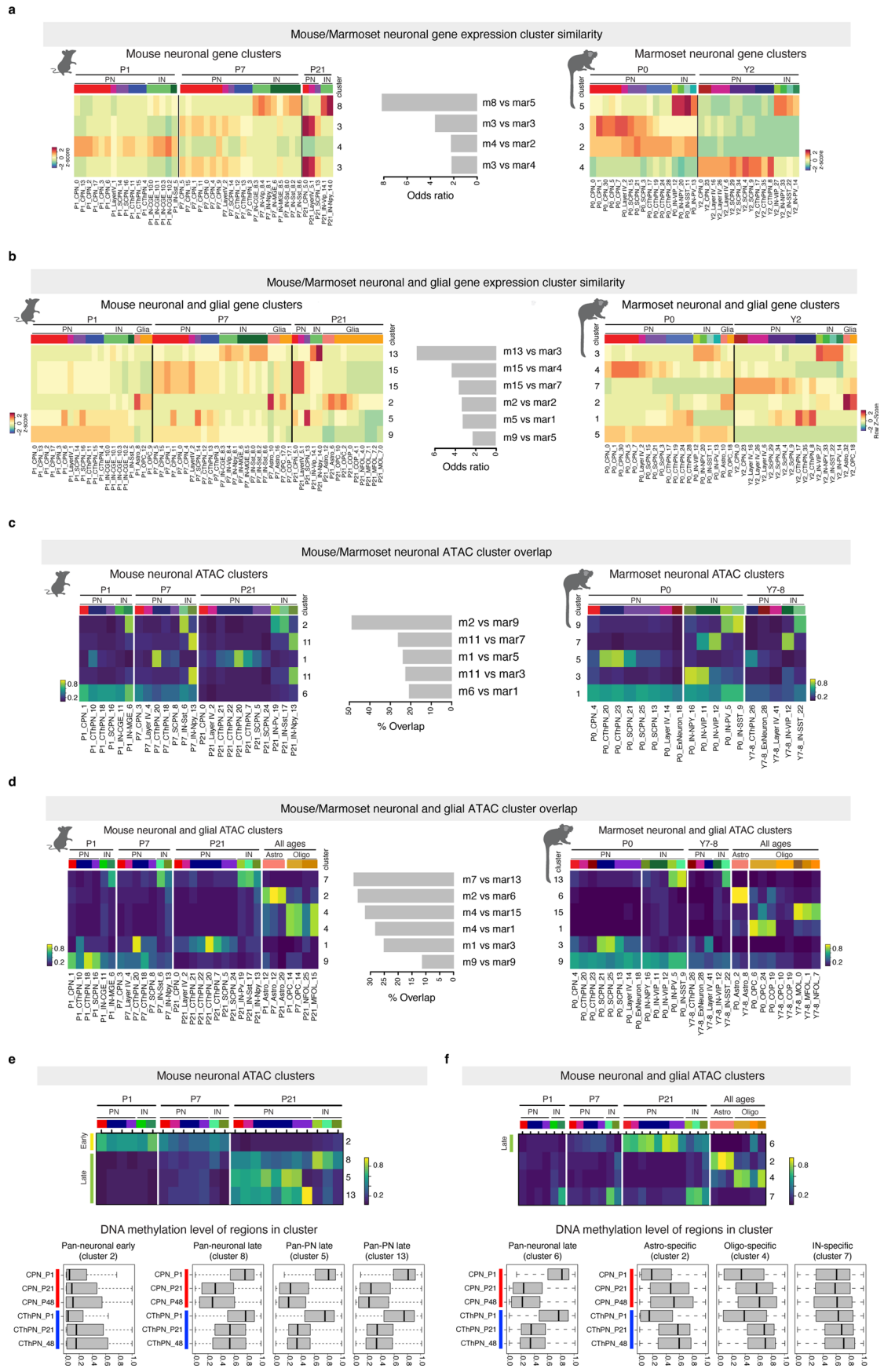


**Extended Data Fig. 8 | Gene cluster specificity.** **a**, Mouse neuronal gene clusters, repeated from Fig. 3 for ease of reference. **b**, Cell-type specificity of all genes in the Allen (left) and FANTOM5 (right) datasets, for each gene cluster. Box and midline: 25, 50 and 75 percentiles; whiskers: 1.5 interquartile range from box. **c**, Cell-type specificity of transcription factors (TFs) in the Allen (left) and FANTOM5 (right) datasets, for each gene cluster. Boxplot as in **b**. **d**, Marmoset neuronal gene clusters, repeated from Fig. 3 for ease of reference. **e**, Cell-type specificity of all genes in the Allen (left) and FANTOM5 (right) datasets, and in our mouse scRNA-seq dataset for each gene cluster. Boxplot as in **b**. **f**, Cell-type specificity of transcription factors in the Allen (left) and FANTOM5 (center) datasets, and in our mouse scRNA-seq dataset (right), for each gene cluster. Boxplot as in **b**.



Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Specificity analyses of mouse and marmoset scATAC peak clusters.** **a**, Mouse ATAC peak clusters, reproduced from Fig. 4 for ease of reference (left), and correlation with age versus cell type for each cluster (right). **b**, Mouse ATAC cluster characteristics as in Fig. 4, for each individual cluster, and number of regions per cluster (cluster sizes). Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5 interquartile range from box. **c**, Summary of characteristics for each cluster in mouse. Each arm of the plot represents an individual metric (center = low, edge = high). Boxplots as in **b**. **d**, Marmoset ATAC peak clusters, reproduced from Fig. 4 for ease of reference (left), and correlation with age versus cell type for each cluster (right). **e**, Marmoset ATAC cluster characteristics as in Fig. 4, for each individual cluster, and number of regions per cluster (Cluster sizes). Boxplots as in **b**. **f**, Summary of characteristics for each cluster in marmoset. Each arm of the plot represents an individual metric (center = low, edge = high). **g**, Percentage of overlapping ATAC-seq regions for scATAC-seq based clusters for all neuronal cell types (y axis) with genomic features derived from E14.5 whole mouse brain and adult mouse cortex H3K27ac and polymerase 2 ChIP-Seq from ENCODE as well as enhancers defined by CAGE-Seq from the FANTOM5 consortium (x axis), excluding all ATAC-seq regions overlapping with annotated ensemble promoters. **h**, Previously defined average open chromatin signal for neuronal mouse scATAC clusters (x axis) shown in a across all neuronal cell types and developmental stages assessed (y axis).



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 | Correlation of mouse versus marmoset gene and ATAC clusters.** **a**, Overlap between mouse and marmoset gene clusters for neuronal cell types only (as presented in text), showing the top 4 pairs. **b**, Overlap between mouse and marmoset gene clusters for neuronal and glial cell types taken together (for comparison as a control), showing the top 6 pairs. **c**, Overlap between mouse and marmoset ATAC clusters for neuronal cell types only, showing the top 5 pairs. Color scale: average peak normalized read count. **d**, Overlap between mouse and marmoset ATAC clusters for neuronal and glial cell types taken together, showing the top 4 pairs. Color scale: average peak normalized read count. **e**, Methylation of ATAC peaks in each mouse shared-developmental neuronal ATAC cluster in the mouse bulk DNAm dataset in Fig. 1. Box and midline: 25th, 50th and 75th percentiles; whiskers: 1.5 interquartile range from box. Color scale for ATAC: average peak normalized read count. **f**, Methylation of ATAC peaks in each mouse shared-developmental neuronal and glial ATAC cluster, determined from the mouse bulk WGBS dataset presented in Fig. 1. Boxplots as in **e**. Color scale for ATAC: average peak normalized read count.



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed   |
|-------------------------------------|---|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

The following programs were used:

Cell Ranger version 3.0.1  
 Trimmomatic version 0.33  
 TopHat2 version 2.0.13  
 Cuffdiff version 2.2.1  
 Bowtie2 version 2.3.2  
 Picard toolkit version 2.7.1  
 MACS2 version 2.1.1  
 MOABS version 1.3.2  
 SAMtools version 1.5  
 sequtils version 1.74  
 bcl2fastq  
 R package DSS version 2.14  
 methylKit version 0.9.5  
 GREAT version 3.0.0  
 R version 3.5.1  
 R package Seurat version 2.0.0 and 3.1.5  
 R package cluster 2.0.7

R package ggplot2 version 2.3.3  
 R package cicero version 1.0.15  
 R package IRanges  
 liftOver (UCSC command-line liftover tool)  
 MEME suite version 4.11.2  
 Scasat  
 bx-python version 0.8.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated in the current study are available through the Gene Expression Omnibus (GEO SuperSeries GSE204851).

The following publicly-available datasets were used:

Mus musculus ENSEMBL genome build NCBI37 (mm9); Callithrix jacchus (common marmoset) genome assembly version ASM275486v1.93; Human developing brain transcriptomic data from BrainSpan ([brainspan.org/rnaseq/search/index.html](http://brainspan.org/rnaseq/search/index.html)); mouse enhancer data from the VISTA enhancer database ([enhancer.lbl.gov](http://enhancer.lbl.gov)); mouse CAGE-tag data from the FANTOM5 database ([fantom.gsc.riken.jp/5/](http://fantom.gsc.riken.jp/5/)); mouse ISH expression data from the Allen Brain Atlas Developing Mouse Brain database ([developingmouse.brain-map.org](http://developingmouse.brain-map.org)); mouse single-cell ATAC-seq data from Cusanovich et al. 2018 (GEO: GSE111586); DNase HS I peak tracks from the mouse ENCODE consortium ([www.encodeproject.org](http://www.encodeproject.org)); human fetal cerebrum single-cell ATAC-seq data from Domcke et al. 2020 ([descartes.brotmanbaty.org](http://descartes.brotmanbaty.org)).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications (Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods, Nat. Biotechnol. 2020).
Data exclusions	Quality-control filtering for sequencing data was performed using standard metrics as described in the Methods.
Replication	All experiments included at least two biological replicates, and differences between replicates were evaluated. All attempts at replication were successful.
Randomization	For mouse experiments, animals were randomized during sample collection where possible. For marmoset experiments, sample selection was dictated by tissue availability; all tissue was collected from healthy, un-manipulated individuals.
Blinding	No blinding was performed as the identity of the samples was central to the analysis.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	<p>rat anti-CTIP2 antibody, 1:1,000 (Abcam ab18465)  rabbit anti-CUX1 (CDP M-222), 1:100 (Santa Cruz sc-13024)  mouse anti-SATB2, 1:50 (Abcam ab51502)  Goat anti-rat Alexa 488, 1:1000 (ThermoFisher A48262)  Donkey anti-mouse Alexa 647, 1:1000 (ThermoFisher A-31571)  Donkey anti-rabbit Alexa 647, 1:1000 (Thermo-Fisher A-31573)</p>
Validation	<p>Validation data for Abcam ab51502 is available on the manufacturer's website (<a href="https://www.abcam.com/satb2-antibody-satba4b10-c-terminal-ab51502.html">https://www.abcam.com/satb2-antibody-satba4b10-c-terminal-ab51502.html</a>), for immunocytochemistry (HT10180 cells), Western blot (NIH/3T3 and HT1080 whole cell lysates), and immunoprecipitation (HeLa cell lysate).</p> <p>Validation data for Abcam ab18465 is available on the manufacturer's website (<a href="https://www.abcam.com/ctip2-antibody-25b6-ab18465.html">https://www.abcam.com/ctip2-antibody-25b6-ab18465.html</a>), for flow cytometry (Jurkat cells), immunocytochemistry (neonatal mouse hippocampal cultured neurons), and Western blot (Jurkat cell nuclear extract; mouse brain tissue lysate).</p> <p>Validation data for Santa Cruz sc-13024 is available in the manufacturer's product datasheet (<a href="https://datasheets.scbt.com/sc-13024.pdf">https://datasheets.scbt.com/sc-13024.pdf</a>), for Western blot (K-562 and BJAB nuclear extracts) and immunocytochemistry (HeLa cells and human urinary bladder tissue).</p>

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mouse, male and female, E18.5 to P48; C57BL/6J (JAX 000664) WT, tamoxifen-inducible tdTomato reporter Ai14 (JAX 007914) heterozygous, Cux2-CreERT2 (MMRRC 032779-MU) heterozygous, and Tle4-2A-CreERT2 (JAX 036298) heterozygous.
Wild animals	No wild animals were used in the study.
Field-collected samples	No field-collected samples were used in the study.
Ethics oversight	Experiments using mice were conducted under protocols approved by the Harvard University Institutional Animal Care and Use Committee and followed the guidelines set forth in the National Institute of Health Guide for the Care and Use of Laboratory Animals. All marmoset experiments were approved by the Institutional Animal Care and Use Committee of Massachusetts Institute of Technology and followed the guidelines from the National Institute of Health's Guide for the Care and Use of Laboratory Animals.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

## Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation	Fresh mouse somatosensory and motor cortex was dissociated by papain digestion at 37°C in medium containing 0.8 mM kynurenic acid and 0.05 mM APV. Digestion was stopped with room-temperature medium containing 10 mg/ml each of ovomucoid protease inhibitor and BSA, and tissue was mechanically dissociated by gentle trituration in ice-cold medium with
--------------------	---

	0.4 mM kynurenic acid and 0.025 mM APV. For full details, see Methods.
Instrument	BD FACSAria II+
Software	BD FACSDiva, FloJo
Cell population abundance	Abundance was between 0.5% and 8.5%, depending on sample. Purity was evaluated by fluorescence imaging of sorted cells (Extended Data Fig 1), and by single-cell sequencing of sorted populations (Supplementary Fig 1).
Gating strategy	DAPI-negative, Vybrant DyeCycle Ruby-positive events were gated on tdTomato fluorescence vs GFP-channel autofluorescence as shown in Extended Data Fig 1b.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.