

# Continuous cell-type diversification in mouse visual cortex development

<https://doi.org/10.1038/s41586-025-09644-1>

Received: 2 September 2024

Accepted: 17 September 2025

Published online: 5 November 2025

Open access

 Check for updates

The mammalian cortex is composed of a highly diverse set of cell types and develops through a series of temporally regulated events<sup>1–3</sup>. Single-cell transcriptomics enables a systematic study of cell types across the entire timeline of cortical development. Here we present a comprehensive and high-resolution transcriptomic and epigenomic cell-type atlas of the developing mouse visual cortex. The atlas is built from a single-cell RNA sequencing dataset of 568,654 high-quality single-cell transcriptomes and a single-nucleus Multiome dataset of 200,061 high-quality nuclei, which were densely sampled across the embryonic and postnatal developmental stages (from embryonic day 11.5 to postnatal day 56). We computationally reconstructed a transcriptomic developmental trajectory map of all excitatory, inhibitory and non-neuronal cell types in the visual cortex. Branching points that mark the emergence of new cell types at specific developmental ages and molecular signatures of cellular diversification are identified. The trajectory map shows that neurogenesis, gliogenesis and early postmitotic maturation in the embryonic stage give rise to all cell classes and nearly all subclasses in a staggered parallel manner. Increasingly refined cell types emerge throughout the postnatal differentiation process, including the late emergence of many cell types during the eye-opening stage and the onset of critical period, suggesting that there is continuous cell-type diversification at different stages of cortical development. Throughout development, there are cooperative dynamic changes in gene expression and chromatin accessibility in specific cell types. We identify cell-type-specific and temporally resolved gene regulatory networks that link transcription factors and downstream target genes through accessible chromatin motifs. Collectively, our study provides a detailed dynamic molecular map directly associated with individual cell types and specific temporal events that can reveal the molecular logic underlying the complex and multifaceted cortical cell type and circuit development.

The cerebral cortex of the mammalian brain controls a wide range of flexible and motivated behaviours and is extensively expanded in species with more advanced cognitive functions (including humans). This brain region has been a prime site for the study of the diverse cell types it contains and how they form functionally specific neural circuits<sup>4–6</sup>. Cell types in the cortex can be defined on the basis of multiple cellular properties, including gene expression, morphology, physiology, connectivity or various combinations thereof<sup>6–9</sup>. Over the past decade, single-cell transcriptomics has provided comprehensive and detailed cell-type classifications that define around 100 transcriptomic cell types (T-types) in each cortical area of the adult brain, and these are markedly consistent across areas and across species (for example, from mouse to human)<sup>10–13</sup>. These T-types can be hierarchically organized into classes and subclasses that reflect their varied relatedness and are likely to be rooted in the evolutionary and developmental histories of the cell types<sup>9</sup>. Specifically, in each cortical area, about 28 cell subclasses have been defined: 9 glutamatergic excitatory neuronal subclasses, 8 GABAergic inhibitory neuronal subclasses, 3 glial subclasses, 3 immune subclasses and 5 vascular subclasses. The glutamatergic subclasses

are organized on the basis of their layer (L) specificity and long-range projections: L2/3 intratelencephalic projecting (IT), L4/5 IT, L5 IT, L6 IT, L6 Car3, L5 extratelencephalic projecting (ET), L5/6 near-projecting (NP), L6 corticothalamic projecting (CT) and L6b. The GABAergic subclasses—Lamp5, Sncg, Vip, Pvalb, Pvalb chandelier, Sst, Sst Chodl and Lamp5 Lhx6—are organized on the basis of their developmental origins. The glial subclasses comprise astrocytes, oligodendrocytes and oligodendrocyte precursor cells (OPCs), whereas the immune subclasses comprise microglia, border-associated macrophages (BAMs) and lymphoid cells. Finally, the vascular subclasses include vascular leptomeningeal cells (VLMCs), arachnoid barrier cells (ABCs), endothelial cells, pericytes and smooth muscle cells (SMCs)<sup>13,14</sup>.

Multimodal integrative approaches have been used to align the different levels of transcriptomic cell types to morphology, physiology and connectivity and, in some cases, to refine cell-type definition<sup>13,15,16</sup>. For example, using Patch-seq, GABAergic neurons in the mouse visual cortex are classified into 28 morphoelectric-transcriptomic types (MET-types), which represent a coarser resolution from the original 61 T-types but with increased cross-modality concordance in each

MET-type<sup>15</sup>. Computational matching of local dendritic and axonal morphology has enabled the assignment of T-type identities to reconstructed neurons in the mouse visual cortex that have long-range projection patterns or synaptic connectivity profiles derived from light or electron microscopy data<sup>17,18</sup>. Cell-type-targeting genetic tools, bar-coded viruses and spatial transcriptomic approaches have also been used to relate transcriptomic identities to connectivity or functional properties<sup>19–21</sup>.

Development of the mammalian cortex has been extensively studied over the years<sup>3,22–25</sup>. It is now known that glutamatergic neurons, astrocytes and oligodendrocytes are generated in the dorsal pallium (which subsequently becomes the cortex), whereas GABAergic neurons are generated in the subpallium and undergo long-distance migration into the cortex following specific routes<sup>26,27</sup>. Immune and vascular cell types originate outside the brain. In both the pallium and the subpallium, progenitors in the ventricular zone (VZ) and the subventricular zone (SVZ) progressively give rise to radial glia (RG), intermediate progenitors (IPs) and immature neurons (IMNs). In the developing cortex, glutamatergic neurons in different layers are thought to be generated sequentially and migrate radially to reach their target layers in an inside-out manner<sup>28,29</sup>. After neurogenesis, RG switch to gliogenesis and generate astrocytes, OPCs and oligodendrocytes (although some oligodendrocytes also come from the subpallium). Postmitotically, all cell types undergo specific maturation processes. Glutamatergic and GABAergic neurons go through dendritic and axonal arborization, synapse formation and activity-dependent circuit refinement. In particular, the visual cortex goes through experience-independent and experience-dependent circuit development to acquire increasingly refined visual response properties<sup>30</sup>.

There are substantial gaps in our understanding of the developmental processes and mechanisms involved in the formation of the mammalian cortex. It is still unclear when specific cell-type identities are established, to what extent cell types observed in the adult cortex are established during the embryonic stage and how lineage-bifurcation decisions occur. In the postnatal developmental period, many processes are in play with overlapping time courses. Such processes include intrinsic neuronal activities, influence of external sensory inputs, incoming and outgoing long-range connections, formation of local excitatory and inhibitory circuit motifs, and neuronal and non-neuronal cell–cell interactions. Consequently, cells are undergoing rapid state transitions. Despite the discovery of many genes, proteins and epigenetic signatures involved in these processes, we have little systematic knowledge about the cell-type-specific events and their dynamics, how cell-type-specific circuits are formed and what mechanisms drive cell-type and circuit maturation. To address these questions, it is important to investigate developmental changes at the single-cell level and to link these changes across time with cell-type specificity.

Here we report a comprehensive transcriptomic and epigenomic cell-type atlas of the developing mouse visual cortex with high temporal resolution from embryonic to postnatal development. We systematically identify the precise timing of the onset of all excitatory, inhibitory and non-neuronal cell subclasses and clusters in the visual cortex and demonstrate that there is a pattern of continuous cell-type diversification. We also systematically categorize large numbers of differentially expressed (DE) gene sets and differentially accessible (DA) chromatin peak modules that are concurrently associated with specific cell types and developmental ages. Together, these data provide a real-time dynamic molecular map associated with individual cell types and specific developmental events that will facilitate future investigations of the mechanisms of cell-type and circuit development.

### Mouse visual cortex developmental cell-type atlas

We generated two datasets of the developing mouse visual cortex using single-cell RNA sequencing (scRNA-seq) and single-nucleus Multiome

(snMultiome, a combination of single-nucleus RNA-seq (snRNA-seq) and single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq)). We used the scRNA-seq data to generate a transcriptomic cell-type atlas and developmental trajectory map. We then used the snMultiome data to reconstruct an epigenetic chromatin-accessibility landscape across development (described below).

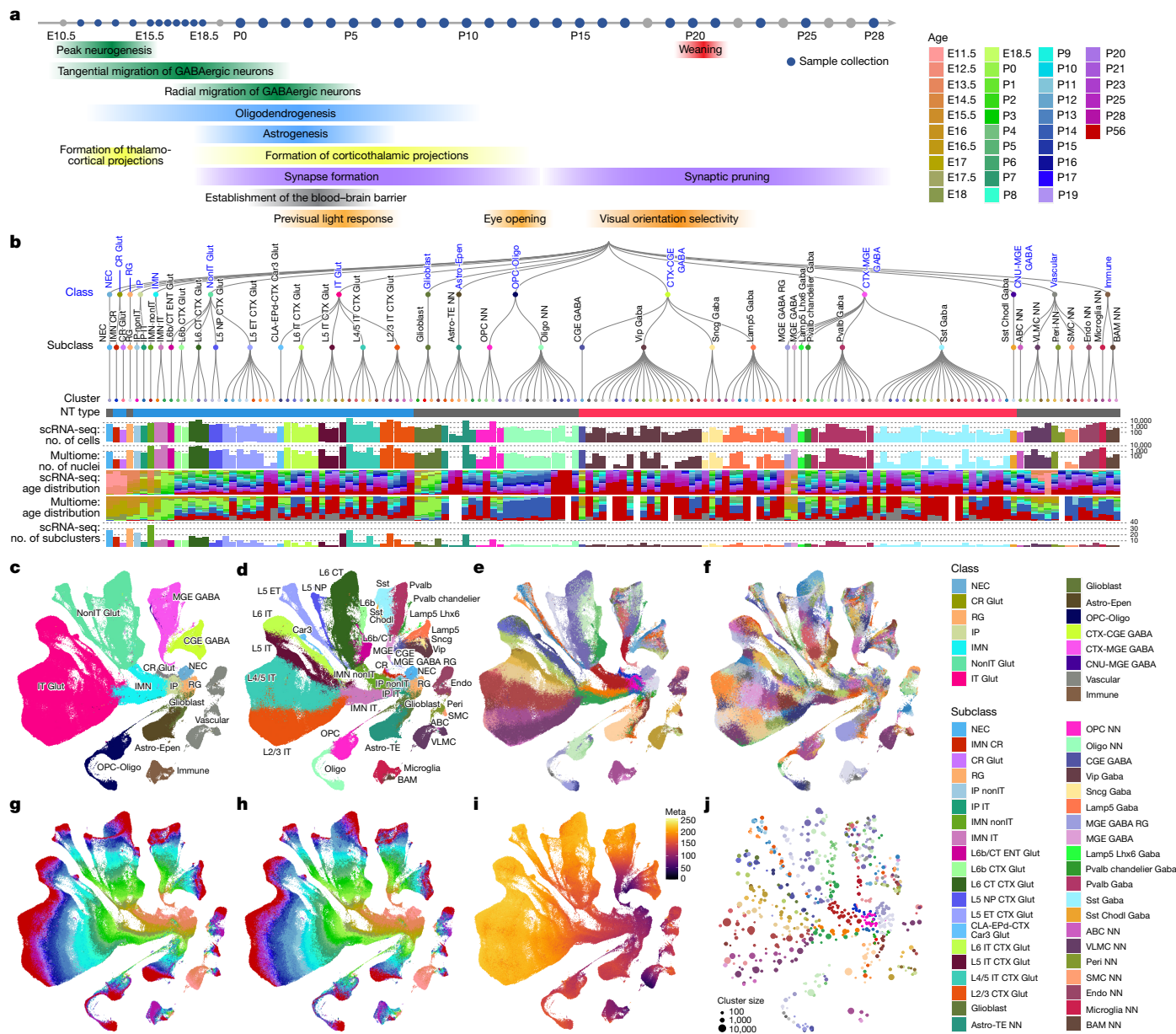
We first generated 91 scRNA-seq libraries using 10x Genomics Chromium v3 (10xv3), which resulted in a dataset of 913,297 single-cell transcriptomes (Supplementary Table 1). The scRNA-seq dataset densely covered the embryonic and postnatal periods over 34 time points from embryonic day 11.5 (E11.5) to postnatal day 28 (P28) and adult stage P56 (Fig. 1a). We established stringent quality control (QC) metrics (Methods and Supplementary Table 2), similar to our previous studies<sup>14</sup>, to remove low-quality single-cell transcriptomes. To overcome natural variation over fixed collection times, we assigned a predicted ‘synchronized age’ to each cell to obtain more homogeneous temporal transcriptomic profiles for some analyses (Methods and Extended Data Fig. 2a,d).

To build the developmental trajectory of the adult cell types, we first conducted label transfer using our recently established adult mouse whole brain taxonomy<sup>14</sup>. The Allen Brain Cell–Whole Mouse Brain (ABC–WMB) Atlas served as the reference for cells at the adult stage to assign cell-type identities at the cluster level. Adult cell-type identities were then propagated to younger cells through sequential cell-type label transfer from older to younger synchronized ages for all postnatal ages (Methods and Extended Data Figs. 2a and 3). Overall, to the P20–28 age bin, we transferred labels from 35 out of the 35 P56 glutamatergic clusters, 60 out of the 61 P56 GABAergic clusters and 16 out of the 20 P56 glial clusters derived from the adult ABC–WMB Atlas to capture nearly all the cell-type diversity in the adult mouse visual cortex.

For the embryonic time points, we mapped cells in prenatal-enriched global clusters to a developmental mouse brain scRNA-seq reference from a previous study<sup>31</sup> to identify broad cell types (Methods). Clusters mapped to RG in that study<sup>31</sup> were further classified as neuroepithelial cells (NECs) or RG, with RG arising from NECs. Previous studies have suggested that a transcriptomic continuum exists for the gradual transition from IPs to IMNs, including migrating neurons, to mature cortical excitatory neurons<sup>32</sup>. Thus, we used a combinatorial set of marker genes to assign clusters to these categories (Methods). We also defined preplate Cajal–Retzius (CR) cells and glioblasts.

After iterative de novo clustering and merging, we conducted further annotation and identified and removed an additional set of ‘noise’ subclusters that had escaped the initial QC process or subclusters that probably originated from outside the cortex. This step resulted in a final set of 568,654 high-quality single-cell transcriptomes that form 714 subclusters (Extended Data Fig. 1a). As part of the annotation, we integrated early developmental ages of our data with three external datasets<sup>31–33</sup> using scVI<sup>34</sup> (Extended Data Fig. 4). Overall, our cell-type assignment was broadly consistent with those from the previous studies at the subclass level while providing finer cell-type and temporal resolutions with additional cluster and subcluster annotations.

We present these complex molecular relationships through a high-resolution transcriptomic cell-type taxonomy for the adult and developing mouse visual cortex, visualized in a dendrogram and a uniform manifold approximation and projection (UMAP) plot (Fig. 1b–j). The taxonomy comprises four nested levels of classification: 15 classes, 40 subclasses, 148 clusters and 714 subclusters (full cluster names are provided in Supplementary Table 3). It includes all known neuronal and non-neuronal cell classes of the developing neocortex from the literature<sup>3</sup> and many transitional cell types and subtypes discovered here. We also generated a list of 6,724 DE genes that differentiate among all clusters and subclusters (Supplementary Table 4).



**Fig. 1 | Transcriptomic developmental cell-type atlas of the mouse visual cortex.** **a**, Schematic timeline of samples collected from scRNA-seq and Multiome in this study along with major developmental events of the isocortex. **b**, The transcriptomic taxonomy tree of 148 clusters organized in a dendrogram (scRNA-seq,  $n = 568,654$  cells; Multiome,  $n = 200,061$  nuclei). The classes and subclasses are marked on the taxonomy tree. Full cluster names are provided in Supplementary Table 3. Bar plots represent (top to bottom): major

neurotransmitter (NT) type, number of scRNA-seq cells, number of Multiome nuclei, age distribution of scRNA-seq cells, age distribution of Multiome nuclei and number of scRNA-seq subclusters for each cluster. **c–i**, UMAP representations of all cell types coloured by class (**c**), subclass (**d**), cluster (**e**), subcluster (**f**), age (**g**), synchronized age (**h**) and pseudotime (**i**). **j**, Constellation plot showing the UMAP centroids of subcluster nodes coloured by cluster.

Of the 148 clusters (Supplementary Table 3), 132 clusters (containing 517 subclusters) aligned with the adult ABC–WMB Atlas<sup>14</sup>, which represent maturing cell types. These clusters belong to 27 of the above-mentioned 28 canonical cortical cell subclasses<sup>13,14</sup> (without lymphoid cells), and 1 destined to the entorhinal cortex (see below), under a total of 9 classes. We used the labels of these 28 subclasses and 132 clusters from the ABC–WMB Atlas while modifying some of their class labels to be more consistent with the embryonic classes. The remaining 16 clusters (containing 197 subclusters) represent progenitor cells and IMNs in embryonic and perinatal stages and belong to 12 subclasses under 8 classes.

Neuronal cell types and their progenitors constitute a large proportion of the developmental atlas and represent 10 classes: NECs, CR Glut,

RG, IPs, IMNs, nonIT Glut, IT Glut, CTX-CGE GABA, CTX-MGE GABA and CNU-MGE GABA (Fig. 1b,c). The 10 classes are further divided into 29 subclasses, 109 clusters and 599 subclusters. The nonIT Glut class consists of four main glutamatergic subclasses—L5 ET, L5 NP, L6 CT and L6b—and a L6b/CT ENT subclass that is mostly present at E17–P3 and belongs to the entorhinal cortex on the basis of our mapping result (Fig. 1b,d). The IT Glut class contains four main subclasses—L2/3 IT, L4/5 IT, L5 IT and L6 IT—and a CLA-EPd-CTX Car3 subclass that consists of a distinct L6 cell type shared with the claustrum and the endopiriform nucleus<sup>12,19</sup>.

Cortical GABAergic neurons are born in three subpallial progenitor zones: the caudal ganglionic eminence (CGE), the medial ganglionic eminence (MGE) and the preoptic area (POA)<sup>24</sup>. Neural progenitor

cells in these regions<sup>35</sup> generate IMNs, and the IMNs migrate to the cortex, where the IMNs mature<sup>36,37</sup>. Our data showed that MGE GABA IMNs differentiate into four subclasses—Sst Gaba, Pvalb Gaba, Pvalb chandelier Gaba and Lamp5 Lhx6 Gaba—in the CTX-MGE class and one subclass, Sst Chodl Gaba, in the CNU-MGE class<sup>14,38</sup> (Fig. 1b,d). The CGE GABA IMNs gradually differentiate into Vip Gaba, Sncg Gaba and Lamp5 Gaba subclasses.

All non-neuronal cell types are classified into 5 classes—glioblast, OPC-Oligo, Astro-Epen, immune and vascular—that are further divided into 11 subclasses (Fig. 1b–d). Glioblasts are one main type of progenitor for the OPC-Oligo and Astro-Epen classes<sup>39</sup>. The OPC-Oligo class contains two subclasses: OPCs (expressing *Olig1*, *Olig2* and *Pdgfra*) and oligodendrocytes (Oligo; *St18* and *Opalin*). The Astro-Epen class contains one subclass of telencephalic astrocytes: Astro-TE (*ApoE*, *Aqp4*, *Aldh1l1* and *Slc1a3*). The immune class consists of two subclasses: microglia (*Siglech*, *Sall1* and *Ifitm10*) and BAMs (*F13a1*, *Pf4* and *Mrc1*). The vascular class consists of five subclasses: ABCs (*Slc47a1*), VLMCs (*Col1a1*, *Col1a2*, *Apod* and *Slc6a13*), pericytes (*Kcnj8*), SMCs (*Acta2* and *Myh11*) and endothelial cells (Endo; *Ly6c1* and *Slco1a4*).

### Building cell-type development trajectories

Trajectory analysis is an essential tool for modelling the dynamic process of cellular development and differentiation. Given the cell-type identities at the adult stage and the dense temporal sampling, we were able to progressively propagate cell-type identities between two adjacent ages (see above). Thus, we defined edge weights of the trajectory tree based on *k*-nearest neighbours (*k*-NN) in the integrated space across synchronized ages for the postnatal trajectory, whereas the *k*-NN approach with Monocle3-based<sup>40</sup> pseudotime was used for the embryonic trajectory (Methods and Extended Data Fig. 2a–c). To demonstrate the robustness of our taxonomy and trajectory map, we used scVI to integrate data between adjacent age bins. The results using this method closely matched those from Seurat (Methods and Supplementary Fig. 1).

Overall, we retained all edges between a cluster and its potential antecedents that have edge weights of >0.2 (Supplementary Table 5). To simplify visualization and conceptualization of the developmental process, we chose the edge with the maximal weight between a cluster and one antecedent to build the developmental trajectory map across the entire timeline from E11.5 to P56 (Figs. 2a and 3). Of note, the total 987 chosen edges with maximal weights to build all the trajectories have an average weight of 0.71 (and more than 85% with weights of >0.5), whereas the 321 non-chosen edges have weights of <0.5 with an average of 0.29 (Supplementary Table 5). This finding indicates that there is a relatively unambiguous trajectory pattern. We then computed the global pseudotime based on the entire developmental trajectory map (Figs. 1i and 3a–d and Methods).

We constructed a branched trajectory tree for the neuronal and glial subclasses in the visual cortex (Fig. 2a). The tree was largely consistent with previous studies<sup>41,42</sup> and supported by the progression of relative proportions of different subclasses with time (Fig. 2b) and by key marker genes for each branching node (Fig. 2c and Extended Data Fig. 5), with a UMAP focused on early developmental cell types (Fig. 2c). The trajectory tree revealed that the earliest cell type emerging from NECs is IMN CR, which probably arises from NECs in the cortical hem<sup>43</sup> before E11.5 and gradually matures into CR cells. Then RG molecular identity emerges at E13, followed immediately by the emergence of IP nonIT and IMN nonIT molecular identities at E13.5. Molecular identities of IP IT cells and glioblasts, as well as IMN IT deep-layer cells, appear around E15.5. IP nonIT cells give rise to more IMN nonIT cells. IMN nonIT cells turn into three subclasses of nonIT neurons (L5 ET, L6 CT and L6b) at E17, whereas the fourth subclass (L5 NP) emerges at E18.5. IP IT cells give rise to more IMN IT cells. IMN IT deep-layer cells turn into L6 IT and L5 IT neurons at E17, and IMN IT upper-layer cells turn into

L4/5 IT and L2/3 IT neurons at E18.5. Meanwhile, glioblasts give rise to astrocytes and OPCs around E17. Separately, for GABAergic neuron classes, MGE RG and MGE IMNs appear before E11.5, and MGE IMNs differentiate into Sst and Pvalb neurons at E14.5. CGE IMNs appear in the cortex later around E15.5, and they differentiate into Vip, Sncg and Lamp5 neurons at E18–P2.

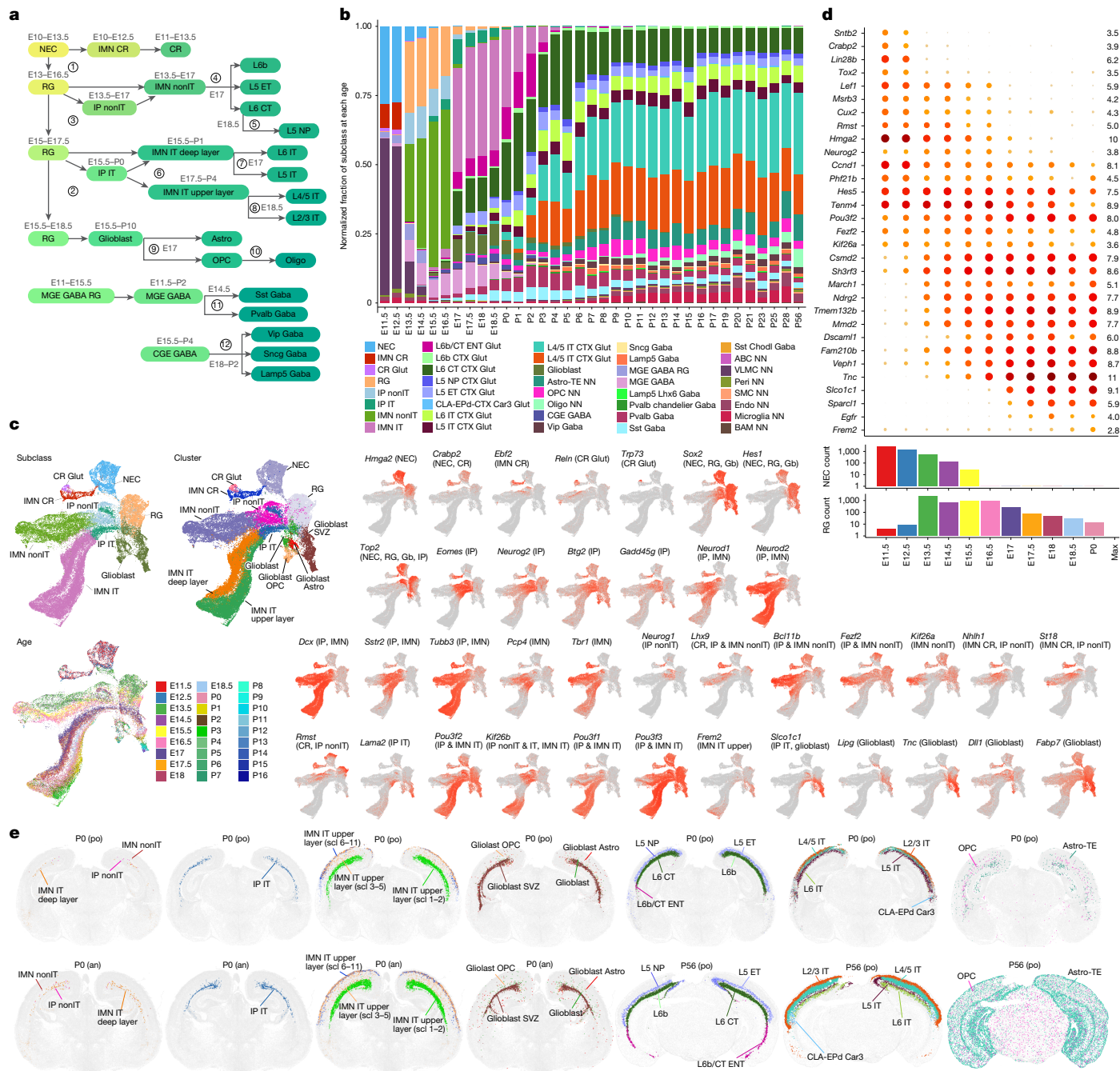
It should be noted that in our data, the identity of a cell is defined by its real-time transcriptional profile (molecular identity) and not by its birth date. At the earliest stage (E11.5–E12.5), cells originating from the pallium are mainly composed of NECs (expressing *Hmga2*), IMN CR cells and the early-born CR cells (expressing *Ebf1*, *Ebf2*, *Ebf3*, *Reln*, *Calb2*, *Crabp2* and *Trp73*) (branching node 1; Fig. 2a–c and Extended Data Fig. 5). The IMN CR cells are antecedents of CR cells, and the expression of *Eomes*, *Neurog2*, *Neurod1* and *Neurod2* decreases as CR cells mature.

Beginning at E13, RG molecular identity (*Sox2*, *Pax6*, *Hes1* and *Hes5*) emerges, and RG simultaneously give rise to IPs (*Eomes*, *Btg2*, *Neurog2* and *Gadd45g*) and IMNs (*Dcx*, *Neurod1*, *Neurod2*, *Neurod6*, *Tubb3* and *Tbr1*), consistent with the co-existence of direct and indirect neurogenesis<sup>42</sup>. Most IPs generated between E13.5 and E16.5 are IP nonIT cells and transition into IMN nonIT cells, whereas most IPs between E17 and P0 are IP IT cells and transition into IMN IT cells (Fig. 2b). IMN IT deep-layer cells are observed at later times than IMN nonIT cells (E15.5–P1 compared with E13.5–E17; Fig. 2a–c), even though they colocalize in deep layers.

We observed clear molecular signatures that distinguish nonIT and IT trajectories at the IP and IMN stages. The nonIT IPs and IMNs express *Neurog1*, *Lhx9*, *Fezf2*, *St18*, *Rmst*, *Nhlh1*, *Nhlh2* and *Kif26a*, whereas the IT IPs and IMNs express *Pou3f1*, *Pou3f2*, *Pou3f3*, *Kif26b*, *Lama2* and *Slco1c1* (node 3; Fig. 2a,c and Extended Data Fig. 5). The canonical deep-layer neuron markers *Fezf2*, *Bcl11b*, *Foxp2* and *Tle4* all have selective expression in nonIT cells but exhibit varying temporal dynamics, with *Bcl11b* and *Fezf2* emerging in IP, whereas *Foxp2* and *Tle4* appear at the late IMN stage.

The transcriptomic difference between IT and nonIT trajectories is not only present in IPs and IMNs but already in RG at different ages, with the nonIT cell marker *Rmst* present in early RG and the IP marker *Pou3f2* in later RG (Fig. 2d). Recent studies have suggested that the transcriptional profile of cortical RG changes as they generate nonIT neurons, IT neurons and glial cells<sup>33,41</sup>. In our data, the divergence of progenitors for glutamatergic neurons (nonIT Glut and IT Glut) and glia (OPC-Oligo and Astro) may start as early as E15.5 (node 2; Fig. 2a–c), and the RG subclass shows a continuum of cells among different transcriptomic states (Fig. 2d). First, earlier-stage RG are enriched for *Neurog2* and *Tenm4* (refs. 44,45), which may represent a committed neurogenic state, whereas expression of *Tnc* is seen in later-stage RG, which may represent a committed gliogenic state. Second, glioblasts emerge at E15.5 and express higher levels of *Fabp7*, *Lipg*, *Slco1c1*, *Tnc*, *Qk* and *Slc1a3* than RG, which indicate their transition towards the glial cell trajectory. Our data suggest that RG already show complex temporal gene expression changes, and they exit the RG states at different ages with these temporal signatures to become IPs and IMNs or glioblasts that are committed to distinct neuronal (nonIT or IT) or glial trajectories. These results are consistent with and may explain the observed heterogeneity in previous lineage tracing and transcriptomic profiling studies<sup>33,46–48</sup>.

We used a MERFISH dataset we recently generated that covers the entire mouse brain at P0 to identify the relative abundance and spatial location of developmental clusters at P0, a critical transitioning time point (Fig. 2e and Methods). The P0 MERFISH data revealed the spatial organization of both embryonic cell types and the emerging subclasses that persist into adulthood. At this time, IP nonIT, IMN nonIT and IMN IT deep-layer cells are scarce, whereas there is still a prominent IP IT population localized in the SVZ and a large number of IMN IT upper-layer cells spread across the cortical depth. These results are consistent with our scRNA-seq findings (Fig. 2a–c). Notably,



**Fig. 2 | Developmental trajectories of visual cortex cell subclasses.**

**a**, Transcriptomic trajectories of visual cortex cell subclasses with estimated timings of onset and major branching nodes. **b**, Relative proportions of cells corresponding to the different subclasses at each age. Note that relative proportions between neuronal and non-neuronal cells do not reflect the actual situation owing to the variable FACS methods used for different scRNA-seq libraries (Methods, Extended Data Fig. 1d and Supplementary Table 1). **c**, UMAP representations of early developmental cell types coloured by subclass,

cluster, age and expression of key marker genes separating different trajectories. **d**, Dot plot showing the expression of DE genes across embryonic ages and P0 in NEC and RG populations. The numbers of NECs and RG at each age point are shown at the bottom. **e**, Representative MERFISH sections at P0 and P56 with specific cell types labelled. an, anterior section that includes the somatosensory cortex; po, posterior section that includes the visual cortex; scl, subclass.

subclusters in the IMN IT upper-layer cluster can be placed into three groups with distinct layer distribution patterns indicative of continued radial migration. Subclusters 1–2 are concentrated in or near the SVZ, whereas subclusters 6–11 are located near the surface of the cortex. Subclusters 3–5 are distributed across the cortical depth between the other two groups. These spatial patterns correspond to the relative locations of the subclusters in UMAP as they progress from more immature to more mature states.

### Developmental trajectories of glutamatergic types

Our analysis indicated that the postmitotic IMNs (IMN nonIT and IMN IT) progressively diversify into more distinct cell subclasses and types (Figs. 1c–f, 2a and 3a,e). In the nonIT trajectory, the IMN molecular identity (*Fezf2*, *Bcl11b* and *Neurod2*) emerges at E13.5, with increasing expression of *Foxp2*, *Tle4* and *Crym* at the late IMN stage. This trajectory splits around E17 into L6 CT, L5 ET and L6b (node 4; Fig. 2a–c and



IT subclasses and *Cux1* and *Cux2* in L2/3 IT and L4/5 IT subclasses. The IMN IT deep-layer cluster differentiates into L5 IT (*Fezf2*) and L6 IT (*Fosl2*) around E17 (node 7, Fig. 2a–c and Extended Data Fig. 5). *Nfia* and *Sox5*, which show strong enrichment in the nonIT trajectory, are also enriched in L6 IT. In the upper-layer IT population, L2/3 IT (*Mdag1* and *Klhl1*) and L4/5 IT (*Rorb*, *Rora* and *Tox*) separate around E18.5 (node 8; Fig. 2a–c and Extended Data Fig. 5).

In the P0 MERFISH data, we observed a divergence in laminar distribution between IT and nonIT subclasses at this age (Fig. 2e). The IT subclasses present a layered profile analogous to that seen in adult brains, but with less clear segregation. By contrast, the nonIT subclasses display a distinct separation into layers. Notably, L5 NP neurons are situated at a deeper cortical depth than L5 ET neurons at P0, whereas they become more intermingled at P56.

In each glutamatergic subclass, cells continue to differentiate and diversify, giving rise to new cell clusters. We derived a cluster trajectory tree of all cell types and conducted DE gene analysis at each branching point (Fig. 3a,e and Extended Data Figs. 6 and 7). In the nonIT class, most of the L5 ET, L5 NP, L6 CT and L6b clusters begin to diverge by P3, except for the L5 ET clusters 371–373 (*Chrna6*), which represent the most distinct subset<sup>10,12,18</sup> and diverge at the onset of critical period (Fig. 3a,e and Extended Data Fig. 6).

In the IT trajectory, many clusters that split off early have a distinct layer distribution (Fig. 3a,e and Extended Data Fig. 7), and many genes with a distinct layer distribution show a specific expression pattern at early stages of IT cell-type divergence. This finding indicates that the cortex has more refined sublayer gradients that are specified by early postnatal age. More clusters arise in later stages of development after eye opening, and these newer clusters have less distinct spatial distributions from sibling clusters. For example, L2/3 IT cluster 109 diverges from cluster 110 at around P11, with increased expression of *Bdnf* and decreased expression of *Adamts2*. By contrast, cluster 118 further diverges from cluster 109 at P21, with increased expression of *Baz1a* and *Tnfaip6*. Spatially in L2/3, clusters 118 and 110 are located more superficially than cluster 109 at P56. Recent studies have shown functional and developmental distinctions of these L2/3 IT clusters in the somatosensory cortex<sup>50</sup> and the visual cortex<sup>51</sup>. We also observed late divergence of L4/5 IT, L5 IT and L6 IT clusters. In L4/5 IT, cluster 100 is the dominant cluster and is L4-specific, whereas clusters 101 and 82 diverge from cluster 100 at P14 and P20, respectively.

Overall, most nonIT clusters already exist before eye opening, except for a few *Chrna6*<sup>+</sup> L5 ET clusters. By contrast, IT clusters continue to emerge from P11, around the time of eye opening, to as late as P21, at the onset of critical period (Fig. 3a,e). This result suggests that IT cells become molecularly distinct at the embryonic stage and continue to diversify throughout the postnatal period.

## Developmental trajectories of glial cell types

RG transition into gliogenesis starting at E15.5, as indicated by the increasing expression of *Tnc* (node 2; Fig. 2a–d). During this time, *Slco1c1* and *Sparcl1* are turned on in RG at E17, with further activation in glioblasts. The glioblast molecular identity emerges at E15.5 and contains two initial clusters: glioblast and a special population we refer to as glioblast SVZ (Fig. 3b,f).

The glioblast SVZ cluster shares expression of *Vephl1*, *Tspan18*, *Tfap2c* and *Adamts18* with RG and shares expression of *Slco1c1* and *Tnc* with astrocytes (Extended Data Fig. 8). *Gja1* is turned on in this population at around P0, and *Thbs4* is turned on at about P9. Our P0 MERFISH data showed that this cluster is located in the SVZ at P0, deeper than other glioblast populations (Fig. 2e).

The glioblast cluster is labelled by both the oligodendrocyte markers *Olig1* and *Olig2* and the astrocyte markers *Tnc*, *Slco1c1* and *Egfr*, and gives rise to both astrocytes and OPCs (node 9; Figs. 2a,c and 3b,f and

Extended Data Figs. 5 and 8). This cluster rapidly splits into clusters glioblast Astro and glioblast OPC starting at E17, with enrichment of *Slco1c1*, *Aldoc*, *Id3* and *Pax6* in the astrocyte trajectory and enrichment of *Dll1*, *Dll3*, *Ascl1* and *ErbB4* in the OPC and oligodendrocyte trajectory. The OPC clusters 5266 and 5271 show strong expression of the cell cycle genes *Mki67* and *Top2a*, which indicates that these cells are still rapidly proliferating; these clusters are reduced after eye opening. Telencephalon spatial-patterning transcription factor (TF) genes *Foxg1* and *Lhx2* are strongly expressed in RG, downregulated but maintained in astrocytes and diminished in OPCs and oligodendrocytes. This pattern may be related to the fact that spatial identity is lost in oligodendrocytes but maintained in astrocytes<sup>14,31</sup>.

During postnatal development, OPCs are predominant, but after P11, their proportion gradually decreases (Extended Data Fig. 8h). Committed oligodendrocyte precursors (COPs) start to appear around P2, marked by the downregulation of *Creb5*, *Etv5*, *Etv4*, *Etv9* and *Pdgfra* and the upregulation of *St18*, *Bmp4*, *Enpp6* and *Plp1* (node 10, Figs. 2a,c and 3b,f and Extended Data Figs. 5 and 8). Newly formed oligodendrocytes (NFOLs) emerge around P11, whereas mature oligodendrocytes (MOLs) appear around P12, and their proportion increases until adulthood. These results are consistent with previous studies showing that neuronal activity influences OPC and oligodendrocyte proliferation, differentiation and myelin remodelling<sup>52</sup>.

Finally, we identified five astrocyte clusters in this dataset (Fig. 3b,f and Extended Data Fig. 8). Cluster 5225 is the most dominant astrocyte cell type in the visual cortex. Clusters 5218 and 5219, with enriched expression of *Gfap*, *Myoc* and *Atoh8*, are interlaminar astrocytes (ILAs) or glia limitans superficialis (GLS) astrocytes<sup>53</sup> localized at the pia of the cortex<sup>14</sup>. Cluster 5228 is a rare astrocyte cell type marked by *Thbs4* and located in the white matter of the lateral cortex and the cortical subplate (CTXsp). Trajectory analysis indicated that cluster 5228 probably originates from the glioblast SVZ cluster, a result supported by shared spatial localization and marker gene expression (Fig. 2e and Extended Data Fig. 8f,g).

## Developmental trajectories of GABAergic types

The earliest GABAergic cell populations at E11.5 express TF genes *Dlx1*, *Dlx2*, *Ascl1* and *Gsx2*, which are required for specification of all GABAergic neurons in the subpallium<sup>24,54,55</sup>. These cells migrate from the ganglionic eminence along tangential paths, with some cell types being pre-specified before reaching the cortex. As our postnatal data collection only includes GABAergic cells in the cortex, it is possible that portions of early trajectory paths outside the cortex were not captured by our analysis. Nonetheless, we still inferred multiple trajectory paths with good confidence (Supplementary Table 5).

At E11.5, we observed the initial emergence of MGE GABAergic progenitors, the MGE GABA RG, which progress to MGE GABA IMNs at E14.5 (Fig. 2a,b). Although previous studies have suggested that Sst and Pvalb cells may originate from different domains in the MGE<sup>56</sup>, we did not see segregation of these two populations in the MGE RG and MGE subclasses, although subtle differential gene signatures might exist. Starting at E14.5, MGE cells differentiate into two subclasses: Sst Gaba (*Shisa6*, *Pou3f3*, *Npas1* and *Tox*) and Pvalb Gaba (*Adamts17*, *Shisa9*, *Tafa2* and *Zfp804b*) (node 11; Fig. 2a and Extended Data Figs. 5 and 9). We identified three additional highly distinct MGE subclasses: Sst Chodl and Pvalb chandelier emerging around P1, and Lamp5 Lhx6 emerging around P5. These subclasses probably have diverged from other MGE cell types before reaching the cortex.

In the Pvalb and Sst subclasses, we identified five primary developmental trajectories for each (Fig. 3c,g and Extended Data Fig. 9). Grouping of these GABAergic cell types by trajectories matches the definition of MET-types previously categorized in mouse visual cortex using Patch-seq<sup>15</sup>, as shown in extended data fig. 5 of our recent study<sup>38</sup>. Clusters in each trajectory group often split at late postnatal

ages, especially during eye opening (at around P11) or critical period (at about P21), which indicates continued diversification.

In the Pvalb subclass (Fig. 3c,g and Extended Data Fig. 9), out of the five Pvalb MET-types<sup>15</sup>, four are fast-spiking basket cells (or cells with related morphologies) located in different layers, and one (Pvalb-MET-5) is the chandelier cell type (corresponding to the Pvalb chandelier cluster 733). Developmental group 1 (*Gpr149*) and group 2 (*Reln* and *Pdlim3*) both correspond to the Pvalb-MET-3 type (in L5). Group 3 (*Tpbp* and *Calb1*) contains clusters 742 and 752, with cluster 742 corresponding to Pvalb-MET-4 (in L2/3) and cluster 752 diverging from cluster 742 at P17 and corresponding to Sst-MET-2. Group 4 (*Sema3e*, *St6galnac5* and *Ptprk*) corresponds to Pvalb-MET-2 (in L6). The *Th*<sup>+</sup> Pvalb cluster 735 corresponds to Pvalb-MET-1 (in L6). However, the developmental trajectory of cluster 735 seems highly ambiguous, with its closest antecedent being Sst cluster 758. This result is consistent with our previous finding that the L6 *Th*<sup>+</sup> Pvalb cells may be a transition type between Pvalb and Sst subclasses<sup>12</sup>.

In the Sst subclass (Fig. 3c,g and Extended Data Fig. 9), there are 13 Sst MET-types<sup>15</sup>. Besides the Sst Chodl subclass (cluster 859, corresponding to Sst-MET-1, long-range projecting neurons), the five developmental groups also correspond to specific MET-types<sup>38</sup>. All Sst clusters exhibit highly restrictive layer distribution. In group 1 (*Crh*, *Crhr2*, *St6galnac5* and *Ptprk*), clusters 757, 758 and 761 correspond to Sst-MET-9-10, and clusters 811, 814, 818, 819 and 820 correspond to Sst-MET-12-13. Sst-MET-9-13 types are all L5/6 non-Martinotti cells. In group 2 (*Cbln4*, *Calb2* and *Tox*), clusters 795 and 797 correspond to Sst-MET-2 (L2/3 fast-spiking basket cell (BC)-like cells), and clusters 803 and 806 correspond to Sst-MET-3-5 types (L2/3 and L5 fanning Martinotti cells). In group 3 (*Hpse*), clusters 792 and 793 correspond to Sst-MET-8 (L4-targeting Martinotti cells), and clusters 799, 800 and 801 correspond to Sst-MET-7 (L5-T-shaped Martinotti cells). In groups 4 and 5, clusters 779, 777 and 780 correspond to Sst-MET-6 (also L5 T-shaped Martinotti cells). Groups 4 and 5 display similar transcriptomic profiles, sharing markers *Nr2f2* and *Myh8*, with *Pdyn* enriched in group 4 and *Kit* enriched in group 5. Notably, we previously found that the Sst-MET-2 type (L2/3 BC-like cells) may be another transition type between Pvalb and Sst subclasses<sup>12,15</sup>. Moreover, Sst clusters 795 and 797 and Pvalb cluster 752 are all mapped to Sst-MET-2, consistent with their relatedness in the transcriptomic space and suggesting convergent differentiation.

CGE-derived neurons emerge in the cortex around E15.5, and they gradually split into the Vip (*Id4*, *Npas1* and *Synpr*), Sncg (*Npas1*, *Synpr*, *Ptprm* and *Id2*) and Lamp5 (*Bcl11b*, *Ptprm* and *Id2*) subclasses during E18–P2 (node 12; Figs. 2a and 3d,h and Extended Data Fig. 5).

In the Vip subclass, we identified five main developmental trajectories, with clusters in each group often splitting at late postnatal ages, a result that suggests there is continued diversification (Fig. 3d,h and Extended Data Fig. 10). Most Vip clusters are present in L2/3, except for cluster 639 enriched in the deep layers. Vip-MET-1-5 types represent L2/3-5 bipolar or bitufted cells<sup>15</sup>. Most clusters in groups 1–3 are mapped to Vip-MET-4 and Vip-MET-5. Cluster 641 corresponds to Vip-MET-2, and cluster 663 and group 4 clusters 660–662 all correspond to the Vip-MET-1 type<sup>38</sup>. Group 5 (*Grin3a* and *Igf1bp6*) is a highly distinct Vip type with almost no Vip expression and no matching MET-type<sup>38</sup>, which suggests that these cells were not sampled in the Patch-seq study<sup>15</sup>. The Sncg subclass has one main trajectory, with three clusters all corresponding to the Sncg-MET-1 type, which is the main type for CCK<sup>+</sup> basket cells<sup>12,15</sup>. In the Lamp5 subclass, we identified four main developmental trajectories. Groups 2–4 clusters 706, 708, 709 and 718 all correspond to the Lamp5-MET-1 type, which represents the L1-5 neurogliaform cells<sup>12,15</sup>. No MET-type matches group 1 clusters, which suggests that these cells were not sampled in the Patch-seq study<sup>15</sup>. The Lamp5 cells are predominantly found in L1, whereas cluster 709 also includes neurons located in deeper layers.

Taken together, the above results reveal a high degree of correspondence between transcriptomic trajectories and the morphoelectrical

properties of highly specific Sst and Pvalb neuronal types, as well as major Vip, Sncg and Lamp5 neuronal types. Most GABAergic MET-types correspond to distinct trajectory paths with early developmental origins, with late-arising clusters (T-types) contributing to diversification in each MET-type. A prominent exception is the many Sst Martinotti cell and non-Martinotti cell MET-types. Each of these corresponds to a specific set of Sst clusters emerging in late postnatal ages, which results in several MET-types with different axon-targeting specificity contained in a trajectory group. This result suggests that the extensive transcriptomic cell-type diversification of Sst neurons is associated with the formation and refinement of the intricate local circuit motifs between Sst types and other inhibitory and excitatory neuron types<sup>17,57</sup>.

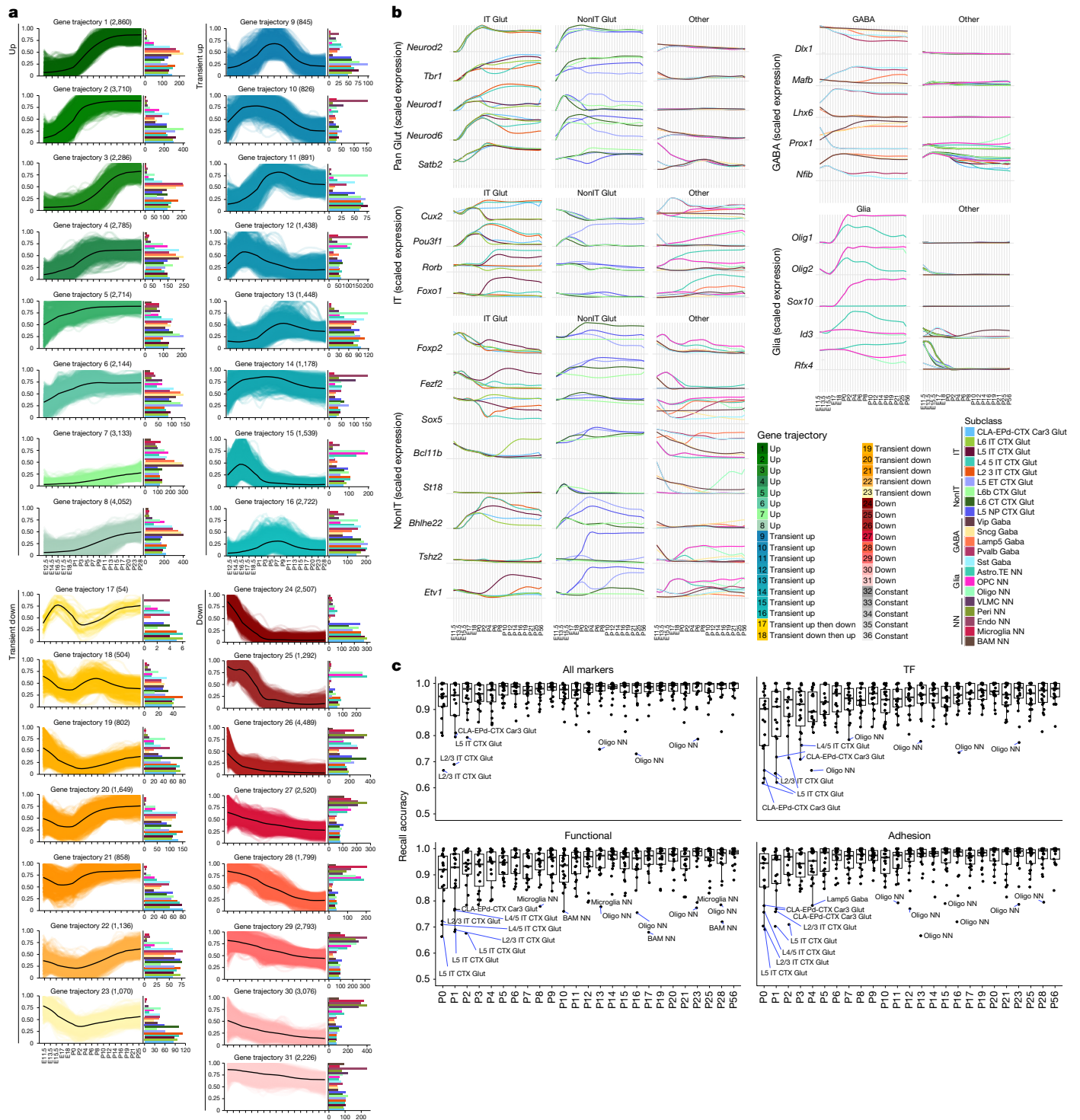
## Gene expression trajectories

To analyse the dynamics of gene expression changes during development, we used generalized additive models to describe temporal profiles of 4,973 developmentally regulated genes in each subclass. Profiles were clustered into 36 trajectory patterns, which were further categorized into 5 general developmental gene trajectories: up, transient up, transient down, down and constant (Fig. 4a and Methods). In the same category, different patterns may show subtle but important differences. For example, trajectory 2 goes up in early embryonic age and plateaus in early postnatal age, whereas trajectory 3 goes up postnatally and plateaus after eye opening.

We further clustered genes on the basis of their similar temporal trajectory in each subclass (Extended Data Fig. 11a). Although most genes exhibit similar temporal expression patterns across related subclasses, many genes also display subtler subclass-specific temporal dynamics. For example, although *Tbr1* expression generally increases during neurogenesis, it begins to diverge between upper and deeper-layer neurons in both IT and non-IT classes after the IMN stage, with particularly substantial downregulation in L5 ET cell types (Fig. 4b). Many important TF genes, including *Cux2*, *Foxp2*, *Nfia* and *Nfib*, show complex temporal dynamics across different subclasses, which suggests that they have tightly regulated, cell-type-specific roles during cortical development. TFs involved in cell-type specification in different subclasses with distinct temporal patterns, as well as some other well-known marker genes, are shown in Fig. 4b and Extended Data Fig. 11b–f.

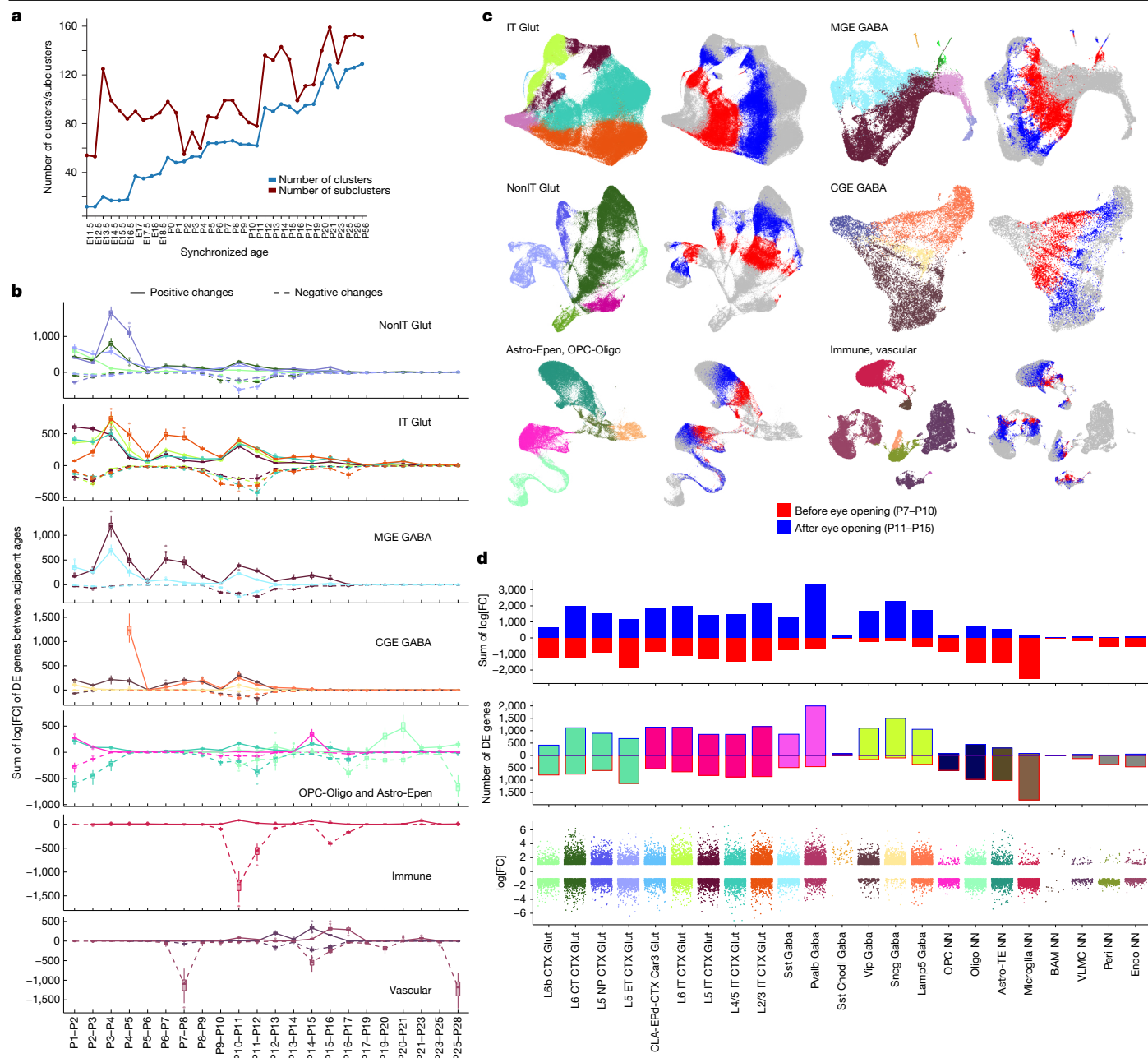
To assess how well the expression of TFs and other gene families at P56 corresponds to transcriptomic cell types during development, we identified DE genes between each pair of subclasses at P56. Using this set of 1,035 adult DE genes, along with 3 curated gene sets (71 TF markers, 139 functional genes and 183 genes encoding adhesion molecules), we trained classifiers to predict subclass identity across postnatal ages (P0–P56) and evaluated their performance using cross-validation (Fig. 4c and Methods). The median subclass recall accuracy is generally lower at earlier ages, as expected, but still maintains medium to high recall accuracy for various subclasses across ages for all gene sets. The variation in recall accuracy is greater during early postnatal ages, with more subclasses showing reduced performance compared with P56. These findings highlight both the transcriptomic distinctiveness of subclasses over postnatal developmental time and the robustness of subclass identities even in early postnatal ages.

We also identified gene modules across cell types and ages, as gene modules could provide a more integrated description of complex biological processes such as cell-type diversification than individual genes (Supplementary Table 6 and Supplementary Fig. 2). Using gene ontology (GO) term enrichment analysis, we assigned biological processes to most modules. The roles of these modules cover several key aspects of brain development, including cell fate determination, cell division, synapse function, immune function and myelination.



**Fig. 4 | Gene expression trajectories across cell types and ages during development.** **a**, Scaled expression of 31 non-constant gene trajectory groups, which are divided into up, transient up, transient down, and down categories. Faint lines represent individual DE genes, whereas the bold central line represents the average trajectory across all DE genes in the group. At each developmental age, the average trajectory was computed as the mean expression value of all genes assigned to the trajectory group. Bar graph to the right of each panel shows number of genes from each subclass that are included in the gene trajectory group. **b**, Representative gene expression trajectories for selected marker genes of pan-glutamatergic (Pan Glut), IT, nonIT, GABAergic or glial cell subclasses, which show higher and more sustained expression (with

variable onsets) in the specific subclasses they represent than other subclasses. **c**, Cross-validation recall accuracy for each subclass at each postnatal age (P0–P56) using classifiers built based on all 1,035 adult marker genes (All markers), 71 TF marker genes (TF), 139 functional marker genes (Functional) and 183 marker genes encoding adhesion molecules (Adhesion). Subclasses with low recall accuracy scores at specific time points are labelled. For the box plots, the central line indicates the median value of accuracy, the box spans the interquartile range (IQR; 25th–75th percentile), and the whiskers extend to 1.5× the IQR, with outliers individually plotted. Labels are displayed for points with accuracy <0.8. The number of cells in each subclass at each age is shown in Supplementary Table 3.



**Fig. 5 | Dynamic gene expression changes before and after eye opening.** **a**, Number of clusters and subclusters at each synchronized age. **b**, Quantification of changes of DE genes in each subclass across development, grouped by classes. For each sliding pair of adjacent synchronized ages, we calculated the sum of log<sub>2</sub>[FC] of DE genes in each subclass, with positive and negative changes (later time minus earlier time) calculated separately (Methods). The box plots show repeated DE analysis 100 times for each pair of adjacent synchronized ages, each time randomly subsampling 70% of the selected cells in each subclass (Methods and Supplementary Table 3). The central line indicates the median value, the box spans the IQR range (25th–75th percentile), and the whiskers extend to 1.5× the IQR, with outliers individually plotted. Microglia are not

represented in the P3–P4, P4–P5 and P25–P28 comparisons owing to an absence of this population in at least one group in each comparison. Similarly, VLMCs are absent across all comparisons from P1–P2 to P4–P5. Subclass colours of the curves are the same as subclass colours shown in **c** and **d**. **c**, UMAP representations of the neuronal and non-neuronal classes, coloured by subclass (left panels of each class) and synchronized age (before eye opening: P7–10; after eye opening: P11–15; right panels). **d**, DE genes between ages before and after eye-opening for all cell subclasses. Top, sum of log<sub>2</sub>[FC] of all DE genes upregulated (blue) or downregulated (red) during eye opening. Middle, number of DE genes upregulated or downregulated during eye opening (coloured by class). Bottom, log<sub>2</sub>[FC] of each DE gene (coloured by subclass).

### Dynamic changes during eye opening

The above trajectory analysis reveals increased cell-type diversity in the visual cortex from early to late developmental ages, as well as extensive transcriptional heterogeneity in a cell type, as shown in the single cluster of RG (Fig. 2d). To quantify diversity and heterogeneity, we plotted the total numbers of clusters and subclusters for all

subclasses across synchronized ages (Fig. 5a). The number of clusters continues to increase with time, with jumps at P11–P13 and P19–P21. By contrast, at the subcluster level, there are several bouts of increased subcluster numbers, which indicates heightened heterogeneity of transitional cell subtypes or cell states at different time periods that are associated with specific developmental events. These events include neurogenesis (E13.5–P1), axon growth and synapse formation (P5–P9),

eye opening (P12–P15) and critical period of experience-dependent plasticity (P20–P28).

To quantify the dynamics of gene expression across cell types during postnatal development (P0–P28), we identified DE genes between each pair of adjacent ages in each subclass. For each comparison, we calculated the sum of log fold changes ( $\log_2[\text{FC}]$ ) of DE genes (Fig. 5b and Methods). Among neuronal cell types such as nonIT Glut, IT Glut and MGE GABA, substantial transcriptional changes are observed during P0–P5 and around the time of eye opening (P10–P15). After this stage, the rate of change considerably reduces. Non-neuronal cell types also exhibit distinct patterns. OPC-Oligo and Astro-TE display marked shifts in gene expression during P0–P4 and P9–P13. After eye opening, OPCs and astrocytes show notable changes during P14–P16, whereas oligodendrocytes undergo substantial changes during P17–P21 and P25–P28. Microglia exhibit downregulation of genes during P10–P12 and P15–P17.

The substantial changes during eye opening and the emergence of new cell types after eye opening spurred our exploration into the molecular characteristics preceding and following this event. We first conducted DE gene analysis before and after eye opening in each subclass or cluster, combining scRNA-seq data during P7–P10 for the before eye-opening period and during P11–P15 for the after eye-opening period (Fig. 5c,d and Supplementary Table 7). Genes with  $|\log_2[\text{FC}]| > 1$  and false discovery rate (FDR)  $< 0.01$  were considered to have significant expression changes. Notably, all neuronal and non-neuronal subclasses have diverse transcriptional changes and there are genes upregulated or downregulated for each subclass and cluster (Fig. 5d and Extended Data Fig. 12a). On average, glutamatergic subclasses, including both IT and nonIT (around 1,200–2,000 DE genes for each subclass), have more DE genes than GABAergic subclasses, except for the Pvalb subclass. Although most neuronal subclasses have more upregulated genes than downregulated genes, all non-neuronal subclasses have more downregulated genes, especially microglia.

The GO term enrichment analysis (Extended Data Fig. 12b) shows that there is strong enrichment in the semaphorin–plexin signalling pathway and the anchoring junction in downregulated genes in glutamatergic neurons after eye opening. Conversely, genes associated with presynapse, synaptic membrane, potassium ion transport, regulation of membrane potential and regulation of neuronal synaptic plasticity are significantly upregulated in glutamatergic neurons. There is also enrichment of specific GO terms in specific subclasses. After eye opening, genes associated with myelin sheath are broadly enriched across different neuronal subclasses, in all IT subclasses except for L4/5 IT, in L6 CT and in all GABAergic subclasses.

Identified DE genes include many immediate-early genes (IEGs), such as *Fos*, *Fosb*, *Fosl2*, *Egr1*, *Arc*, *Bdnf* and *Nr4a3* (Extended Data Fig. 12c–h), consistent with previous findings<sup>58</sup>. These IEGs often have different temporal patterns among different subclasses, which suggests that different IEGs may have different effects in cortical microcircuits. Neuronal populations exhibit a higher number of significantly expressed IEGs, whereas non-neuronal populations, particularly immune and vascular cells, also express IEGs albeit generally more weakly.

## Chromatin accessibility landscape across development

Chromatin accessibility and transcriptomic profiles for each nucleus can be simultaneously obtained using the snMultiome dataset. The dataset contains a total of 331,831 nuclei from 35 libraries collected from 13 embryonic and postnatal time points (Extended Data Fig. 1 and Supplementary Table 1). We applied a similar QC strategy as for the scRNA-seq dataset (Methods and Supplementary Table 2). We conducted global clustering and integration with the scRNA-seq and Multiome snRNA-seq data across all ages, as well as integration of P0 Multiome snRNA-seq data with P0 whole brain MERFISH data. During these processes, we further identified and removed lower-quality clusters or clusters present outside the cortex, which resulted in a

final set of 200,061 high-quality nuclei for further analysis (Extended Data Fig. 1).

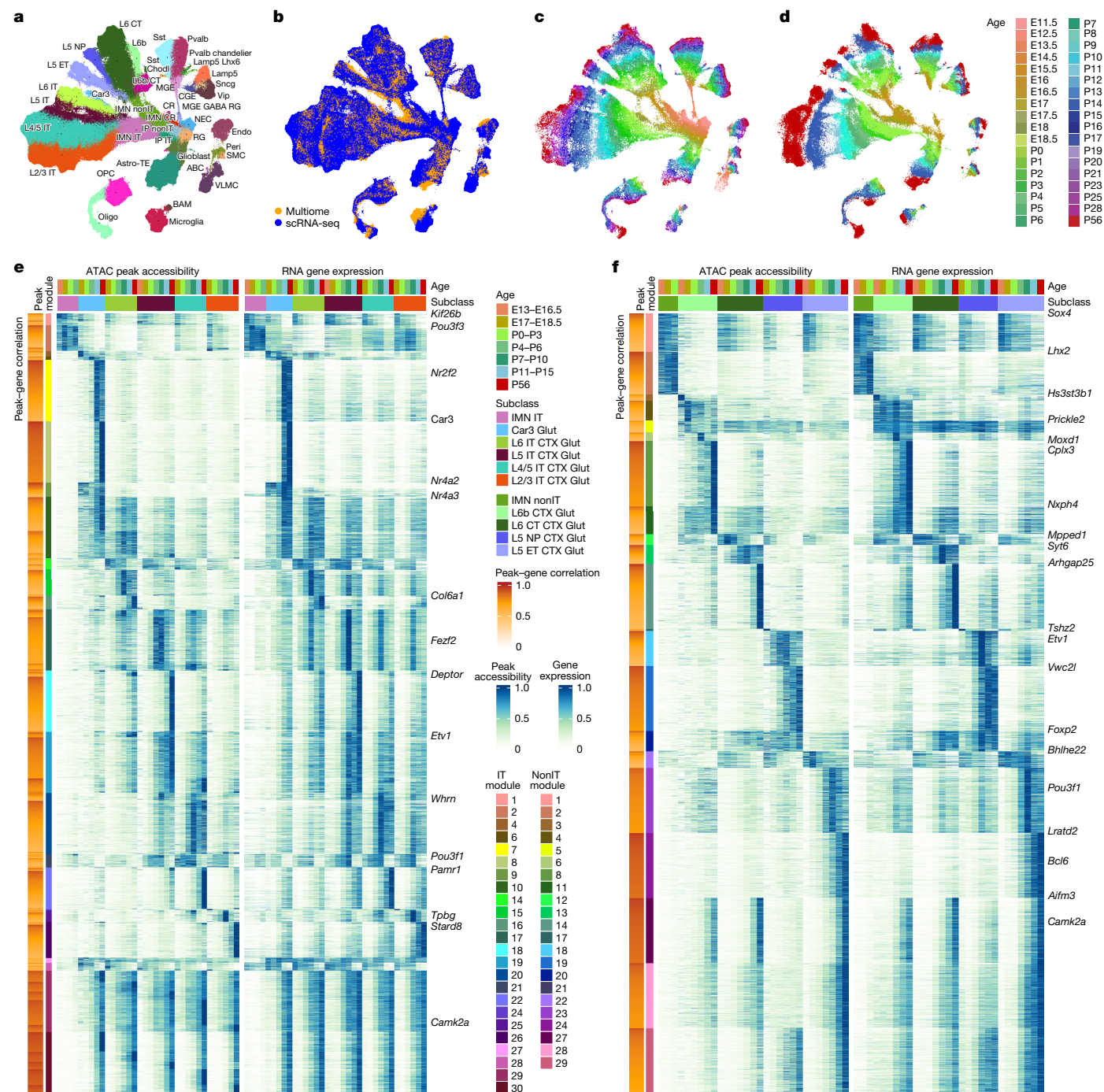
Through integration between Multiome snRNA-seq and scRNA-seq datasets using scVI (Methods), we obtained the transferred cell class, subclass and cluster labels from the reference scRNA-seq atlas for each nucleus (Supplementary Table 8). Owing to sparser postnatal time points for the snMultiome data, we combined the timepoints into age groups that are consistent with the synchronized age bins. The UMAPs based on the integrated scVI latent space show high intermixing of the scRNA-seq and snRNA-seq data, and clear delineation of subclasses and age groups that are consistent between the two datasets (Fig. 6a–d).

The snMultiome analysis pipeline is summarized in Extended Data Fig. 13a. We called chromatin accessibility peaks (total 882,075 peaks) using ArchR<sup>59</sup> based on pseudobulk sets composed of mapped subclasses, clusters and categories defined by both subclass and age group (Methods). We then performed pairwise DA peak analysis between all subclasses and between all subclass-by-age groups using Chi-squared tests. To study the peaks involved in the regulation of cell types and their temporal dynamics, we identified peak modules with similar subclass specificity and temporal patterns among the DA peaks based on their average accessibility across subclass-by-age groups. To associate peak accessibility with gene expression, we identified all the DA peak and DE gene pairs such that the DE gene is within a 5-Mb window centred at the DA peak and the corresponding gene expression and peak accessibility have a correlation of  $>0.5$ . For visualization purposes, for each peak module, we selected the top 500 peak–gene pairs with the strongest correlations (Supplementary Table 9).

We first applied this approach to study the subclass specificity in the IT Glut and nonIT Glut classes separately (Fig. 6e,f), as many genes are re-used to specify different cell types in these two classes. For the IT Glut class, we identified early and late peak–gene pairs for each subclass (Fig. 6e). Peak modules 1 and 2 are specific to the IMN IT subclass and are linked to genes such as *Kif26b* and *Pou3f3*. Peak modules 29 and 30 have increasing accessibility for all IT subclasses, with module 30 activated later than module 29. *Nr4a2*, *Nr2f2* and *Car3*, along with their corresponding peaks, are linked to the CLA-EPd-CTX Car3 Glut subclass, whereas *Nr4a3* and *Col6a1* are linked to L6 IT. *Fezf2*, *Etv1* and *Deptor* are linked to L5 IT, *Whrn* and *Pamr1* to L4/5 IT, *Tpbp* and *Stard8* to L2/3 IT, and *Pou3f2* and *Pou3f1* preferentially linked to upper layers, with various temporal dynamics. Overall, cell-type specific TFs tend to turn on early, whereas other functional genes turn on late.

For the nonIT Glut class (Fig. 6f), *Lhx2* is specific to the IMN nonIT subclass. For the L6b subclass, *Hs3st3b1* is an early marker, *Moxd1* and *Cplx3* turn on late, and *Nxph4* expression remains relatively stable in development. Similarly, *Syt6* and *Arhgap25* are the early and late L6 CT markers, respectively, *Tshz2* and *Vw2cl* are the early and late L5 NP markers, respectively, and *Pou3f1* and *Lratd2* are the early and late L5 ET markers, respectively. These marker genes all have matching chromatin accessibility profiles with similar subclass and temporal specificities. There are also other peak modules that are specific, but shared by multiple subclasses. For example, module 11 is shared by L6b and L6 CT, module 20 is shared by L6 CT and L5 NP, and module 29 is shared by L5 NP and L5 ET. Peak modules 27 and 28, which are generally upregulated during development, are specifically silenced in L5 NP. It is interesting that these peak modules are more distinct among different subclasses than those for the IT Glut class, which show more gradient differences among subclasses.

The GABAergic and glia populations show similar patterns (Extended Data Fig. 13b,c). For each subclass, we identified early and late gene markers and corresponding accessibility peaks. Many well-known GABAergic subclass markers, such as *Lamp5*, *Sncg*, *Vip* and *Pvalb*, as well as glial markers like *Mbp* and *Aqp4*, along with their associated accessibility peaks, are activated at relatively late developmental ages. An exception is *Sst* and its corresponding peaks, which are prenatally



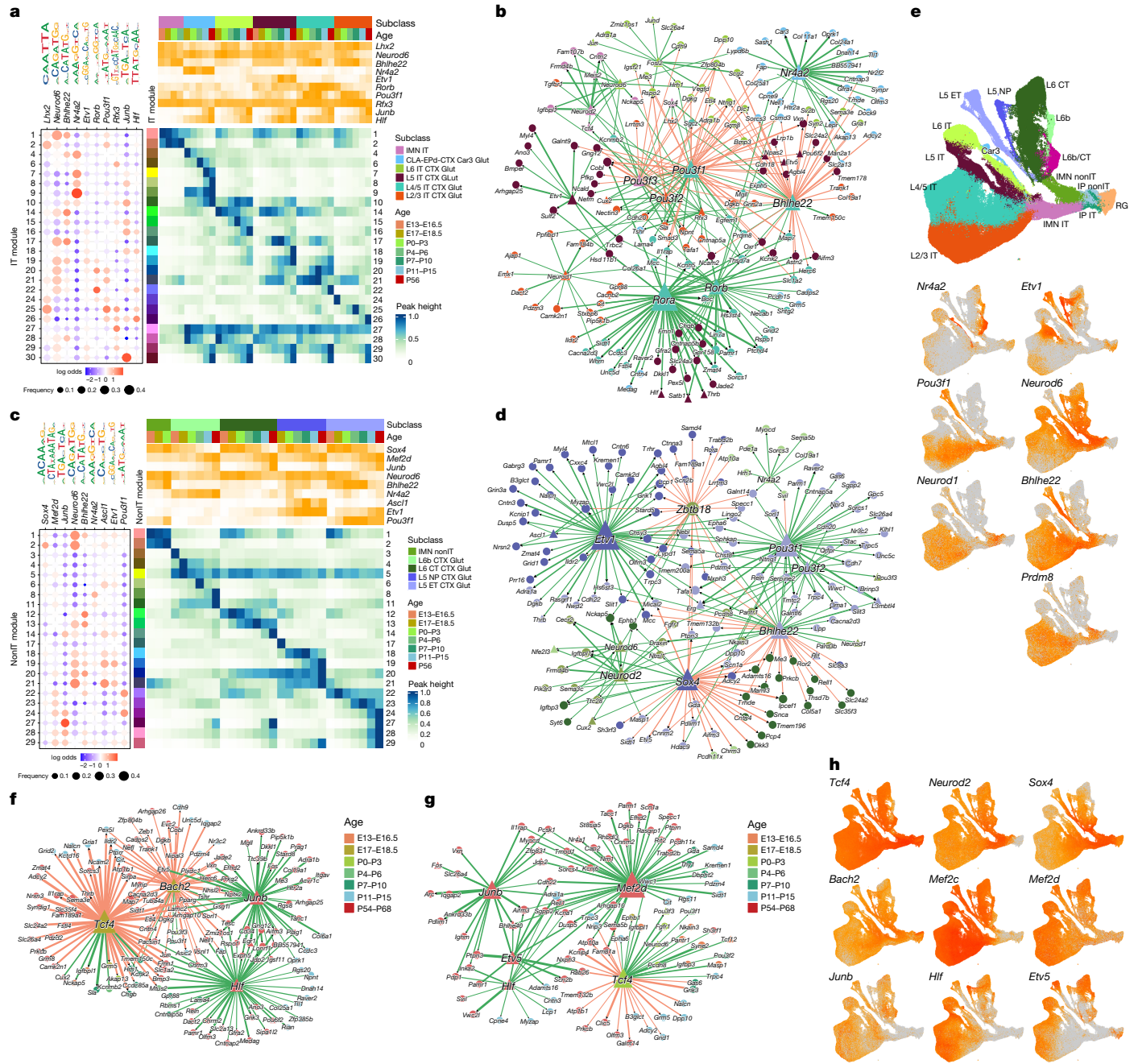
**Fig. 6 | Integration of scRNA-seq and snMultiome data and correspondence of chromatin accessibility and gene expression across glutamatergic neuron types and ages during development.** **a–d**, UMAP representations of scRNA-seq cells and snMultiome nuclei in the integrated space, coloured by subclass (**a**), data modality (**b**) and age in scRNA-seq (**c**) or snMultiome (**d**) datasets. The scRNA-seq cells shown in the UMAP are the subsampled ones (up to 200 cells per cluster) used for scVI integration. **e, f**, Heatmap representations of correspondence of chromatin accessibility and gene expression across IT

subclasses (**e**) and nonIT subclasses (**f**) and ages during development. In each panel, each row corresponds to a peak–gene pair, ordered by peak module and peak–gene correlation, and each column corresponds to a subclass-by-age group. The left-hand and right-hand heatmaps show the average peak accessibility and average gene expression, respectively, in each subclass-by-age group. Accessibility and expression values are normalized, with a maximum value of 1 per peak or gene and 0 indicating no accessibility or expression.

activated and continue to increase until plateauing around P10. Temporal profiles with greater resolution of many of these genes based on scRNA-seq are also shown in Extended Data Fig. 11.

We examined the epigenomic landscape of several genes that are expressed in multiple cell types in different ages. Genes with long gene bodies are regulated by distinct peaks at different ages and across various

cell types, such as *Cux2* (Extended Data Fig. 14) and *Grik1* (Extended Data Fig. 15). Both genes have long gene bodies (193 kb for *Cux2* and 394 kb for *Grik1*) and are associated with highly distinct peaks in different cell types and at different developmental ages. By contrast, the TF gene *Fezf2* has only a 6-kb gene body, and most of the regulatory elements are packed within a 10-kb window around the gene body (Supplementary Fig. 3).



**Fig. 7 | GRNs for IT and nonIT glutamatergic neuron types. a,c, TF DNA-binding motif enrichment in chromatin accessibility peak modules with different cell-type and temporal specificities in IT subclasses (a) and nonIT subclasses (c). In each panel, the dot plot to the left shows the average motif frequency for each peak module (as shown in Fig. 6), dot size indicates the frequency and colour corresponds to the log odds of motif occurrence in each module relative to random chance. The large heatmap at the bottom shows the average accessibility for each peak module (in rows) across each subclass-by-age group (in columns). The heatmap at the top shows the average expression of specific TFs corresponding to the motifs across each subclass-by-age group. The values in the heatmaps are normalized per peak module or gene, with 1**

indicating the maximum value and 0 indicating no accessibility or expression. **b,d**, GRNs for IT subclasses (b) and non-IT subclasses (d). Nodes represent genes, with triangles denoting TFs and circles denoting other genes. Each node is coloured according to the subclass in which the gene is most highly expressed. Activation interactions are in green, repression in orange, and edge widths reflect interaction strengths. **e**, Expression of shared cell-type TF regulators between IT and nonIT subclasses. **f,g**, GRNs involved in temporal regulation for IT subclasses (f) and nonIT subclasses (g). Each node is coloured by the age group in which each gene is most highly expressed. **h**, Expression of shared temporal TF regulators between IT and nonIT subclasses.

## TF regulators and gene regulatory networks

To identify the potential TF regulators for each peak module, we performed differential TF DNA-binding motif analysis between peak modules using all pairwise comparisons (Methods). For the TF motifs that seem significant in any pairwise comparison, we plotted the motif

presence frequencies across all the peak modules together with the average subclass and temporal accessibility pattern of each module (Fig. 7a,c and Extended Data Fig. 16a,d). We used TFs associated with differential TF motifs as potential regulators to construct a gene regulatory network (GRN) by using the SCENIC+ framework<sup>60</sup> (Methods and Extended Data Fig. 13a). We aimed to identify both activating and

repressing TF interactions. Each TF–peak–target triplet was assigned a confidence score, and only the top-ranked predictions are presented (Methods and Supplementary Table 10). Given that motifs from the same TF family often exhibit similar enrichment patterns, and their associated TFs can be strongly correlated, we highlight only the most influential regulators from each TF family as determined by GRN analysis (see the additional notes in the caption of Extended Data Fig. 16).

In the IT Glut class (Fig. 7a,b), the *Lhx2* motif is enriched in peak modules specific to IMN IT and L2/3 IT subclasses, whereas *Lhx2* expression decreases during IT neuron maturation except for L2/3, thereby potentially contributing to the maintenance of upper-layer identity. *Nr4a2* is identified as a key regulator of the CLA-EPd-CTX Car3 subclass, with its motif enriched in both early and late Car3 subclass-specific peak modules. *Rora* and *Rorb* are identified as key regulators of the L4/5 IT subclass, with both showing the strongest motif enrichment in L4/5 IT-specific peak modules and regulating the L4/5 IT subclass markers *Whrn* and *Rspo1*. *Rora* and *Rorb* have highly similar binding motifs. *Rora* shows wider expression, whereas *Rorb* is highly specific to L4/5 IT subclass, and they may cooperate to regulate the same set of target genes. The POU-III class TF genes *Pou3f1*, *Pou3f2* and *Pou3f3* are also identified as key regulators for upper-layer neurons. This finding is consistent with their crucial roles in specifying and maintaining the identity of these neuronal populations<sup>61</sup>. The analysis also predicts *Cux2* as a downstream target.

For the nonIT Glut class (Fig. 7c,d), we identified *Nr4a2* as a regulator of L6b, *Ascl1* and *Etv1* as regulators of L5 NP, and *Pou3f1* as a regulator of L5 ET, along with their cell-type specific targets. Similar to the IT subclasses, *Bhlhe22* and *Neurod6* have similar peak module enrichment, but opposite expression patterns in the nonIT subclasses. *Bhlhe22* and *Neurod1* show stronger expression, whereas *Neurod6* shows weaker expression in L5 ET. Our analysis suggests that *Bhlhe22* represses IMN nonIT and L6 CT markers in L5 ET while activating L5 ET markers, possibly through cooperation with *Neurod1*.

Comparisons of the GRNs for both IT and nonIT subclasses identified several shared regulators (Fig. 7e): *Nr4a2* in Car3 and L6b, *Etv1* in L5 IT and L5 NP, *Pou3f1* in L2/3 IT and L5 ET, *Neurod6* in deep-layer IT neurons and nonIT subclasses except for L5 ET, and *Neurod1* and *Bhlhe22* in upper-layer IT neurons and L5 ET. Considering that L5 ET is the most superficial layer in the nonIT class (especially at P0; Fig. 2e), this TF code may reflect relative laminar positioning in each class.

In addition to TFs that regulate cell-type specificity, we identified a conserved temporal regulation network shared by most cortical glutamatergic cell types (Fig. 7f–h). *Tcf4* and *Sox4* are upregulated in the IP stage and gradually decrease after the IMN stage. We predict that *Sox4* simultaneously represses the premature expression of certain neuronal markers, particularly L6 CT markers (Fig. 7c,d), to help maintain cells in the IMN state. *Mef2c* and *Mef2d*, which show a steady increase throughout development, have crucial roles in neuronal differentiation, synaptic connectivity and neuronal survival<sup>62</sup>.

We identified the AP-1 complex, including *Junb* and *Fos*, as a central regulator of activity-dependent gene expression after eye opening (Fig. 7a,c,f,g and Extended Data Fig. 12). Activated by synaptic activity, AP-1 regulates other IEGs such as *Arc*, which is involved in synaptic remodelling and maturation<sup>63</sup>. Our analysis reveals that many genes with peak expression and chromatin accessibility in adulthood are probably AP-1 targets. We also identified *Bach2*, a transcriptional repressor, as a key temporal regulator that binds AP-1-like motifs (Fig. 7f,h). Unlike *Junb* and *Fos*, the expression of which increases into adulthood, *Bach2* expression declines after eye opening. We predict that *Bach2* counteracts AP-1 activity by competing for shared binding sites, consistent with its known function in repressing AP-1-driven transcription in the immune system<sup>64</sup>.

Moreover, we identified *Hlf* and *Etv5*, both upregulated after eye opening, as important regulators of late-stage neuronal development (Fig. 7f–h). *Etv5*, which is involved in axonal growth, dendritic

arborization and circuit formation, is regulated by the neurotrophic factor BDNF in dorsal root ganglion and hippocampal development<sup>65,66</sup>, whereas BDNF itself is upregulated after eye opening (Extended Data Fig. 12). Together, these findings reveal a core temporal transcriptional network that coordinates the sequential stages of cortical glutamatergic neuron maturation. This network balances early neurogenesis, differentiation and synaptic development with the repression of premature gene activation to ultimately ensure proper circuit formation, activity-dependent plasticity and long-term neural stability.

We conducted a similar GRN analysis to identify key TF regulators of GABAergic neuron development. In addition to temporal regulators shared with glutamatergic cell types such as *Sox4*, *Mef2c* and *Junb*, we identified *Mafb* and *Sox6* as major MGE regulators and *Nfib* and *Nr2f2* as CGE regulators (Extended Data Fig. 16a–c). Notably, despite their opposing expression patterns, *Sox6* and *Nfib* exhibit highly similar motif enrichment profiles in peak modules specific to CGE subclasses. These are in turn anticorrelated with the *Mafb* motif, which is enriched in MGE-specific modules. *Sox6*, known as a transcriptional repressor downstream of *Lhx6*, is essential for the differentiation and diversity of MGE-derived cortical interneurons<sup>67,68</sup>. Our analysis reveals that many CGE subclass marker genes activated by *Nfib* are repressed by *Sox6*, which highlights its role in reinforcing MGE identity. Moreover, we predict that *Sox6* promotes the expression of markers enriched in Pvalb neurons, consistent with the role of *Sox6* in regulating their synaptic function postnatally<sup>69</sup>. Finally, we identified *Esrrg* as a strong regulator of Pvalb subclass markers, including *Pvalb* itself and some markers shared with the Sst subclass.

We analysed the GRNs involved in gliogenesis using a similar approach, incorporating the IP subclass into the analysis to explore the bifurcation between glial and neuronal development (Extended Data Fig. 16d–f). As expected, we identified neurogenic factors such as *Neurod2* as the main regulator of neurogenesis. We identified *Tfap2c* mainly as a RG regulator, but it is also expressed in glioblast SVZ cells. *Tfap2c* is known to promote neuronal fate by directly regulating *Neurod4* and *Eomes*<sup>70</sup>.

We identified *Rfx4*, *Rora*, *Rorb* and *Nr3c2* as key regulators of astrocyte development (Extended Data Fig. 16d–f). *Rfx4* is expressed in RG and glioblasts, with its expression increasing in astrocytes and decreasing in neuronal and oligodendrocyte populations. *Rora* and *Rorb* are similarly upregulated in glioblasts and continue to increase during postnatal astrocyte maturation. *Nr3c2* is activated later in the astrocyte trajectory. Our analysis suggests that *Rorb* and *Rora* may act upstream of both *Rfx4* and *Nr3c2*, forming a feed-forward regulatory network that promotes astrocyte maturation. In the oligodendrocyte trajectory, we observed significant enrichment of SOX motifs, particularly *Sox6*, *Sox8* and *Sox9*, in oligodendrocyte-specific but not OPC-specific peak modules (Extended Data Fig. 16d–f). *Sox8* and *Sox10* are selectively expressed in the OPC-Oligo class. By contrast, *Sox6* and *Sox9* are broadly expressed in RG, glioblasts, astrocytes and OPCs, but are turned off during oligodendrocyte maturation, *Sox9* in late OPCs and COPs and *Sox6* in NFOLs. On the basis of these patterns, we propose that *Sox8* and *Sox10* promote oligodendrocyte maturation, whereas *Sox6* and *Sox9* act as stage-specific repressors of maturation.

## Epigenomic changes before and after eye opening

Motivated by the transcriptomic differences observed before and after eye opening (Fig. 5), we investigated epigenomic changes during this developmental stage. For each subclass, we computed DA peaks between P7–P10 and P11–P15. A total of 32,865 DA peaks were identified and then sorted by the subclass with the highest accessibility and the age group (Extended Data Fig. 17a). More DA peaks are detected in IT subclasses than other subclasses. Among glutamatergic subclasses, more increasing peaks than decreasing peaks are seen after eye opening, particularly in the L5 IT and L6 CT subclasses. Considerable overlap

in DA peaks occurs among glutamatergic subclasses, especially among the IT subclasses (Extended Data Fig. 17a,b). Strong correlations of chromatin accessibility changes are evident among L2/3 IT, L4/5 IT and L5 IT subclasses, followed by L6 IT, L6 CT and L5 ET subclasses. By contrast, L5 NP, non-neuronal and GABAergic subclasses show minimal correlated changes with any other subclasses (Extended Data Fig. 17c). To mitigate the bias stemming from difference in cell numbers, we also computed separately the sum of positive and negative changes in the common set of 32,865 DA peaks for each subclass (Extended Data Fig. 17d). The metric solely assesses the absolute changes without factoring significance, which makes it less sensitive to sample size. The overall amount of positive and negative changes shows the same trend as the number of DA peaks across different subclasses, albeit with smaller variations. For example, the amount of change for L6 IT subclass, which has few DA peaks (presumably owing to smaller cell numbers), is now a lot more comparable with other IT and L6 CT subclasses.

## Discussion

Several important insights are obtained in the current study. We find transcriptional heterogeneity in each of the embryonic cell populations (that is, RG, glioblasts, IPs and IMNs), which suggests that early specification of cell fates becomes increasingly apparent and distinct with time (Figs. 1–3). After neurogenesis, both excitatory and inhibitory neurons exhibit gradually increased complexity, with new subclasses and types emerging along the developmental timeline, including a burst of new cell types after eye opening and at critical period, especially for the IT and ET excitatory neurons and the Sst inhibitory neurons (Figs. 2, 3 and 5 and Extended Data Figs. 6–10). Throughout development, there are cooperative dynamic changes in gene expression and chromatin accessibility in specific cell types. We identify both chromatin peaks that potentially regulate the expression of specific genes and TFs that potentially regulate specific peaks, which form extensive GRNs through cell-type specific and temporally resolved TF–peak–target gene interactions (Figs. 6 and 7 and Extended Data Figs. 13 and 16).

A widely accepted concept in early cortical development is a sequential inside–out model. Namely, radial glia progenitors generate deep-layer glutamatergic neurons first, then upper-layer glutamatergic neurons and finally glial cells (astrocytes and oligodendrocytes). Although our data are largely consistent with this developmental paradigm, with nonIT IP and IMN cells appearing the earliest at E13.5, we also observe the emergence of both IT IP and IMN cells and glioblasts at E15.5, and the appearance of astrocytes and OPCs at E17 is also concurrent with that of the adult-like nonIT and deep-layer IT neurons (Fig. 2a,b). Furthermore, our transcriptomic trajectory map shows that the primary division of glutamatergic neurons is between nonIT and IT cell types, which already exists at the IP stage (Fig. 2c), rather than between deep-layer and upper-layer neurons. Later in the transition from IMNs to adult-like neurons, deep-layer neurons, including L6 CT, L5 ET and L6b subclasses from the nonIT class and L6 IT and L5 IT subclasses from the IT class, do appear earlier at E17, whereas the L5 NP (of nonIT), L4/5 IT and L2/3 IT subclasses appear at E18.5. Therefore, the transcriptomic profiles suggest a more nuanced view; that is, a staggered parallel process for the generation of glutamatergic neuron and glial cell types from the common pool of RG. The transcriptomic heterogeneity observed in the RG population, with nonIT and IT neuronal markers and glial markers appearing in a staggered overlapping manner, further supports this view (Fig. 2d). Our findings are compatible with the results from recent studies that revealed extensive heterogeneity in the repertoire of cortical cell types that each RG progenitor generates, which may be due to a series of probabilistic decisions in individual RG progenitors leading to varied lineage progression<sup>28,46</sup>.

Perhaps a more notable finding is the extensive diversification of cell types after birth, with the total number of neuronal clusters increasing from 40 at PO to 48 at P8, 60 at P16 and 93 at P25 (Fig. 5a). Although

nearly all cell subclasses are generated prenatally, the vast majority of cell clusters emerge postnatally. This diversification coincides with the maturation of neurons, the formation of synaptic connections, myelination and activity-dependent plasticity, among other processes. During the eye-opening stage (P11–P14) and around the onset of critical period (P21), many new clusters emerge, especially for the IT excitatory neurons and the Sst and Vip inhibitory neurons (Fig. 3 and Extended Data Figs. 6–10). Using Patch-seq multimodal MET types, we are able to relate the developmental transcriptomic clusters defined here to the traditionally defined GABAergic neuron types based on axon-projection patterns<sup>15,17,38</sup> (Extended Data Figs. 9d and 10d). We demonstrate that the Sst Martinotti and non-Martinotti cells with extensive axon-projection diversity correspond to several specific and distinct Sst trajectories diverging from synaptogenesis (after P7), thereby linking temporally precise transcriptional specificity with connectional specificity. *Cbln4*, a gene that plays critical roles in the synaptic targeting of excitatory neuron dendrites by Sst interneurons<sup>57</sup>, has the highest expression in the Sst clusters that are L2/3–5 fanning Martinotti cells, L4-targeting Martinotti cells or L2/3 fast-spiking-like cells.

We find that eye opening is also associated with broad-ranging, cell-type specific gene expression changes and activated chromatin accessibility peaks (Fig. 5 and Extended Data Figs. 12 and 17), a result that extends beyond previous studies<sup>51,58</sup>. In these changes, the activation of many IEGs in excitatory, inhibitory and non-neuronal types is highly significant, presenting a mechanism for broad gene expression regulations to refine cell-type specific functions. It should be noted that the changes that occur in all cell types after eye opening could be induced by the arrival of external visual sensory information, be just part of the intrinsic circuit maturation or, most probably, due to the interaction between the two. Future studies under normal development or sensory-deprivation conditions, as well as comparison of the circuit development processes across different cortical areas, could help to resolve these contributing factors.

Our GRN analysis uncovers a complex network shaped by major TF families, including bHLH, MEF2, SOX, POU, AP-1 and numerous nuclear receptors such as *Nr4a2*, *Rora*, *Rorb*, *Nr2f2*, *Esrrg* and *Nr3c2* (Figs. 6 and 7 and Extended Data Figs. 13 and 16). Nuclear receptors stand out as key regulators of cell-type specificity. With or without a known ligand, they convert extracellular signals into context-dependent transcriptional responses. Their ability to bind precise DNA motifs, to recruit co-activators or co-repressors based on ligand presence or cell state and to respond dynamically over time enables them to fine-tune gene expression with both cell-type and temporal precision. The GRN analysis can be further enhanced in the future with better characterization of the DNA-binding motifs of more TF genes and with a better understanding of the transcriptional activation and repression mechanisms.

## Online content

Any methods, additional references, Nature Portfolio 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/s41586-025-09644-1>.

1. Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).
2. Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G. & Nowakowski, T. J. Development and arealization of the cerebral cortex. *Neuron* **103**, 980–1004 (2019).
3. Bella, D. J. D., Domínguez-Iturza, N., Brown, J. R. & Arlotta, P. Making Ramón y Cajal proud: development of cell identity and diversity in the cerebral cortex. *Neuron* **112**, 2091–2111 (2024).
4. Petilla Interneuron Nomenclature Group. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
5. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
6. Yuste, R. et al. A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nat. Neurosci.* **23**, 1456–1468 (2020).

7. Fishell, G. & Heintz, N. The neuron identity problem: form meets function. *Neuron* **80**, 602–612 (2013).
8. Huang, Z. J. & Paul, A. The diversity of GABAergic neurons and neural communication elements. *Nat. Rev. Neurosci.* **20**, 563–572 (2019).
9. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
10. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
11. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
12. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
13. Brain Initiative Cell Census Network. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
14. Yao, Z. et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).
15. Gouwens, N. W. et al. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 935–953 (2020).
16. Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
17. Gamlin, C. R. et al. Connectomes of predicted Sst transcriptomic types in mouse visual cortex. *Nature* **640**, 497–505 (2025).
18. Sorensen, S. A. et al. Connecting single-cell transcriptomes to projectomes in mouse visual cortex. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.11.25.568393> (2023).
19. Peng, H. et al. Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).
20. Klingler, E. et al. Temporal controls over inter-areal cortical projection neuron fate diversity. *Nature* **599**, 453–457 (2021).
21. Bugeon, S. et al. A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330–338 (2022).
22. Jabaudon, D. Fate and freedom in developing neocortical circuits. *Nat. Commun.* **8**, 16042 (2017).
23. Favuzzi, E. & Rico, B. Molecular diversity underlying cortical excitatory and inhibitory synapse development. *Curr. Opin. Neurobiol.* **53**, 8–15 (2018).
24. Lim, L., Mi, D., Llorca, A. & Marin, O. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
25. Vanderhaeghen, P. & Polleux, F. Developmental mechanisms underlying the evolution of human cortical circuits. *Nat. Rev. Neurosci.* **24**, 213–232 (2023).
26. Bandler, R. C. & Mayer, C. Deciphering inhibitory neuron development: the paths to diversity. *Curr. Opin. Neurobiol.* **79**, 102691 (2023).
27. Kessar, N. & Denaxa, M. Cortical interneuron specification and diversification in the era of big data. *Curr. Opin. Neurobiol.* **80**, 102703 (2023).
28. Hippenmeyer, S. Principles of neural stem cell lineage progression: insights from developing cerebral cortex. *Curr. Opin. Neurobiol.* **79**, 102695 (2023).
29. Klingler, E. Temporal controls over cortical projection neuron fate diversity. *Curr. Opin. Neurobiol.* **79**, 102677 (2023).
30. Espinosa, J. S. & Stryker, M. P. Development and plasticity of the primary visual cortex. *Neuron* **75**, 230–249 (2012).
31. La Manno, G. et al. Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021).
32. Di Bella, D. J. et al. Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).
33. Telley, L. et al. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019).
34. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
35. Turrero Garcia, M. & Harwell, C. C. Radial glia in the ventral telencephalon. *FEBS Lett.* **591**, 3942–3959 (2017).
36. Marin, O. Cellular and molecular mechanisms controlling the migration of neocortical interneurons. *Eur. J. Neurosci.* **38**, 2019–2029 (2013).
37. Toudji, I., Toumi, A., Chamberland, E. & Rossignol, E. Interneuron odyssey: molecular mechanisms of tangential migration. *Front. Neural Circuits* **17**, 1256455 (2023).
38. van Velthoven, C. T. J. et al. Transcriptomic and spatial organization of telencephalic GABAergic neurons. *Nature* <https://doi.org/10.1038/s41586-025-09296-1> (2025).
39. Schitine, C., Nogaroli, L., Costa, M. R. & Hedin-Pereira, C. Astrocyte heterogeneity in the brain: from development to disease. *Front. Cell. Neurosci.* **9**, 76 (2015).
40. Cao, J. et al. The single cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
41. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
42. Huilgol, D., Russ, J. B., Srivas, S. & Huang, Z. J. The progenitor basis of cortical projection neuron diversity. *Curr. Opin. Neurobiol.* **81**, 102726 (2023).
43. Causeret, F., Moreau, M. X., Pierani, A. & Blanquie, O. The multiple facets of Cajal–Retzius neurons. *Development* **148**, dev199409 (2021).
44. Tucker, R. P. & Chiquet-Ehrismann, R. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev. Biol.* **290**, 237–245 (2006).
45. Zhang, X., Lin, P.-Y., Liakath-Ali, K. & Südhof, T. C. Teneurins assemble into presynaptic nanoclusters that promote synapse formation via postsynaptic non-teneurin ligands. *Nat. Commun.* **13**, 2297 (2022).
46. Llorca, A. et al. A stochastic framework of neurogenesis underlies the assembly of neocortical cytoarchitecture. *eLife* **8**, e51381 (2019).
47. Zahr, S. K. et al. A translational repression complex in developing mammalian neural stem cells that regulates neuronal specification. *Neuron* **97**, 520–537 (2018).
48. Ruan, X. et al. Progenitor cell diversity in the developing mouse neocortex. *Proc. Natl Acad. Sci. USA* **118**, e2018866118 (2021).
49. Hoerder-Suabedissen, A. & Molnár, Z. Development, evolution and pathology of neocortical subplate neurons. *Nat. Rev. Neurosci.* **16**, 133–146 (2015).
50. Condyly, C. et al. Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022).
51. Cheng, S. et al. Vision-dependent specification of cell types and function in the developing cortex. *Cell* **185**, 311–327 (2022).
52. Gibson, E. M. et al. Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science* **344**, 1252304 (2014).
53. Hasel, P. et al. Defining the molecular identity and morphology of glia limitans superficialis astrocytes in vertebrates. *Cell Rep.* **44**, 115344 (2025).
54. Flames, N. et al. Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
55. Wang, B. et al. Loss of Gsx1 and Gsx2 function rescues distinct phenotypes in Dlx1/2 mutants. *J. Comp. Neurol.* **521**, 1561–1584 (2013).
56. McKenzie, M. G. et al. Non-canonical Wnt signaling through Ryk regulates the generation of somatostatin- and parvalbumin-expressing cortical interneurons. *Neuron* **103**, 853–864 (2019).
57. Favuzzi, E. et al. Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits. *Science* **363**, 413–417 (2019).
58. Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
59. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
60. Bravo González-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
61. Sugitani, Y. et al. Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. *Genes Dev.* **16**, 1760–1765 (2002).
62. Assali, A., Harrington, A. J. & Cowan, C. W. Emerging roles for ME2F2 in brain development and mental disorders. *Curr. Opin. Neurobiol.* **59**, 49–58 (2019).
63. Kumar, V., Fahey, P. G., Jong, Y.-J. I., Ramanan, N. & O'Malley, K. L. Activation of intracellular metabotropic glutamate receptor 5 in striatal neurons leads to up-regulation of genes associated with sustained synaptic transmission including Arc/Arg3.1 Protein. *J. Biol. Chem.* **287**, 5412–5425 (2012).
64. Jang, E. et al. Bach2 represses the AP-1-driven induction of interleukin-2 gene transcription in CD4<sup>+</sup> T cells. *BMB Rep.* **50**, 472–477 (2017).
65. Liu, D. et al. Brain-derived neurotrophic factor promotes vesicular glutamate transporter 3 expression and neurite outgrowth of dorsal root ganglion neurons through the activation of the transcription factors Etv4 and Etv5. *Brain Res. Bull.* **121**, 215–226 (2016).
66. Fontanet, P. A., Ríos, A. S., Alsina, F. C., Paratcha, G. & Ledda, F. Pea3 transcription factors, Etv4 and Etv5, are required for proper hippocampal dendrite development and plasticity. *Cereb. Cortex* **28**, 236–249 (2018).
67. Valakh, V. et al. A transcriptional constraint mechanism limits the homeostatic response to activity deprivation in mammalian neocortex. *eLife* **12**, e74899 (2023).
68. Batista-Brito, R. et al. The cell-intrinsic requirement of Sox6 for cortical interneuron development. *Neuron* **63**, 466–481 (2009).
69. Munguba, H. et al. Postnatal Sox6 regulates synaptic function of cortical parvalbumin-expressing neurons. *J. Neurosci.* **41**, 8876–8886 (2021).
70. Pinto, L. et al. AP2γ regulates basal progenitor fate in a region- and layer-specific manner in the developing cortex. *Nat. Neurosci.* **12**, 1229–1237 (2009).

**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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Yuan Gao<sup>1</sup>, Cindy T. J. van Velthoven<sup>1</sup>, Changkyu Lee<sup>1</sup>, Emma D. Thomas<sup>1</sup>, Rémi Mathieu<sup>1</sup>, Angela P. Ayala<sup>1</sup>, Stuard Barta<sup>1</sup>, Darren Bertagnoli<sup>1</sup>, Jazmin Campos<sup>1</sup>, Trangthanh Cardenas<sup>1</sup>, Daniel Carey<sup>1</sup>, Tamara Casper<sup>1</sup>, Anish Bhaswanth Chakka<sup>1</sup>, Rushil Chakrabarty<sup>1</sup>, Megan Chiang<sup>1</sup>, Lindsey Ching<sup>1</sup>, Michael Clark<sup>1</sup>, Marie J. Desierto<sup>1</sup>, Rebecca Ferrer<sup>1</sup>, Jessica Gloe<sup>1</sup>, Jeff Goldy<sup>1</sup>, Nathan Guilford<sup>1</sup>, Junitta Guzman<sup>1</sup>, Carliana R. Halterman<sup>1</sup>, Samantha D. Hastings<sup>1</sup>, Daniel Hirschstein<sup>1</sup>, Windy Ho<sup>1</sup>, Katelyn James<sup>1</sup>, Zoe Juneau<sup>1</sup>, Naomi Martin<sup>1</sup>, Rachel McCue<sup>1</sup>, Emma Meyerdierrks<sup>1</sup>, Amanda C. Mitchell<sup>1</sup>, Josh S. Nagra<sup>1</sup>, Beagan Nguy<sup>1</sup>, Thuc Nghi Nguyen<sup>1</sup>, Paul Olsen<sup>1</sup>, Alana A. Oyama<sup>1</sup>, Nick Pena<sup>1</sup>, Jacob Quon<sup>1</sup>, Qingzhong Ren<sup>1</sup>, Augustin Ruiz<sup>1</sup>, Nadiya V. Shapovalova<sup>1</sup>, Josef Sulc<sup>1</sup>, Amy Torkelson<sup>1</sup>, Alex Tran<sup>1</sup>, Herman Tung<sup>1</sup>, Nasmil Valera Cuevas<sup>1</sup>, Justin Wang<sup>1</sup>, Jeanelle Ariza<sup>1</sup>, Delissa A. M. McMillen<sup>1</sup>, Jack Waters<sup>1</sup>, Michael Kunst<sup>1</sup>, Kara Ronellenfitch<sup>1</sup>, Boaz Levi<sup>1</sup>, Michael J. Hawrylycz<sup>1</sup>, Chelsea Pagan<sup>1</sup>, Nick Dee<sup>1</sup>, Kimberly A. Smith<sup>1</sup>, Bosiljka Tasic<sup>1</sup>, Zizhen Yao<sup>1,2,3</sup> & Hongkui Zeng<sup>1,2,3</sup>

<sup>1</sup>Allen Institute for Brain Science, Seattle, WA, USA. <sup>2</sup>e-mail: [zizheny@alleninstitute.org](mailto:zizheny@alleninstitute.org); [hongkui@alleninstitute.org](mailto:hongkui@alleninstitute.org)

## Methods

### Mouse breeding and husbandry

All experimental procedures related to the use of mice were approved by the Institutional Animal Care and Use Committee of the Allen Institute for Brain Science (AIBS) in accordance with NIH guidelines. Mice were housed in rooms with controlled temperature (21–22 °C) and humidity (40–51%) conditions at no more than five adult animals of the same sex per cage. Mice were provided food and water ad libitum and were maintained on a regular 14:10 h light–dark cycle. Mice were maintained on the C57BL/6J (RRID: IMSR\_JAX:000664) background. We excluded any mice with anophthalmia or microphthalmia.

The presence of vaginal plugs was monitored at 12-h intervals (6:00 and 18:00). To collect embryos with accuracy to 0.5 days, only dams with visible plugs were used to obtain embryonic time points. For postnatal time points, births were recorded at 12-h intervals (6:00 and 18:00). Animal handling was reduced as much as possible until weaning at P21. At weaning, animals were separated from their mothers and opposite-sex siblings. Weaned mice were group-housed, kept separate from the opposite sex and maintained under normal housing conditions until dissection.

All animals used for data generation are listed in Supplementary Table 1. No statistical methods were used to predetermine sample sizes. In total we used 53 mice to collect scRNA-seq data from 913,297 cells across 35 time points between E11.5 and adulthood: embryonic day E11.5, E12.5, E13.5, E14.5, E15.5, E16.5, E17.0, E17.5, E18.0 and E18.5; postnatal day P0, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, P17, P19, P20, P21, P23, P25 and P28; and adult stage P54–P68 (collectively simplified as P56). Brain dissections for all groups took place in the morning. From ages E11.5 to E12.5, we collected whole brain tissue, from ages E13.5 to E14.5, we collected the cerebrum and the brainstem, and from other ages we dissected the visual cortex (VIS). Tissue dissection was performed on about 300- $\mu$ m coronal sections, and the VIS was identified and dissected out using the Allen CCFv3 reference atlas. The adult samples were taken from the ABC–WMB dataset by selecting scRNA-seq libraries with regions of interest (ROIs) of VIS-PTLp or VISp. For snMultiome data generation, we collected data from 331,831 nuclei from 28 mice across 13 time points: E15.5, E16.0, E17.0, E18.0, P0, P2, P4, P5, P8, P9, P11, P14 and P56. At embryonic time points, we dissected the cerebrum and the brainstem and at postnatal time points, we collected either the VIS or the isocortex.

In some cases, transgenic mice were used for fluorescence-positive cell isolation by FACS. To enrich neurons profiled by scRNA-seq, cells were isolated from the pan-neuronal Snap25-IRES2-Cre line (RRID: IMSR\_JAX:023525) crossed to the Ai14-tdTomato reporter (RRID: IMSR\_JAX:007914) (31 out of 53 mice, Supplementary Table 1).

### Single-cell isolation

Single cells were isolated following a cell-isolation protocol developed at the AIBS<sup>71</sup>. The brain was dissected, submerged in artificial cerebrospinal fluid (ACSF), embedded in 2% agarose and sliced into 350- $\mu$ m coronal sections on a compresstome (Precisionary Instruments). Block-face images were captured during slicing. ROIs were then microdissected from the slices and dissociated into single cells.

Dissected tissue pieces were digested with 30 U ml<sup>-1</sup> papain (Worthington PAP2) in ACSF for 30 min at 30 °C. Owing to the short incubation period in a dry oven, we set the oven temperature to 35 °C to compensate for the indirect heat exchange, with a target solution temperature of 30 °C. Enzymatic digestion was quenched by exchanging the papain solution 3 times with quenching buffer (ACSF with 1% FBS and 0.2% BSA). Samples were incubated on ice for 5 min before trituration. The tissue pieces in the quenching buffer were triturated through a fire-polished pipette with a 600- $\mu$ m-diameter opening approximately 20 times. The tissue pieces were allowed to settle and the supernatant, which now contained suspended single cells, was transferred to a new

tube. Fresh quenching buffer was added to the settled tissue pieces, and trituration and supernatant transfer were repeated using 300  $\mu$ m and 150  $\mu$ m fire-polished pipettes. The single-cell suspension was passed through a 70  $\mu$ m filter into a 15 ml conical tube with 500  $\mu$ l high-BSA buffer (ACSF with 1% FBS and 1% BSA) at the bottom to help cushion the cells during centrifugation at 100g in a swinging-bucket centrifuge for 10 min. The supernatant was discarded, and the cell pellet was resuspended in the quenching buffer. We collected 547,092 cells without performing FACS. The concentration of the resuspended cells was quantified, and cells were immediately loaded onto a 10x Genomics Chromium controller.

To enrich neurons or live cells in some samples, cells were collected by FACS (BD Aria II running FACSdiva v.8 (RRID: SCR\_001456)) using a 130  $\mu$ m nozzle, following a FACS protocol developed at AIBS<sup>72</sup>. Cells were prepared for sorting by passing the suspension through a 70  $\mu$ m filter and adding Hoechst or DAPI (to a final concentration of 2 ng ml<sup>-1</sup>). The sorting strategy with example images has been previously described<sup>72</sup>. We collected 30,833 calcein-positive and Hoechst-positive cells, 18,992 Hoechst-positive cells, 13,912 RFP-positive cells and 302,468 RFP-positive and Hoechst-positive cells (Extended Data Fig. 1d, Supplementary Table 1 and Supplementary Fig. 4). Around 30,000 cells were sorted within 10 min into a tube containing 500  $\mu$ l quenching buffer. Each aliquot of sorted 30,000 cells was gently layered on top of 200  $\mu$ l high-BSA buffer and immediately centrifuged at 230g for 10 min in a centrifuge with a swinging-bucket rotor (the high-BSA buffer at the bottom of the tube slows down the cells as they reach the bottom, which minimizes cell death). No pellet could be seen with this small number of cells, so we removed the supernatant and left behind 35  $\mu$ l of buffer, in which we resuspended the cells. Immediate centrifugation and resuspension allowed the cells to be temporarily stored in a high-BSA buffer with minimal ACSF dilution. The resuspended cells were stored at 4 °C until all samples were collected, usually within 30 min. Samples from the same ROI were pooled, the cell concentration quantified and the samples immediately loaded onto a 10x Genomics Chromium controller.

### Single-nucleus isolation

Mice were anaesthetized with 2.5–3% isoflurane and transcardially perfused with cold pH 7.4 HEPES buffer containing 110 mM NaCl, 10 mM HEPES, 25 mM glucose, 75 mM sucrose, 7.5 mM MgCl<sub>2</sub> and 2.5 mM KCl to remove blood from the brain<sup>73</sup>. After perfusion, the brain was rapidly dissected, frozen for 2 min in liquid nitrogen vapour and then transferred to –80 °C for long-term storage using a freezing protocol developed at AIBS<sup>74</sup>.

For VIS dissections, frozen mouse brains were sectioned using a cryostat, with the cryochamber temperature set at –20 °C and the object temperature set at –22 °C. Brains were securely mounted by the cerebellum or by the olfactory region onto cryostat chucks using OCT (Sakura FineTek 4583). Tissue was trimmed at a thickness of 20–50  $\mu$ m, and once at the desired location, slices with a thickness of 300  $\mu$ m were generated to dissect out ROIs following the reference atlas. Images were taken while leaving the dissection in the cutout section. Nuclei were isolated using the RAISINS<sup>75</sup> method, but with a few modifications as described in a nucleus isolation protocol developed at AIBS<sup>76</sup>. In brief, excised tissue samples were transferred to a 12-well plate containing CST extraction buffer. Mechanical dissociation was performed by chopping the sample using spring scissors in ice-cold CST buffer for 10 min. The entire volume of the well was then transferred to a 50 ml conical tube while passing through a 100  $\mu$ m filter and the walls of the tube were washed using ST buffer. Next the suspension was gently transferred to a 15 ml conical tube and centrifuged in a swinging-bucket centrifuge for 5 min at 500 r.c.f. and 4 °C. After centrifugation, the majority of supernatant was discarded, pellets were resuspended in 100  $\mu$ l 0.1 $\times$  lysis buffer and incubated for 2 min on ice. After the addition of 1 ml wash buffer, samples were gently filtered using a 20  $\mu$ m filter and centrifuged as

# Article

before. After centrifugation, most of the supernatant was discarded, pellets were resuspended in 10  $\mu$ l chilled nucleus buffer and nuclei were counted to determine the concentration. Nuclei were diluted to a concentration targeting 5,000 nuclei per  $\mu$ l.

## cDNA amplification and library construction

For 10x scRNA-seq, the cell suspensions were processed using a Chromium Single Cell 3' Reagent Kit v.3 (1000075, 10x Genomics)<sup>77</sup>. We followed the manufacturer's instructions for cell capture, barcoding, reverse transcription, cDNA amplification and library construction. We loaded  $8,876 \pm 2,981$  cells per port. We targeted a sequencing depth of 120,000 reads per cell; the actual average achieved was  $64,719 \pm 60,037$  reads per cell across 91 libraries (Supplementary Table 1).

For 10x snMultiome processing, we used a Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (1000283, 10x Genomics). We followed the manufacturer's instructions for transposition, nucleus capture, barcoding, reverse transcription, cDNA amplification and library construction<sup>78</sup>. For the snMultiome libraries, we loaded  $9,276 \pm 3,883$  nuclei per port. For snRNA-seq, we targeted a sequencing depth of 120,000 reads per nucleus. The actual average achieved, for the nuclei included in this study, was  $108,297 \pm 65,116$  reads per nucleus across 35 libraries (Supplementary Table 1). For snATAC-seq, we targeted a sequencing depth of 85,000 reads per nucleus. The actual average achieved, for the nuclei included in this study, was  $122,312 \pm 74,159$  reads per nucleus across 35 libraries.

## Sequencing data processing and QC

To remove low-quality cells, we developed a stringent QC process. Cells were first classified into broad cell classes after mapping to our established ABC-WMB Atlas<sup>14</sup>, and cell quality was assessed on the basis of gene detection, QC score and doublet score. The QC score was calculated by summing the log-transformed expression of a set of genes for which the expression level was significantly decreased in poor-quality cells. Doublets were identified using a modified version of the DoubletFinder algorithm (available in `scrattch.hicat`; <https://github.com/AllenInstitute/scrattch.hicat>, v.1.0.9) and removed when the doublet score was  $>0.3$ . In prenatal time points, neuronal precursors of non-cortical origin were excluded on the basis of low expression of *Foxg1*, *Emx1* or *Emx2*. Using different QC scores and gene-count thresholds among different cell classes (Supplementary Table 2), we filtered out 151,945 cells and kept 761,352 cells for 10xv3 scRNA-seq data (Extended Data Fig. 1a,b).

We used a similar strategy to filter low-quality nuclei for the 10x Multiome snRNA-seq dataset. Nuclei were first classified into broad cell classes after mapping to the ABC-WMB Atlas, and cell quality was assessed on the basis of gene detection and the doublet score. For the 10x Multiome snRNA-seq dataset, although the overall gene counts were lower compared with the 10xv3 scRNA-seq dataset, they showed a more distinct bimodal distribution, which enabled us to retain stringent QC criteria (Supplementary Table 2). For 10x Multiome snATAC-seq data, we used the default criteria implemented in ArchR (RRID: SCR\_020982)<sup>59</sup>: the number of unique nuclear fragments ( $n_{\text{Fragments}} > 1,000$ ) and signal-to-background ratio (transcription start site (TSS)  $> 3$ ). For the 10x Multiome dataset, only nuclei that passed both snRNA-seq and snATAC-seq QC criteria (a total of 261,380 nuclei) were included in downstream analyses (Extended Data Fig. 1a,c).

## Inferring synchronized developmental age

To estimate the synchronized developmental age for each single cell, we trained  $k$ -NN models (Extended Data Fig. 2a). We first performed global de novo clustering for 10xv3 single-cell datasets across all time points using the R package `scrattch.bigcat`<sup>14</sup> (<https://github.com/AllenInstitute/scrattch.bigcat>). The whole gene-count matrices were chunked to smaller parquet files that could be loaded efficiently and concurrently using the `arrow` package (v.12.0.1; <https://github.com/apache/arrow>)

and <https://arrow.apache.org/docs/r/>). The automatic iterative clustering method `iter_clust_big` was used with stringent differential gene expression criteria as described in a previous study<sup>12</sup>:  $q1.th = 0.5$ ,  $q.diff.th = 0.7$ ,  $de.score.th = 150$ ,  $min.cells = 50$ . We then performed principal component analysis (PCA) based on the gene expression matrix of 5,824 marker genes derived from the de novo clusters. We downsampled up to 200 cells per cluster to avoid memory limitations. The principal components (PCs) based on the sampled cells were then projected to the entire datasets. We selected the top 100 PCs and removed 1 PC with more than 0.7 correlation with the technical bias vector, defined as  $\log_2[\text{gene count}]$  for each cell. The  $k$ -NN algorithm identified 10 nearest neighbours to each of the single cells in the input data based on their distances computed using the selected 99 PCs. The inference of synchronized developmental age using the  $k$ -NN algorithm was iteratively run: in the first iteration, each cell was assigned a predicted age on the basis of the most common age among its ten neighbours. In subsequent iterations, the predicted age of each cell was assigned on the basis of the most common predicted age from the previous iteration of its ten neighbours. Ten iterations were run until convergence into the final synchronized ages (Fig. 1h and Extended Data Fig. 2a,d).

## Label transfer and clustering

For adult cells (P56), cell-type identities were assigned by mapping to the ABC-WMB Atlas<sup>14</sup> using the R package `scrattch.mapping` (v.0.55; <https://github.com/AllenInstitute/scrattch.mapping>)<sup>79</sup>. Mapped clusters were merged on the basis of DE criteria to define final cluster identities. For earlier developmental ages (P0–P28), cell types were assigned using Seurat<sup>80</sup> (RRID: SCR\_016341) with the reciprocal PCA (RPCA) label transfer approach (Extended Data Fig. 2b). Specifically, cell-type labels were transferred from the P56 reference to P20–P28, and from P20–P28 to P17–P19, and so forth. Clusters with fewer than ten cells in an age bin were reassigned to the nearest cluster using the  $k$ -NN approach ( $k = 10$ ). After cell-type assignment, additional iterative clustering was performed in each cluster and synchronized age bin to identify subclusters (Extended Data Fig. 2a).

For global clusters that were dominantly from the embryonic stage (E11.5–E18.5), we used `scrattch.mapping` to assign cell types on the basis of previously published data<sup>31</sup>, a mouse development study that covered E7–E18 using a list of 2,947 marker genes derived from the study's cluster-specific markers. Global clusters that were mapped to RG were assigned as a NEC subclass (dominated by cells from E11.5 to E12.5 and expressing *Hmga2*) or RG (dominated by cells from E13.5 to E16.5 and expressing *Sox2*, *Pax6*, *Hes1* and *Hes5*). Neurons born early at E11.5 and E12.5, characterized by enrichment in *Reln*, *Trp73* and *Calb2*, were classified as a CR Glut subclass. According to the trajectory analysis, clusters at E11.5 and E12.5 that were enriched in *Crabp2*, *Ebf2* and *Eomes*, and which give rise to CR Glut, were categorized as the IMN CR subclass. Global clusters mapped to neuronal IPs were identified as the IP class (with high expression of *Eomes*, *Btg2*, *Neurog2* and *Gadd45g*). IP clusters enriched in *Neurog1*, *Lhx9*, *Rmst* and *Nhlh1* were classified as the IP nonIT subclass, whereas those with higher levels of *Pou3f2*, *Lama2* and *Slco1c1* were classified as the IP IT subclass. Embryonic global clusters that were highly enriched in the neuronal markers *Dcx*, *Neurod1*, *Neurod2*, *Neurod6*, *Tubb3* and *Tbr1* were annotated as the IMN class. In the IMN class, clusters enriched in *Fezf2* were labelled as the IMN nonIT subclass, whereas those enriched in *Pou3f2* were labelled as the IMN IT subclass. We also defined the glioblast subclass (which expresses *Tnc*, *Fabp7*, *Qk*, *Lipg* and *Slco1c1*). Cells in each embryonic subclass were merged into one or a few clusters, followed by iterative clustering in each cluster and each synchronized age bin to identify subclusters. Finally, we merged the subclusters in each cluster that did not pass the DE gene criteria:  $q1.th = 0.4$ ,  $q.diff.th = 0.7$ ,  $de.score.th = 150$ ,  $min.cells = 10$ .

The final developmental cell-type taxonomy with annotations at the class, subclass, cluster and subcluster levels is summarized in Supplementary Table 3.

## Differential gene expression analysis and marker selection

We performed differential gene expression analysis both at the clustering step for each iteration and after clustering between all pairs of subclusters (or global clusters). We applied the limma package (implemented in `scratch.bigcat` package) to perform this analysis. For each pairwise comparison, we computed DE genes ( $\text{padj.th} = 0.01$ ,  $\text{q1.th} = 0.4$ ,  $\text{q.diff.th} = 0.7$ ,  $\text{de.score.th} = 150$ , and  $\log_2[\text{FC}] \geq 1.5$ ). We selected the top 15 DE genes in each direction and pooled such genes from all pairwise comparisons. All DE genes are shown in Supplementary Table 4.

## Reconstruction of the developmental trajectory

Popular computational methods such as Monocle<sup>81</sup>, PAGA<sup>82</sup>, Slingshot<sup>83</sup> and RNA Velocity<sup>84</sup> leverage the gradients in the transcriptomic space to infer a cell-type trajectory. However, a primary challenge that these tools face is the deconvolution of the temporal gradient from other gradients associated with cell-type heterogeneity. We found that these tools were not able to derive the trajectory of cortical development with the desired cell-type resolution. For example, the trajectories inferred by Monocle3 (ref. 40) switched back and forth between different layers and different ages for IT cells (Extended Data Fig. 2c), which made the results difficult to interpret.

Given the cell-type identities at the adult stage and the dense temporal sampling, we were able to progressively propagate cell-type identities between two adjacent ages (see above). As all cells in the later age evolve from cells in the earlier age, we found that identifying corresponding cell types in two adjacent time points that have only subtle transcriptomic differences could be readily solved using existing methods such as Seurat label transfer. Here instead of using actual age, we used synchronized age to derive trajectories for clusters with adult cell-type labels. This strategy worked well until earlier developmental time points, when cells develop more rapidly and cells in the same age can be present in different developmental states. During this period, when cells mainly belong to the RG, IP, IMN or glioblast classes, the transcriptomic gradient corresponding to differentiation is dominant, whereas cell-type diversity is much lower compared to later ages. We found that established methods such as Monocle3 worked well in this case. Therefore, we defined the embryonic trajectory using the same  $k$ -NN approach but using Monocle3-based pseudotime instead of synchronized age.

For postnatal stages, we connected each cluster in a synchronized age bin with its most likely antecedent in the previous bin using the  $k$ -NN approach (Extended Data Fig. 2b). Cells from two consecutive synchronized age bins were integrated using the Seurat RPCA approach. To assign edge weights between clusters from adjacent age bins, we applied a bootstrapping strategy: for cells of each cluster in the later age bin, we identified their 50 closest neighbour cells from the earlier age bin in the integrated latent space and then calculated the proportion belonging to each candidate antecedent cluster. Clusters with fewer than ten cells in a given age bin during label transfer were reassigned, and in such cases, we identified their nearest neighbours from the current and all preceding age bins in the global PCA space. We repeated these steps 100 times with a subsampling of 90% of cells, and median proportions were used as edge weights. Edge weights of  $>0.2$  were retained and shown in Supplementary Table 5, and we chose the edge with maximal weight for each cluster for the resulting trajectory (Supplementary Table 5).

For the embryonic stages, cells are changing substantially in the same age. We used the above strategy in pseudotime that was computed using Monocle3 (ref. 40) (Extended Data Fig. 2c). For cells in each cluster, we identified their 50 closest neighbour cells from clusters at an earlier median pseudotime using a bootstrapping strategy. Same as for the postnatal stages, edge weights of  $<0.2$  were removed. The developmental trajectory across the entire timeline from E11.5 to P56 is summarized in Supplementary Table 5 (Figs. 2a and 3).

## Comparison of Seurat RPCA and scVI integration

To assess the robustness and consistency of our developmental cell-type taxonomy, we compared the integration results from two different methods: Seurat RPCA and scVI (Supplementary Fig. 1). We applied both methods to integrate the scRNA-seq data from adjacent age bins (P0–P56) with age as a batch effect. For scVI, we used three sets of highly variable genes (HVGs; 1,000, 2,000 and 3,000 genes) between adjacent age bins. For Seurat RPCA, we performed data integration using the parameters  $\text{nfeatures} = 2,000$ ,  $\text{dims} = 1:30$ ,  $\text{k.anchor} = 50$  and  $\text{k.weight} = 100$ . For scVI, we tested three model architectures with varying complexity: default ( $\text{n\_hidden} = 128$ ,  $\text{n\_layer} = 1$ ,  $\text{n\_latent} = 10$ ); medium ( $\text{n\_hidden} = 128$ ,  $\text{n\_layer} = 2$ ,  $\text{n\_latent} = 32$ ); and large ( $\text{n\_hidden} = 256$ ,  $\text{n\_layer} = 3$ ,  $\text{n\_latent} = 32$ ). This resulted in a total of nine scVI model settings. Both Seurat RPCA and scVI produced highly consistent results, with small differences observed between the two methods. The integration facilitated alignment of the cell-type annotations across adjacent age bins, which further confirmed the robustness of the taxonomy. The slight variations among the nine scVI models were mainly attributed to the differences in model complexity and the number of HVGs, with more HVGs demonstrating a slightly higher resolution in capturing subtle transcriptional changes between cell types. Moreover, Seurat RPCA was able to capture rare cell types, for example, *Sst Chodl*. Both methods were able to reliably recover the majority of developmental transitions and cellular heterogeneity.

## Pseudotime

We computed the overall pseudotime in the global PCA space (Fig. 1i) using the entire developmental trajectory described above separately for the three independent trajectories: excitatory neurons and glia derived from NECs; MGE GABAergic neurons derived from MGE GABA RG; and CGE GABAergic neurons starting at CGE GABA. Each node of the trajectory was a cluster-by-synchronized-age-bin group of cells. The node centroid was defined as the median of all the corresponding cells in each PCA dimension. Pseudotime was computed iteratively over ten independent runs. In each iteration, one random cell in the starting node of each trajectory was set at pseudotime = 0. For each cell in the starting node, the pseudotime was the Euclidean distance to the randomly selected cell in the global PCA space. The Euclidean distance of the starting node centroid to the randomly selected cell was also computed. For each of the subsequent nodes, we computed the cumulative Euclidean distance to the starting cluster by summing the edge lengths along the trajectory, where each edge was defined as the distance between consecutive node centroids in the global PCA space. The pseudotime of each cell in each subsequent node was computed as the sum of the following parameters: (1) the distance from the cell to its closest antecedent node centroid; (2) the closest cumulative Euclidean distance of the antecedent node to the starting node; and (3) the distance from the starting node to the randomly selected cell. Finally, we took the mean of the pseudotime for each cell computed from the ten independent runs.

## Clustering temporal gene expression trajectories

To capture dynamic changes of each gene in each subclass over developmental time, we identified major temporal patterns using an unsupervised clustering approach. First, to identify cell-type-specific and temporal-specific marker genes, we computed pairwise DE genes between subclasses in each synchronized age and between synchronized ages in each subclass. We then computed the average expression of each marker gene identified above for each subclass-by-age group and normalized the values, with a maximum value of 1 for each gene across all subclass-by-age groups. To extend the analysis to prenatal time points, we also included cells from the antecedent clusters that give rise to the given postnatal subclass. Afterwards, we fit the normalized expression trajectory of each gene in each subclass using a generalized additive model with synchronized age as the predictor.

# Article

The model, implemented using the 'gam' function from the R package `mgcv`<sup>85</sup>, used cubic regression splines ( $bs = "cr"$ ), with a basis dimension of 12 ( $k = 12$ ) to capture gradual temporal changes without overfitting. To optimize smoothing estimates, we used the restricted maximum likelihood method, which provides robust and accurate performance in noisy and complex datasets. The fitted gene trends were hierarchically clustered using the R package `fastcluster`<sup>86</sup> with a Ward linkage and Euclidean distance metric. The hierarchical tree was cut at a height of 10 using the 'cutree' function, resulting in 394 clusters. The clusters were further merged using  $k$ -means, which resulted in 36 distinct gene trajectory patterns (Fig. 4).

## Assessing cell-type predictive power of adult marker genes across developmental ages

We performed fivefold cross-validation to classify subclasses at each individual postnatal age using different sets of marker genes derived from adult VIS subclasses: all 1,035 marker genes, 71 TF marker genes, 139 functional genes (including neuropeptides, GPCRs, ion channels and transporters) and 183 genes coding for adhesion molecules. We used the 'map\_cells\_knn\_big' function in the `scratth.bigcat` package, the same method implemented in `scratth.mapping`<sup>79</sup>, to perform the cross-validation.

## Quantification of DE genes along developmental trajectory

To quantify the developmental rate of each subclass, we computed the sum of log fold changes of DE genes ( $|\log_2 FC| > 1$ , Benjamini–Hochberg adjusted  $P < 0.05$ ) between each pair of adjacent synchronized ages, separately for DE genes with positive or negative changes (Fig. 5b). To control differences in sample sizes (that is, number of cells per age), we used a sliding window approach with a bandwidth of 2. Specifically, for a given comparison (for example, P1 versus P2), we compared combined samples from adjacent time points (for example, P0 + P1 versus P2 + P3). For each group in the comparison, we sampled up to 500 cells (for example, up to 1,000 cells for P0 + P1; Supplementary Table 3). To assess uncertainty and account for sampling variability, we randomly subsampled 70% of the selected cells and repeated the DE analysis 100 times. This enabled us to estimate the variability and mean in  $\log_2[FC]$  and provided more robust measures of developmental rate across subclasses.

## P0 MERFISH data generation

**Brain dissection and freezing.** Mice were transferred from the vivarium to the procedure room with efforts to minimize stress during transfer. Mice were anaesthetized with 0.5% isoflurane. Brains were rapidly dissected and selected on the basis of the absence of dissection damage. The selected brain was flash-frozen in OCT using 2-methylbutane chilled with liquid nitrogen and stored at  $-80^\circ\text{C}$ .

**Cryosectioning.** The freshly frozen brain was sectioned at  $14\ \mu\text{m}$  using Leica 3050 S cryostats. The OCT block was trimmed in the cryostat until the desired starting section was reached. Sections were collected every  $100\ \mu\text{m}$  to evenly sample the brain from posterior to anterior. Each section was mounted onto a functionalized 20-mm coverslip treated with yellow–green fluorescent microspheres (Vizgen, 2040003).

**Probe design.** A 500-gene panel was designed as previously described<sup>14</sup>. In brief, we chose the in-house P0 whole brain transcriptomic taxonomy (unpublished observations) as the reference and excluded any genes that had shown poor performance in previous MERFISH experiments. Starting with a default set of well-established marker genes curated from previous studies, we expanded the panel by selecting additional genes to ensure that there are at least two DE genes in both directions for each pair of clusters, using the 'select\_N\_markers' function from the `scratth.bigcat` package, and selected the top 350 genes (including the default genes) from this list.

To evaluate performance, we conducted 5-fold cross-validation using these 350 genes using `scratth.bigcat` package as described above. For clusters with cross-validation accuracy values below 0.7, we further refined the panel by selecting one additional DE gene in each direction using the same function. Our goal was to build a solid gene panel with strong predictive power at the subclass level and be opportunistic at resolving finer cell types. Except for the default gene set, the remaining genes were largely ordered with decreasing predictive power. We submitted a total of 729 genes to the Vizgen portal and selected the top 500 genes that passed the additional filters applied by Vizgen. The final gene set provided an overall cross-validation accuracy of 85.1% at cluster level and 98.0% at subclass level.

**Fixation and dehydration.** After air drying on the coverslips for 10–15 min, the tissue sections were loaded into a Leica Autostainer XL (Leica ST5010). They were washed in  $1\times$  PBS for 1 min, fixed in 4% paraformaldehyde for 15 min, washed in  $1\times$  PBS for 5 min 3 times, washed in 70% ethanol and then stored in 70% ethanol at  $4^\circ\text{C}$ . They were stored for at least 1 day and no more than 6 weeks before subsequent analyses.

**Hybridization.** For staining the tissue with MERFISH probes, a modified version of instructions provided by the manufacturer was used. All solutions were prepared according to the instructions provided by the manufacturer. For hybridization, samples were removed from the 70% ethanol and washed in a Petri dish containing Vizgen sample prep buffer (Vizgen, 20300001). Sample prep buffer was aspirated, and the samples were equilibrated with 5 ml Vizgen formamide wash buffer (Vizgen, 20300002) in a humidified incubator at  $37^\circ\text{C}$  for 30 min. Formamide wash buffer was removed by aspiration and a 50  $\mu\text{l}$  droplet of MERSCOPE Gene Panel Mix was added onto the centre of the tissue section. Next, the tissue section was covered with Parafilm and stored in a humidified  $37^\circ\text{C}$  cell culture incubator for 36 h.

**Gel embedding.** Parafilm covering the sections was removed, and 5 ml of the Vizgen formamide wash buffer was immediately added. Sections were incubated at  $47^\circ\text{C}$  for 30 min. Formamide wash buffer was aspirated and the previous step repeated. Sections were washed with Vizgen sample prep wash buffer after the second formamide wash for 2 min. Next, 110  $\mu\text{l}$  Vizgen gel embedding solution (Vizgen 20300004) with APS and TEMED was added onto the centre of a Gel Slick-coated XL microscope slide (Ted Pella, 260231) and any excess embedding solution was gently removed. To enable the gel to fully polymerize, the sections were incubated at room temperature for 1.5 h. To clear the tissue, the section was incubated in 5 ml Vizgen Clearing solution (Vizgen 20300003) with proteinase K (NEB P8107S) according to the manufacturer's instructions for at least 16–18 h in a humidified incubation oven at  $37^\circ\text{C}$ .

**Imaging.** Following clearing, sections were washed twice for 5 min in sample prep wash buffer (Vizgen, 20300001). Vizgen DAPI and PolyT stain (Vizgen, 20300021) was applied to each section for 15 min followed by a 10 min wash in formamide wash buffer. Formamide wash buffer was removed and replaced with sample prep wash buffer during MERSCOPE set up. Next, 100  $\mu\text{l}$  RNase inhibitor (New England Biolabs M0314L) was added to 250  $\mu\text{l}$  imaging buffer activator (Vizgen, 203000015) and this mixture was added through the cartridge activation port to a pre-thawed and mixed MERSCOPE Imaging cartridge (Vizgen, 1040004). Fifteen millilitres of mineral oil (Millipore-Sigma m5904-6X500ML) was added to the activation port, and the MERSCOPE fluidics system was primed according to Vizgen instructions. The flow chamber was assembled with the hybridized and cleared section coverslip according to Vizgen specifications, and the imaging session was initiated after collection of a  $10\times$  mosaic DAPI image and selection of the imaging area. For specimens that passed the minimum count threshold,

imaging was initiated, and processing was completed according to Vizgen's proprietary protocol.

**MERFISH data analysis.** Raw MERSCOPE data were decoded using Vizgen software (v.231). For cell segmentation, we used an in-house model to segment cells by applying the human-in-the-loop approach introduced in Cellpose (v.2.0)<sup>87</sup>. Starting with the 'cyto2' pretrained model in Cellpose, we trained our own model using the Cellpose GUI with the human-in-the-loop method, in which model predictions were iteratively corrected by the user and incorporated into training. Our model was trained on 24 two-channel images (200 × 200 μm) that represented a range of cellular densities, developmental stages and both sexes.

We used DAPI as the nuclear channel. For the cytoplasmic channel, we generated a post hoc stain from the measured mRNA transcripts. Transcripts were binned into a 2D histogram aligned with the DAPI channel, convolved with a Gaussian filter ( $\sigma_z = 1$ ,  $\sigma_x/y = 3$ ) and processed with a 3D median filter ( $z = 2$ ,  $x/y = 10$ ). This produced a stain-like signal that improved cytoplasmic boundary detection. Segmentation was performed in 3D using Cellpose's volumetric mode, which computes flows across the  $yx$ ,  $zx$  and  $zy$  planes and averages them before running 3D dynamics.

**MERFISH QC metrics.** To ensure high quality in our P0 MERFISH dataset, we retained cells that met the following criteria: at least 6 detected genes and at least 30 detected mRNA transcripts. We also used the percentage of blank barcodes per cell as a marker for low-quality cells. Blank barcodes do not encode for any gene targeted by the panel and thus represent a measure of false-positive detection. We excluded cells with a blank barcode percentage greater than 2%. Next, we applied the Solo algorithm to identify doublets<sup>88</sup> independently on each section. Solo outputs a probability score for each cell being a singlet or doublet. We computed the difference (dif) between singlet and doublet scores of the predicted doublets. To define a threshold for doublet classification, we calculated the 0.9 and 0.1 quantiles of the dif distribution among the predicted doublets. The threshold was set as  $q_{0.9}(\text{dif}) - q_{0.1}(\text{dif})$ , and all cells with dif values above this threshold were classified as doublets. These cells were excluded from the dataset.

### Spatial mapping of developmental VIS cell types at P0

To identify spatial distribution of cell types at P0, we used the unpublished P0 whole brain scRNA-seq taxonomy (P0 WB) as a reference to bridge the P0 whole brain MERFISH dataset. First, we mapped the MERFISH dataset to P0 WB using the `scrattch.mapping`<sup>79</sup> package based on the 500 gene panel. In parallel, cells from E18.5, P0 and P1 in this study (referred to as P0 VIS) were also mapped to P0 WB using a combined marker set from both datasets. E18.5 and P1 cells were included to increase cell numbers, as their transcriptomic differences from P0 were minor. We selected P0 WB clusters to which P0 VIS clusters were mapped and then extracted MERFISH cells mapped to those clusters with mapping scores of  $\geq 0.5$ . These MERFISH cells were then directly mapped to the P0 VIS reference. Owing to substantial transcriptomic heterogeneity in IMN types and a gradual continuum between IMN and IP populations, we first integrated selected MERFISH and P0 VIS reference cells labelled as IMN and IP using `scVI`<sup>34</sup> ( $n_{\text{hidden}} = 256$ ,  $n_{\text{layer}} = 3$ ,  $n_{\text{latent}} = 32$ ), and trained a random forest classifier on the scRNA-seq clusters. Then we performed flat mapping with bootstrapping, randomly sampling 80% of the genes in each of 100 iterations. This enabled subcluster-level mapping of MERFISH cells, which improved cell-type resolution and provided reliable confidence estimates essential for distinguishing closely related cell types.

### Identification of gene modules

Synchronized age-associated co-regulated genes specific to each subclass or class were determined using an unsupervised clustering approach. First, we computed pairwise DE genes ( $\text{padj.th} = 0.01$ ,

$q_{1.\text{th}} = 0.5$ ,  $q_{\text{diff.th}} = 0.7$ ,  $\text{de.score.th} = 150$ , and  $\log_2[\text{FC}] \geq 1.5$ ) between subclasses or classes in each synchronized age bin or across synchronized age bins in each subclass or class. We selected the top 15 DE genes in each direction and pooled such genes from all pairwise comparisons. We then computed the average expression of each DE gene among cells in each subcluster. We computed for each gene the  $k$ -NN ( $k = 5$ ) using cosine similarity metrics, then computed the Jaccard similarity graph based on the number of shared nearest neighbours between every pair of genes. The Louvain clustering algorithm (resolution = 2) based on the Jaccard graph was used to identify gene co-expression modules. All the gene modules are summarized in Supplementary Table 6.

### GO enrichment analysis

To relate various gene modules to known biological processes, we performed gene set enrichment analyses using the R package `clusterProfiler` 4.0 (RRID:SCR\_016884)<sup>89</sup> and `gProfiler` (RRID:SCR\_006809)<sup>90</sup>. The function 'gconvert' from `gProfiler2` (RRID:SCR\_018190) was used to convert gene identifiers to their Ensembl identifiers. The functions 'enrichGO' and 'simplify' from `clusterProfiler` were then used to enrich GO terms from all three GO databases (Molecular Function, Biological Process and Cellular Component). A Benjamini-Hochberg adjusted  $P$  value cut-off of 0.01 was used to determine significant GO terms.

### Integration of snMultiome and scRNA-seq datasets and label transfer

We performed global de novo clustering of the Multiome snRNA-seq dataset following nucleus-level QC (gene count  $\geq 1,000$ , TSS enrichment  $\geq 3$ ,  $n_{\text{frags}} \geq 1,000$  and doublet score  $\leq 0.3$ ). Clusters outside the cortex were removed on the basis of low expression of the dorsal forebrain markers *Foxg1*, *Emx1* and *Emx2* (Supplementary Table 2). To further identify and remove non-cortical clusters at early developmental ages, we applied the same MERFISH mapping strategy described above, using the P0 whole brain taxonomy as a bridge. Specifically, selected P0 MERFISH cells were mapped to P0 snMultiome clusters through shared P0 whole brain clusters, which enabled the identification and removal of additional non-cortical snMultiome clusters. For assigning identities of snMultiome nuclei, we mapped their transcriptomes to the scRNA-seq developmental cell-type taxonomy. We integrated scRNA-seq data (subsampling up to 200 cells per cluster) and Multiome snRNA-seq data (all nuclei) using `scVI`<sup>34</sup> ( $n_{\text{hidden}} = 512$ ,  $n_{\text{layer}} = 4$ ,  $n_{\text{latent}} = 50$ ) using a combined set of DE genes based on the scRNA-seq subclusters and snMultiome global clusters (Supplementary Table 4). In the integrated latent space, we applied a random forest classifier to predict each nucleus' most probable cell-type identity, using the scRNA-seq taxonomy as a reference. For this, we used the `RandomForestClassifier` implementation from the `sklearn.ensemble` module in Python, with default parameters except for  $n_{\text{estimators}} = 100$ . Last, we performed further annotation and QC of each predicted cluster and filtered out a small set of clusters deemed to be low quality or outside the cortex based on additional mapping to the adult ABC-WMB atlas, which resulted in a final set of 200,061 Multiome nuclei for further analysis. The final Multiome snRNA-seq to scRNA-seq developmental cell-type taxonomy mapping result is shown in Supplementary Table 8 (Figs. 1b and 6a-d).

### Multiome peak calling

To call chromatin accessibility peaks in the snATAC-seq data, we first categorized snMultiome cells (nuclei) according to both subclass and age group. To accumulate enough samples with sufficient statistical power for comparative analysis, we combined consecutive ages into the following groups: E13-E16.5, E17-E18.5, P0-P3, P4-P6, P7-P10, P11-P15 and P56, which are broadly consistent with the synchronized age bins (Extended Data Fig. 2d). We kept only the subclass-by-age groups with more than 50 nuclei. We generated pseudobulk replicates using the `ArchR`<sup>59</sup> function 'addGroupCoverages'. We created a reproducible

# Article

merged peak set using function 'addReproduciblePeakSet'. Finally, we built the peak-by-cell matrix, which contains insertion counts in the merged peak set, using function 'addPeakMatrix'.

## Identification of DA peaks

To identify DA peaks, as the peak presence in each cell is mostly binary, we chose the Chi-squared test to evaluate the significance of DA peaks across all 882,075 peaks identified above between every pair of subclass-by-age groups. In addition to the  $\log_2[\text{FC}]$  and adjusted  $P$  value ( $\text{adj.}P$ ) based on the Chi-squared test, we also computed the fraction of cells in each category with non-zero counts for each peak. To choose significant DA peaks, we required  $\log_2[\text{FC}] > 1$ ,  $\text{adj.}P < 0.05$  and fraction of cells with non-zero value in the foreground category to be  $> 0.05$ . This method was implemented in 'de\_all\_pairs' function in the `scratch.bigcat` package with extensive parallelization for efficiency. Because of the extensive diversity in cell types overall, we opted for pairwise comparisons instead of one-versus-all comparisons. This decision was made because the cell types in the background group exhibit high heterogeneity in one-versus-all comparison scenarios, which poses challenges in detecting subtle differences. The all-pairwise approach offers enhanced accuracy in identifying DA peaks across both similar and dissimilar pairs of cell categories.

## Identification of peak modules with similar cell-type and temporal specificity

To identify peaks that regulate different cell types at different developmental ages, we first extracted the DA peaks for each age group across different subclasses. We then pooled all the DA peaks identified between different subclasses across all age groups and clustered them to identify peak modules. To do so, we first computed the peak-by-category matrix as the average number of reads in each peak per subclass-by-age group, divided by the total number of reads across all peaks per subclass-by-age group, then multiplied by 30,000. The clustering was performed on peak-by-category matrix, subset to the DA peaks, using the Jaccard-Leiden clustering algorithm. We first computed for each peak the  $k$ -NN ( $k = 10$ ) using cosine similarity metrics, then computed the Jaccard similarity graph based on the number of shared nearest neighbours between every pair of peaks, and finally performed the Leiden clustering algorithm based on the Jaccard graph. In most cases, we used resolution index = 2. In cases when we observed more heterogeneity in the peak module, we increased the resolution index accordingly. This method is robust, efficient and scalable, and generates peak modules with high cell-type and temporal specificity.

## Identification of peak-gene pairs with matching accessibility and gene expression

We first extracted all the peak and gene pairs such that the gene is located within the 5 Mb window centred at the peak. Then we computed the correlation between the average peak accessibility and average gene expression based on the Multiome dataset across subclass-by-age groups. Given that a gene can be regulated by different peaks in different cell types and/or different ages, the correlation is computed only in different subsets in different contexts, for example, in IT subclasses only. We chose a minimal correlation of 0.5 to select such peak-gene pairs. Furthermore, we computed the average accessibility profile across subclass-by-age groups for all the peaks in each peak module. Subsequently, we calculated the correlation between the average expression in each subclass-by-age group of each gene and the peak module average profile described above. We then filtered and retained only those peak-gene pairs if the gene has the strongest correlation with the peak module corresponding to the respective peak. To accommodate space constraints, for each peak module, only the top 500 selected peak-gene pairs with the strongest peak-gene correlations were included for visualization (Supplementary Table 9).

## Differential motif analysis

We first scanned all the peak sequences using motif database with ArchR 'addMotifAnnotation' function, which produced a matrix that included the number of motif occurrences in each peak. We used JASPAR 2024 CORE non-redundant motif database (<https://jaspar.elixir.no/downloads/>), which enabled us to associate each motif to a corresponding TF. To perform differential motif analysis on peaks in different modules, we again used Chi-squared tests between all pairs of modules using 'de\_all\_pairs' function, using cutoff  $\log_2[\text{FC}] > 2$ ,  $\text{adj.}P < 0.05$ , and fraction of peaks with non-zero motif occurrences in the foreground of  $> 0.1$ . The absolute  $\log_2$  odds against random chance considering all the peaks should also be  $> 1$ . Once more, we conducted pairwise comparison across all peak modules, as we did not have sufficient previous knowledge of which peak modules might share common or distinct motifs. This strategy enabled us to identify enriched motifs in different combinations of peak modules. When multiple similar motifs were identified, we chose to report the ones associated with stronger regulators based on the GRN analysis described below.

## Inference of gene regulatory networks

Using SCENIC+ (ref. 60), we identified triplets consisting of a TF, a target peak and a target gene that met the following criteria: (1) TF expression should predict target gene expression; (2) the peak lies within 150 kb of the TSS of the target gene or in its gene body; (3) peak accessibility should predict target gene expression; (4) the peak contains the TF-binding motif; and (5) the TF motif should be enriched in the peak module to which the peak belongs. The choice of 150 kb to target the TSS was adopted from SCENIC+. We aimed to identify both activating and repressing TF interactions. For activation, we required that the TF positively correlates with both the accessibility of the target peak and the expression of the target gene and is expressed at the time of peak or gene activation in the same cell type. For repression, we required a negative correlation between the TF and both peak accessibility and gene expression, and to be expressed at or before the activation of the peak or gene, although not necessarily in the same cell type. As anticorrelation alone does not confirm repression, owing to potential motif matches by chance, or binding by other TFs from the same family, we further limited candidate repressors to TFs with known repressive functions from the literature. Each triplet was assigned a confidence score, and only the top-ranked predictions are presented (Supplementary Table 10).

Building on the SCENIC+ framework<sup>60</sup>, we integrated Multiome data (snRNA-seq and snATAC-seq) with motif analysis to identify regulatory relationships through triplet scores that link TFs, accessible chromatin peaks and target genes. We reimplemented the core concepts of SCENIC+ with modifications tailored to developmental datasets and to enable better integration with the ArchR pipeline. This approach enabled us to use a consistent pre-processed dataset while flexibly applying regulatory network analysis across distinct trajectories to capture cell-type specific TF-target interactions in context rather than assuming static relationships across the entire dataset.

The key steps of triplet score computation are outlined in Extended Data Fig. 13a. We began by selecting candidate TFs that are associated with differential motifs and exhibit significant differential expression in the trajectory of interest (using counts per million (CPM) as a measure of gene expression level, with maximum  $\log_2[\text{CPM}]$  difference of  $> 2.5$  across subclasses and age groups). Target genes were filtered to include only the top ten DE genes ( $|\log_2[\text{FC}]| > 2$ ) from all subclass and age group pairwise comparisons in the trajectory. For each target gene, triplets were computed by scoring all relevant TFs and their associated accessible peaks, which resulted in a final score that reflects the confidence of each TF-peak-gene interaction.

Following the SCENIC+ approach, we inferred TF-to-gene relationships by combining XGBoost-based prediction of TF influence on gene expression. We retained interactions with importance gain of >0.01 and absolute correlation of >0.05. Gene-gene correlations were computed at the single-cell level using imputed values averaged from the 20 nearest neighbours in the scVI latent space, which provides greater robustness than raw expression values. However, for the XGBoost model, we used raw  $\log_2$ [CPM] data, as such data better highlight distinct marker genes and produce more interpretable results.

To infer peak-to-gene relationships, we also used the XGBoost model to predict target gene expression using all the peaks within 150 kb of the TSS or in the gene body. Owing to sparsity of the snATAC-seq data, imputed values were used to fit the model. We filtered the interactions using the same importance and correlation criteria described above. We filtered the peaks based on the occurrence of TF motif and enrichment of the TF motif in the corresponding peak module. We also computed a TF and peak correlation based on imputed TF expression and peak accessibility. The final score was computed as follows: Triplet score =  $\text{abs}(\text{TF target gain} \times \text{TF target correlation} \times \text{peak target gain} \times \text{peak target correlation} \times \text{TF peak correlation})$ .

Finally, we considered the timing of TF and target gene expression. To infer a regulatory relationship, the TF must be expressed when the target is activated. We focused on the subclass where the target reaches peak expression and confirmed that the TF is already active in that context. For repressive interactions, for which TF and target expression are anticorrelated, we still required the TF to be expressed before target activation, although not necessarily in the same subclass.

Triplets with scores  $<10^{-5}$  were filtered out. For activating interactions, we required a TF-target correlation of >0.1, a peak-target correlation of >0.05 and a TF-peak correlation of >0.05. For repressive interactions, we required a TF-target correlation of  $<-0.3$ , a peak-target correlation of >0.05 and a TF-peak correlation of  $<-0.1$ . Activation and repression scores were aggregated across peaks to produce a single interaction score per TF-target pair, and all TF-target pairs were normalized between 0 and 1. Each gene's network weight was calculated as the sum of its interaction scores, and only nodes with weights > 1 were visualized. Top TFs in each family were selected on the basis of the node weights. These empirical thresholds were chosen to emphasize strong interactions and to reduce network complexity, but they may need to be re-examined in future work.

## Reporting summary

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

## Data availability

Primary data are being made available through BRAIN Initiative Cell Atlas Network (BICAN, RRID: SCR\_022794; <https://www.portal.brain-bican.org/>) and the Neuroscience Multi-omic Data Archive (NeMO, RRID: SCR\_016152; <https://nemoarchive.org/>). The following resources are also available from NeMO: the identifier containing links to all primary data (<https://assets.nemoarchive.org/dat-5y9mf0h>); the 10x scRNA-seq dataset (<https://assets.nemoarchive.org/dat-0oyried>); the Multiome snRNA-seq dataset (<https://assets.nemoarchive.org/dat-bbqchpq>); and the Multiome snATAC-seq dataset (<https://assets.nemoarchive.org/dat-5ke3d8i>). The following resources are also available online: the processed 10x scRNA-seq dataset ([https://allen-developmental-mouse-atlas.s3.amazonaws.com/scRNA/DevVIS\\_scRNA\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.amazonaws.com/scRNA/DevVIS_scRNA_processed.h5ad)); the processed Multiome snRNA-seq dataset ([https://allen-developmental-mouse-atlas.s3.us-west-2.amazonaws.com/Multiome/DevVIS\\_multiome\\_snRNA\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.us-west-2.amazonaws.com/Multiome/DevVIS_multiome_snRNA_processed.h5ad)); and the processed Multiome snATAC-seq dataset ([https://allen-developmental-mouse-atlas.s3.amazonaws.com/Multiome/DevVIS\\_multiome\\_snATAC\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.amazonaws.com/Multiome/DevVIS_multiome_snATAC_processed.h5ad)).

## Code availability

The data analysis code used in the study is available from GitHub (<https://github.com/AllenInstitute/scrattch.bigcat> and <https://github.com/AllenInstitute/MouseDevVIS>).

- Allen Institute for Brain Science. Mouse whole cell tissue processing for 10x Genomics Platform v2. *protocols.io* <https://doi.org/10.17504/protocols.io.q26g7b52klwz/v9> (2022).
- Allen Institute for Brain Science. FACS single cell Sorting v4. *protocols.io* <https://doi.org/10.17504/protocols.io.be4cjgsw> (2020).
- Allen Institute. HEPES-sucrose cutting solution v2. *protocols.io* <https://doi.org/10.17504/protocols.io.5jyl8peq8g2w/v2> (2023).
- Allen Institute. Mouse brain perfusion and flash freezing v2. *protocols.io* <https://doi.org/10.17504/protocols.io.j8nlkodr6v5r/v2> (2023).
- Drokhlyansky, E. et al. The human and mouse enteric nervous system at single-cell resolution. *Cell* **182**, 1606–1622 (2020).
- Allen Institute. RAISiNs (RNA-seq for profiling intact nuclei with ribosome-bound mRNA) nuclei isolation from mouse CNS tissue protocol v1. *protocols.io* <https://doi.org/10.17504/protocols.io.4r3l22n5pl1y/v1> (2023).
- Allen Institute for Brain Science. 10xV3 Genomics sample processing protocol v2. *protocols.io* <https://doi.org/10.17504/protocols.io.bq7cmziw> (2021).
- Allen Institute for Brain Science. 10x Multiome sample processing v2. *protocols.io* <https://doi.org/10.17504/protocols.io.bp2l6l1mqrvq/v2> (2022).
- Johansen, N., Miller, J., Lee, C. & ikapen-alleninst. AllenInstitute/scrattch.mapping: v0.55. *Zenodo* <https://doi.org/10.5281/zenodo.10939013> (2024).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
- Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
- Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
- La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
- Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).
- Müller, D. fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53**, 1–18 (2013).
- Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
- Bernstein, N. J. et al. Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101 (2020).
- Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
- Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
- Uchiyama, Y. et al. Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme. *Proc. Natl Acad. Sci. USA* **107**, 9240–9245 (2010).
- Ng, S.-Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
- Britanova, O. et al. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron* **57**, 378–392 (2008).
- Thompson, C. L. et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* **83**, 309–323 (2014).
- Kim, E. J., Juavinett, A. L., Kyubwa, E. M., Jacobs, M. W. & Callaway, E. M. Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron* **88**, 1253–1267 (2015).
- Tran, L. N., Loew, S. K. & Franco, S. J. Notch signaling plays a dual role in regulating the neuron-to-oligodendrocyte switch in the developing dorsal forebrain. *J. Neurosci.* **43**, 6854–6871 (2023).
- Marin, O., Anderson, S. A. & Rubenstein, J. L. Origin and molecular specification of striatal interneurons. *J. Neurosci.* **20**, 6063–6076 (2000).
- Fragkouli, A., van Wijk, N. V., Lopes, R., Kessar, N. & Pachnis, V. LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons. *Development* **136**, 3841–3851 (2009).
- Ross, S. E. et al. Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron* **73**, 292–303 (2012).
- Goodall, J. et al. Brn-2 represses microphthalmia-associated transcription factor expression and marks a distinct subpopulation of microphthalmia-associated transcription factor-negative melanoma cells. *Cancer Res.* **68**, 7788–7794 (2008).
- Ellmann, L., Joshi, M. B., Resink, T. J., Bosserhoff, A. K. & Kuphal, S. BRN2 is a transcriptional repressor of CDH13 (T-cadherin) in melanoma cells. *Lab. Invest.* **92**, 1788–1800 (2012).
- Yang, J. et al. SOX4-mediated repression of specific tRNAs inhibits proliferation of human glioblastoma cells. *Proc. Natl Acad. Sci. USA* **117**, 5782–5790 (2020).
- Sarkar, D. et al. Adult brain neurons require continual expression of the schizophrenia-risk gene *Tcf4* for structural and functional integrity. *Transl. Psychiatry* **11**, 494 (2021).
- Pai, E. L.-L. et al. Mafb and c-Maf have prenatal compensatory and postnatal antagonistic roles in cortical interneuron fate and function. *Cell Rep.* **26**, 1157–1173 (2019).
- Huang, Y.-H., Jankowski, A., Cheah, K. S. E., Prabhakar, S. & Jauch, R. SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. *Sci. Rep.* **5**, 10398 (2015).
- Stolt, C. C. et al. SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. *Dev. Cell* **11**, 697–709 (2006).

# Article

**Acknowledgements** We are grateful to staff at the Transgenic Colony Management, the Lab Animal Services and the Molecular Biology, Histology and Data and Technology departments at the Allen Institute for technical support; D. J. Di Bella for advice on marker genes for apical progenitors, IPs and IMNs; and the Allen Institute founder P. G. Allen for his vision, encouragement and support. The research was funded by U19MH114830 and U01MH130962 grants from National Institute of Mental Health to H.Z. under the BRAIN Initiative of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH and its subsidiary institutes. This work was also supported by AIBS.

**Author contributions** Conceptualization: H.Z. Data analysis lead and coordination: Y.G. and Z.Y. Data generation (scRNA-seq and Multiome): C.T.J.vV., E.D.T., D.B., T. Cardenas, D.C., T. Casper, M. Chiang, M. Clark, M.J.D., R.F., J. Gloe, N.G., J. Guzman, C.R.H., D.H., W.H., K.J., R. McCue, E.M., A.C.M., T.N.N., N.P., Q.R., N.V.S., J.S., A. Torkelson, A. Tran, H.T., K.R., B.L., N.D., K.A.S., Z.Y. and H.Z. Data processing and analyses (scRNA-seq and Multiome): Y.G., C.T.J.vV., C.L., A.B.C., R.C., J. Goldy, B.N., J. Wang, M.J.H., K.A.S., Z.Y. and H.Z. Data generation (PO MERFISH): A.P.A., S.B., J.C., S.D.H., Z.J., N.M., J.S.N., P.O., A.A.O., A.R., N.V.C., J.A. and D.A.M.M. Data analysis

(PO MERFISH): Y.G., R. Mathieu, L.C., J.Q. and M.K. Project management: C.P. and K.A.S. Management and supervision: C.T.J.vV., S.D.H., J.A., D.A.M.M., J. Waters, M.K., K.R., B.L., M.J.H., N.D., K.A.S., B.T., Z.Y. and H.Z. Manuscript writing and figure generation: Y.G., C.T.J.vV., Z.Y. and H.Z. Manuscript review and editing: Y.G., C.T.J.vV., R. Mathieu, J. Waters, M.K., M.J.H., B.T., Z.Y. and H.Z.

**Competing interests** H.Z. is on the scientific advisory board of MapLight Therapeutics. The other authors declare no competing interests.

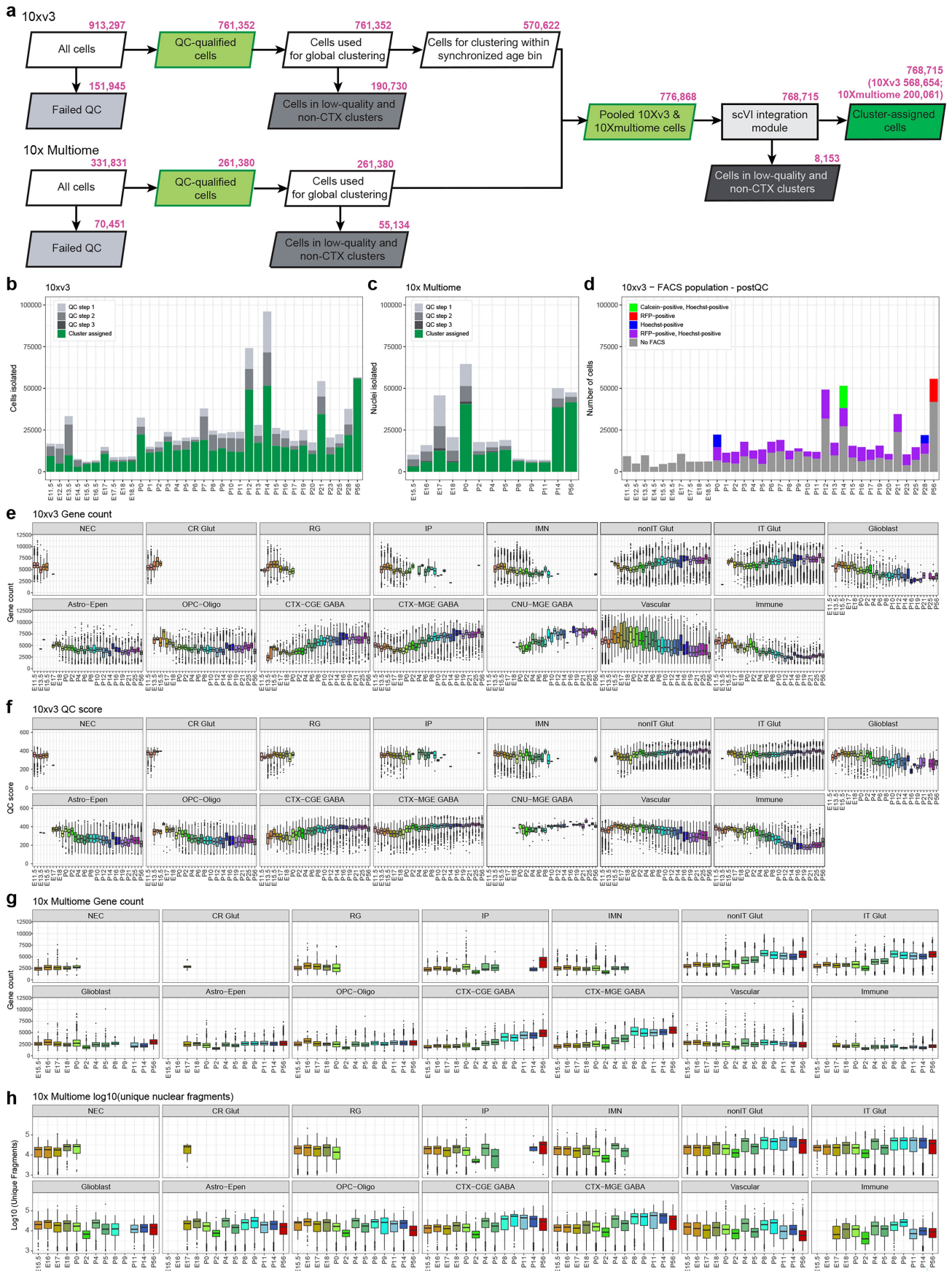
**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09644-1>.

**Correspondence and requests for materials** should be addressed to Zizhen Yao or Hongkui Zeng.

**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



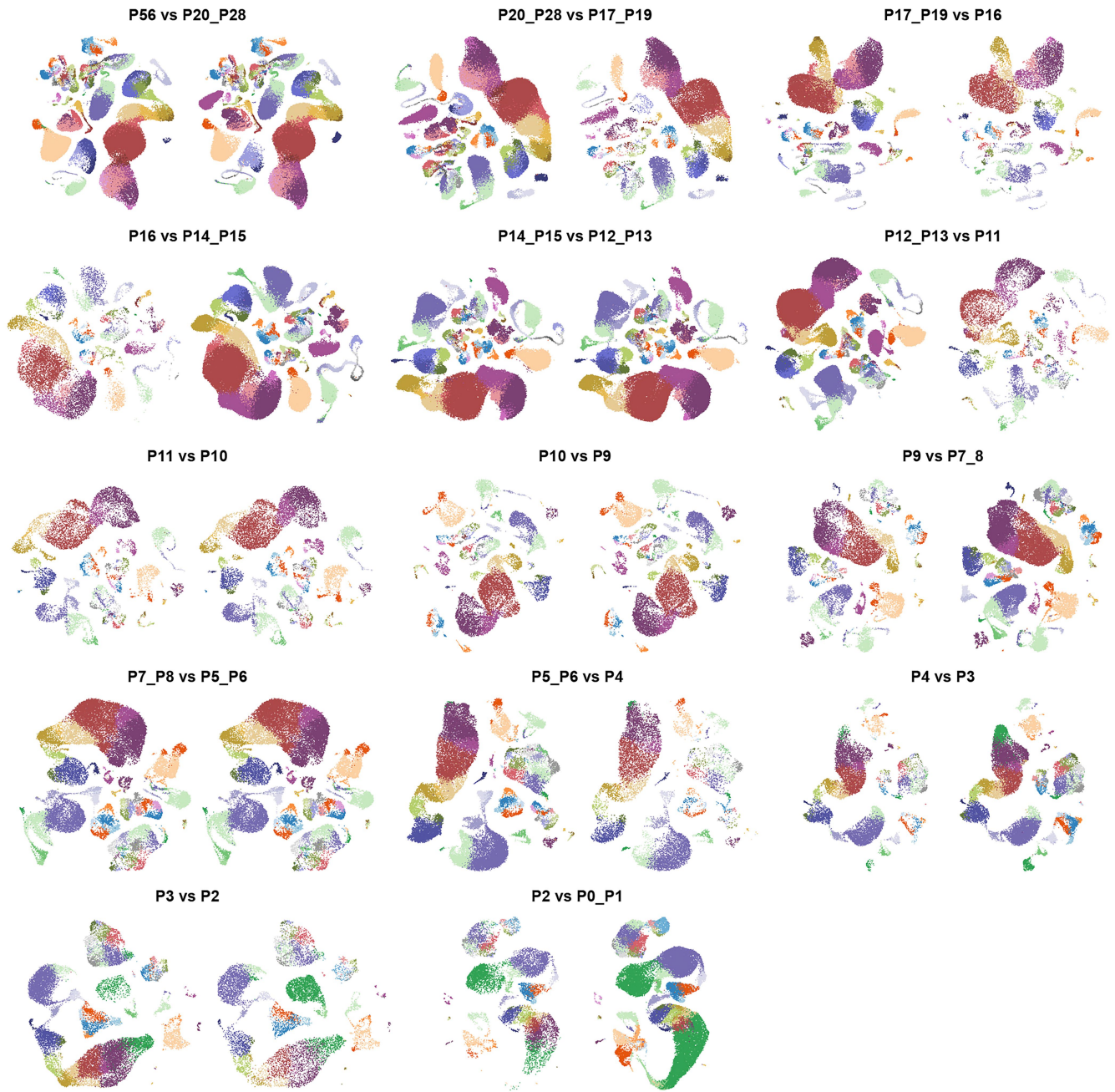
Extended Data Fig. 1 | See next page for caption.

# Article

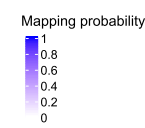
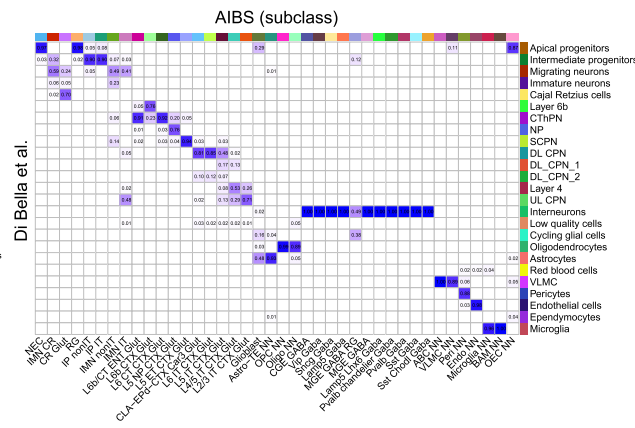
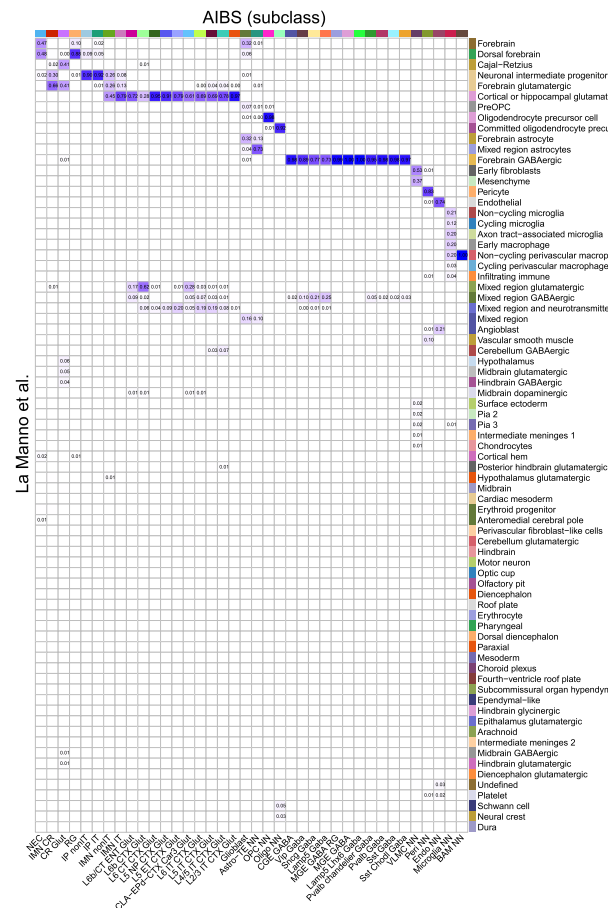
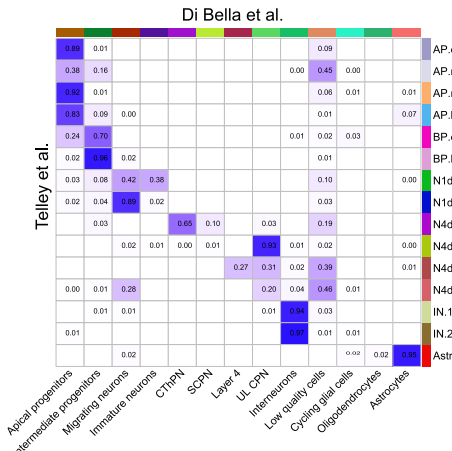
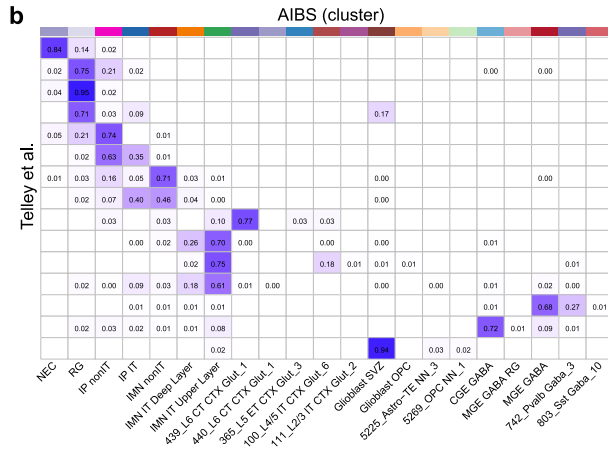
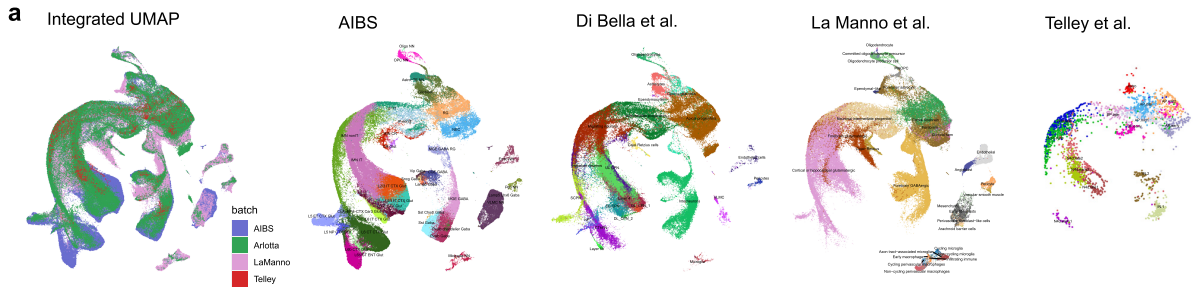
**Extended Data Fig. 1 | scRNA-seq and Multiome data processing and analysis workflow and quality control.** (a) Number of cells at each step in the scRNA-seq and snMultiome data processing and analysis pipeline. The identification of doublets and low-quality cells and clusters is described in detail in Methods. The 10xv3 and 10x Multiome data were first QC-ed and analyzed separately. After initial clustering the datasets were combined and QC-ed again before and after integration. (b-c) Number of cells after each QC step in scRNA-seq (b) and snMultiome data (c). The color codes of QC steps correspond to the colored QC boxes in (a). (d) Number of cells from each FACS

population in scRNA-seq data. (e-h) Box plots of gene detection (e) and QC score (f) for 10xv3, and gene detection (g) and number of unique fragments (h) for 10x Multiome, per cell across different cell classes and ages. In the box plots, the central line indicates the median value (gene count, QC score, or number of unique fragments per cell), the box spans the interquartile range (IQR; 25th–75th percentile), and the whiskers extend to  $1.5 \times$  IQR, with outliers plotted individually. The number of cells for 10xv3 in each subclass each age is shown in Supplementary Table 3, and the detail for 10x Multiome is shown in Supplementary Table 8.





**Extended Data Fig. 3 | Integration between adjacent age bins for label transfer.** UMAP comparison of each synchronized age bin with its adjacent younger age bin after integration and label transfer, showing common clusters.



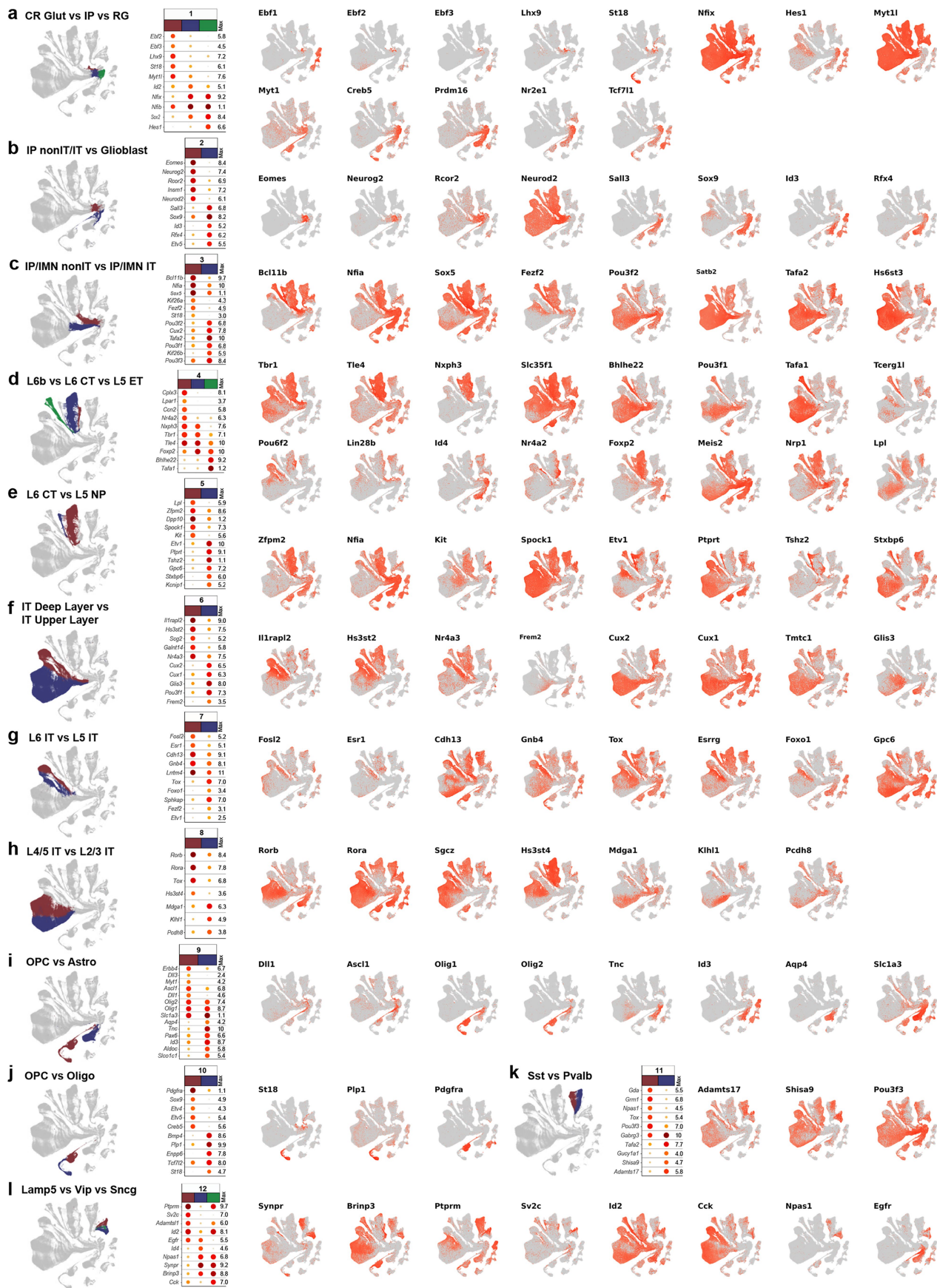
Extended Data Fig. 4 | See next page for caption.

# Article

## Extended Data Fig. 4 | Integration between AIBS data and external datasets.

**(a)** UMAPs showing the integrated embedding of AIBS developmental VIS scRNA-seq data and external datasets using scVI followed by label transfer using a Random Forest classifier. Subsets of the leftmost integrated UMAP are shown on the right, with each panel highlighting one dataset and colored by its original cell-type annotations: AIBS developmental VIS scRNA-seq (E11.5 to P4; colored and labeled at subclass level), Di Bella et al. (E10 to P4), La Manno et al. (E7 to E18; cells outside the cortex are excluded from the integrated UMAP), and Telley et al. **(b)** Corresponding matrices between AIBS VIS developmental taxonomy and external datasets. Values in the matrices denote mapping probabilities. An empty value in the matrices represents a mapping probability of zero or less than 0.01. For instance, the Midbrain subclass from La Manno's data appears in the matrix but with all empty values, indicating that at least one cell has a probability of less than 0.01. If a subclass from La Manno's data does not appear in the matrix, it means no cell types were mapped to it. Specifically, in comparison with Di Bella et al., our NEC and RG subclasses were mapped to Apical progenitors, while IP nonIT and IP IT corresponded to Intermediate

progenitors. The IMN nonIT subclass aligns with Migrating and Immature neurons, and IMN IT aligns with Migrating neurons and upper-layer Callosal Projection Neurons (UL CPN). Our excitatory subclasses, L6b, L6 CT, L5 NP, and L5 ET, show strong correspondence with Layer 6b, CThPN, NP, and SCPN, respectively. The CLA-EPd-CTX Car3 Glut, L6 IT, and L5 IT subclasses align with Deep Layer CPN (DL CPN), while L4/5 IT and L2/3 IT correspond to Layer 4 and UL CPN. At the cluster level, comparison with Telley et al. reveals that our NEC cells correspond to their AP.early cluster, while our RG cells align with AP.mid1, AP.mid2, and AP.late. The BP.early primarily corresponds to RG and IP nonIT populations, while BP.late mainly aligns with our IP nonIT and IP IT clusters. The N1d.early cluster was mapped to our IMN nonIT cluster, whereas the N1d.late cluster aligns partially with IMN nonIT (46%) and IP IT (40%). Notably, N4d.early1 mainly corresponds to our L6 CT cluster, whereas N4d.early2, N4d.late1, and N4d.late2 show strong alignment with our IMN IT Upper Layer cluster. We also compared the datasets from Di Bella et al. and Telley et al., and the results are consistent with the three-way comparisons between our data and the two published studies.



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

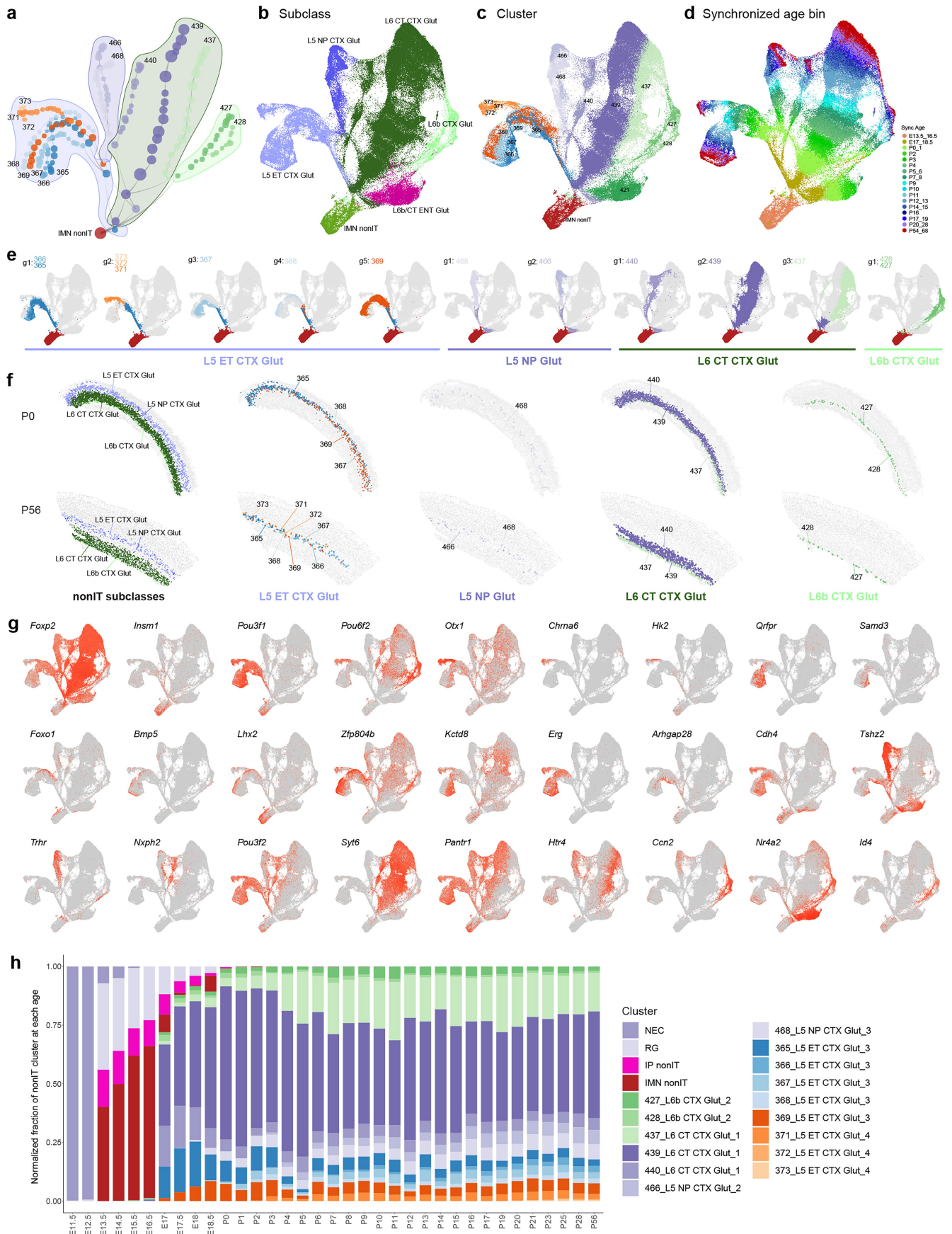
# Article

## Extended Data Fig. 5 | Expression of branching marker genes on UMAP.

(a-l) UMAP representations of major branching nodes shown in Fig. 2a and dot plots showing marker gene expression in each descendant branch of each branching node. Dot size and color indicate proportion of expressing cells and average expression level of a marker gene in each subclass, respectively.

Expression of marker genes at each branching node corresponding to Fig. 2a. As an example, while *Kif26a* is expressed in RG and *Kif26b* in IP cells, their expressions are transiently turned off before being turned on again specifically in IMN nonIT or IMN IT, respectively (Fig. 2c), and both genes are downregulated again in adulthood. These two closely related paralogs in the kinesin family

have mirrored temporal progression in distinct lineages, and *Kif26b* is known to play an important role in regulating adhesion of the embryonic kidney mesenchyme<sup>91</sup>. *Rmst* is a long non-coding RNA, previously reported to interact with *Sox2* to regulate neurogenesis<sup>92</sup>. Our results suggest that it might do so in a time and state dependent manner and is likely involved in nonIT fate specification. Well-known upper-layer regulators<sup>93</sup> *Satb2* and *Cux2* show modest enrichment in IT at IP stage and stronger enrichment at IMN stage. *Cux2* expression is further restricted to upper layer IT neurons at early postnatal stage, while the expression difference of *Satb2* between IT and nonIT gradually decreases at late postnatal stages.

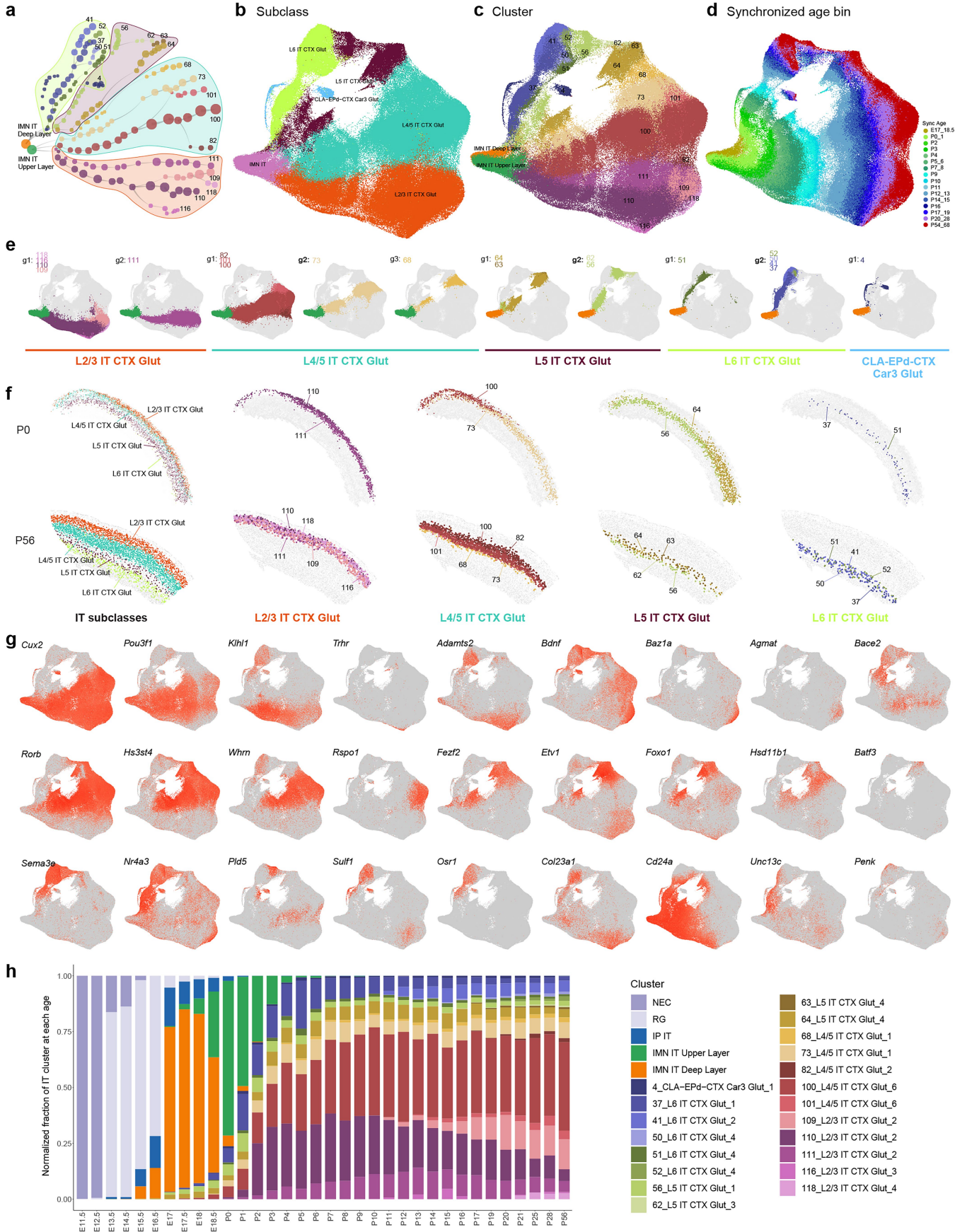


Extended Data Fig. 6 | See next page for caption.

# Article

**Extended Data Fig. 6 | Developmental trajectories of visual cortex nonIT Glut cell types.** (a) Transcriptomic trajectory tree for nonIT clusters starting from the common IMN nonIT antecedent. Nodes are clusters subdivided by synchronized age bins, and edges represent antecedent-descendant relationship between adjacent nodes, with thicker end at the antecedent node and thinner end at the descendant node. Nodes are grouped by subclass, and adult clusters are labeled. Nodes from L6b/CT ENT subclass are not included. (b-d) UMAPs for nonIT cells colored by subclass (b), cluster (c) and synchronized age bin (d). (e) Clusters are grouped together based on similar trajectories. Within each cluster group, all cells along their trajectories, including all antecedent nodes, are shown and are colored by cluster membership. (f) Spatial distribution of nonIT subclasses and clusters within each subclass in P0 and P56 MERFISH data, based on the ABC-WMB Atlas<sup>14</sup>. (g) Marker genes illustrating cell type diversification along trajectories. (h) Cluster composition of all nonIT cells at each age. There is a distinct population of L6b like cells with shared expression of subplate markers *Cplx3*, *Lpar1*, *Nr4a2*, but not *Ccn2*, *Nxph4* and *Pappa2*. This population is more abundant than L6b at E17-P3 (Fig. 2b), with expression of *Cyp26b1* and *Cobll1*, and mapped to adult L6b/CT ENT subclass. Based on Allen Developing Brain Atlas<sup>94</sup>, *Cyp26b1* is expressed specifically in the entorhinal cortex at E18.5, and our MERFISH data confirms the localization of L6b/CT ENT neurons in entorhinal cortex at P0 and P56 (Fig. 2e). For the L5 ET subclass, clusters 371–373 (*Chrna6*) represent the most distinct subset<sup>10,12,18</sup>, emerging at P3 with specific expression of TFs *Pou6f2* and *Otx1*. Expression of marker gene *Chrna6* begins relatively late, around P9, and peaks in adulthood. Clusters 372 and 373 diverge from 371 after P21, with 373

specifically expressing *Hk2*. Based on our trajectory analysis, *Chrna6*<sup>+</sup> clusters 371–373 share a common origin with clusters 365 and 366, with shared expression of *Kctd8*. We have identified multiple TFs potentially involved in regulation of different L5 ET clusters, including *Foxo1*, *Bmp5*, *Lhx2*, *Zfp804b* and *Erg*. There is no apparent spatial segregation of different L5 ET clusters in visual cortex at P56, while cluster 369 shows enrichment ventrally at P0. The L5 NP subclass contains two clusters, 466 and 468, which are diverged around P3, with *Su2c* and *Nxph2* enriched in each cluster respectively. *Nxph2*<sup>+</sup> cluster 466 appears to be slightly deeper than cluster 468 at P56, while only cluster 468 is present at P0. Unlike most other subclasses of cortical glutamatergic neurons, L5 NP cells do not have long-range projections, and their functions remain elusive<sup>10,95</sup>. The L6 CT subclass has three major clusters, 440, 439 and 437, diverging at E17. Interestingly, *Nxph2*<sup>+</sup> L6 CT cluster 440 is very distinct from the other L6 CT clusters but more related to L5 NP subclass based on trajectory analysis, with shared expression of TF gene *Pou3f2* with L5 ET and L5 NP. The separation between L6 CT clusters 437 and 439 (the dominant L6 CT cluster) is quite subtle transcriptomically, marked by enrichment of *Pantr1* and *Htr4*, respectively, but very distinct spatially: cluster 437 is clearly deeper than 439 at both P56 and P0, and is co-localized with L6b cells. *Pantr1*, a noncoding RNA gene adjacent to TF gene *Pou3f3*, is absent in the deep L6 CT cluster 437 and L6b but present in all other more superficial nonIT clusters. In L6b subclass, two major clusters 427 and 428 diverge around E17, with TFs *Foxp2*, *Nr4a2* and *Id4* enriched in 427 and *Tox* enriched in 428. There is no apparent difference in spatial distribution of these two clusters, but 427 is more closely related to L6 CT subclass transcriptomically.

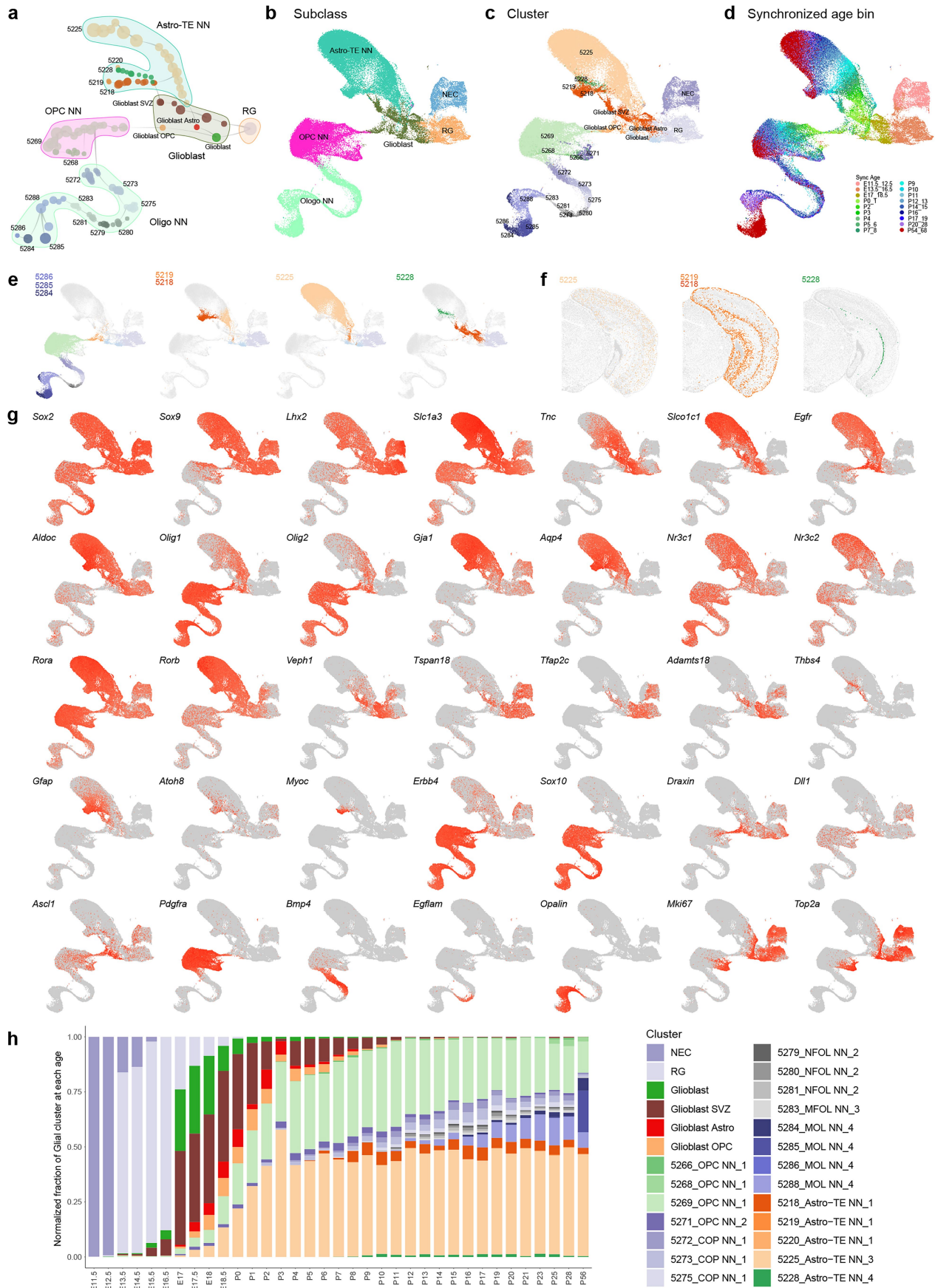


Extended Data Fig. 7 | See next page for caption.

# Article

**Extended Data Fig. 7 | Developmental trajectories of visual cortex IT Glut cell types.** (a) Transcriptomic trajectory tree for IT clusters starting from the common IMN IT antecedents. Nodes are clusters subdivided by synchronized age bins, and edges represent antecedent-descendant relationship between adjacent nodes, with thicker end at the antecedent node and thinner end at the descendant node. Nodes are grouped by subclass, and adult clusters are labeled. (b-d) UMAPs for IT cells colored by subclass (b), cluster (c) and synchronized age bin (d). (e) Clusters are grouped together based on similar trajectories. Within each cluster group, all cells along their trajectories, including all antecedent nodes, are shown and are colored by cluster membership. (f) Spatial distribution of IT subclasses and clusters within each subclass in P0 and P56 MERFISH data, based on the ABC-WMB Atlas<sup>14</sup>. (g) Marker genes illustrating cell type diversification along trajectories. (h) Cluster composition of all IT cells at each age. In the IMN IT subclass, *Frem2* is enriched in the late upper layer IMN cluster, with this enrichment persisting until P10. In the IT trajectory, many clusters that split off early have distinct layer distribution. In the L5 IT subclass, clusters 64 and 56 diverge around E17, and 64 is more

superficial than 56 at P56. In the L4/5 IT subclass, clusters 100 and 73 diverge at E18.5, with 100 being more superficial than 73 at P56. In the L2/3 IT subclass, clusters 110 and 111 separate around P1, with 110 located more superficially than 111 at both P0 and P56. More clusters arise in later stages of development after eye opening, and these newer clusters usually have less distinct spatial distribution from sibling clusters. For example, L2/3 IT cluster 109 diverges from 110 at -P11 with increased expression of *Bdnf* and decreased expression of *Adamts2*, while cluster 118 further diverges from cluster 109 at P21 with increased expression of *Baz1a* and *Tnfrsf6*. Spatially within L2/3, clusters 118 and 110 are located more superficially than 109 at P56. We also observe late divergence of L4/5 IT clusters 101 and 82 from cluster 100 at P14 and P20 respectively, which display subtle differences in spatial distribution, with cluster 82 located more superficially than 100 while 101 located deeper than 100. There are also new cell types emerging for L5 IT and L6 IT subclasses, with L6 IT clusters 41 and 50 emerging from 37 around P11, cluster 52 emerging from 41 around P20, and L5 IT clusters 62 and 63 emerging from 56 and 64, respectively, around P19-P21.

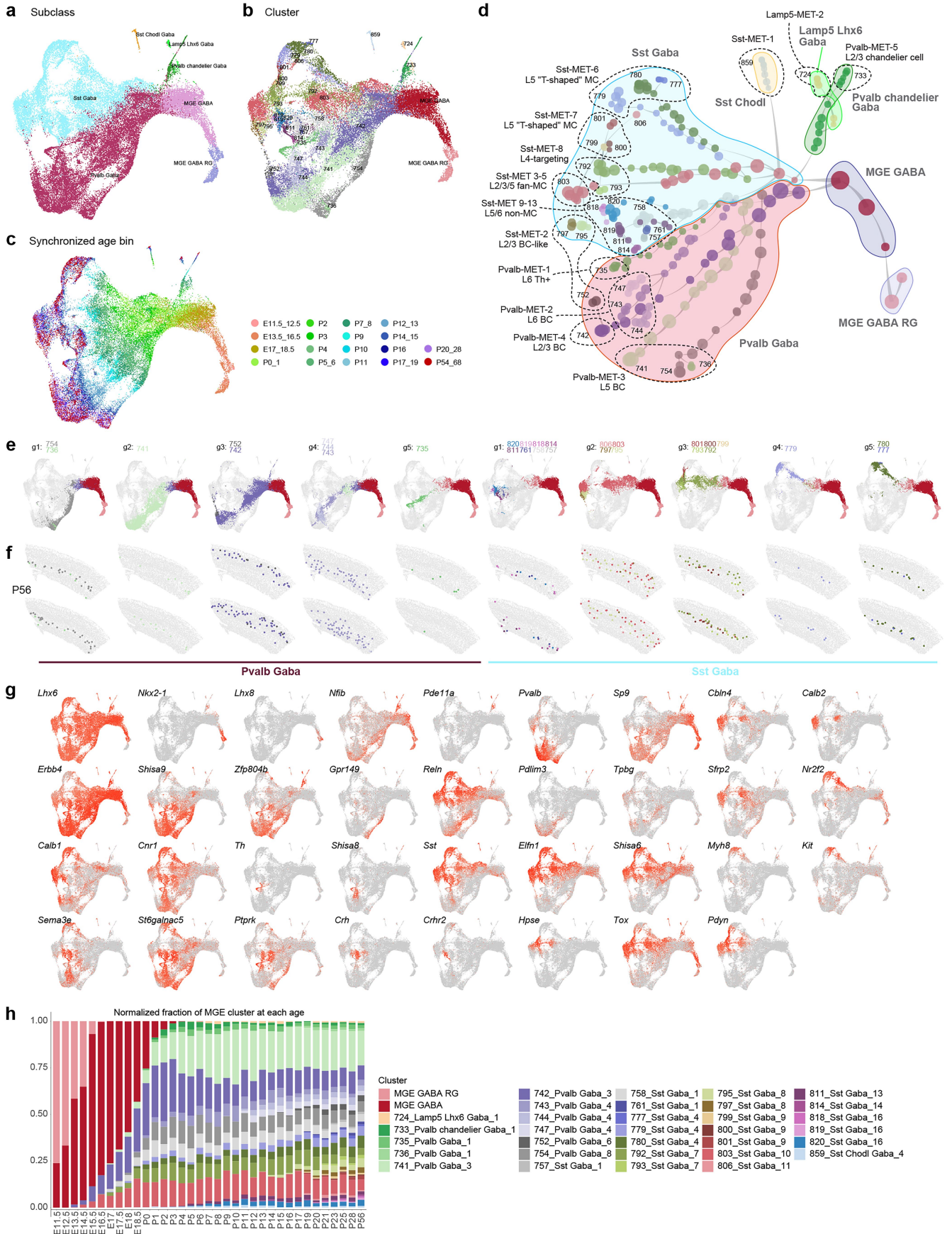


Extended Data Fig. 8 | See next page for caption.

# Article

**Extended Data Fig. 8 | Developmental trajectories of visual cortex Glia cell types.** **(a)** Transcriptomic trajectory tree for glia clusters starting from the common RG antecedent. Nodes are clusters subdivided by synchronized age bins, and edges represent antecedent-descendant relationship between adjacent nodes, with thicker end at the antecedent node and thinner end at the descendant node. Nodes are grouped by subclass, and adult clusters are labeled. **(b-d)** UMAPs for glial cells colored by subclass (b), cluster (c) and synchronized age bin (d). **(e)** Clusters are grouped together based on similar trajectories. Within each cluster group, all cells along their trajectories, including all antecedent nodes, are shown and are colored by cluster membership. **(f)** Spatial distribution of astrocyte clusters in P56 MERFISH

data, based on the ABC-WMB Atlas<sup>14</sup>. **(g)** Marker genes illustrating cell type diversification along trajectories. **(h)** Cluster composition of all glial cells at each age. Specifically, Notch ligands *Dll1*, *Dll3* and *Ascl1* are expressed transiently and downregulated as the cells transition from glioblasts to OPCs, while *ErbB4* maintains its expression. It has recently been shown that Notch signaling plays a dual role in both promoting and inhibiting oligodendrogenesis to fine-tune regulation of oligodendrocyte generation<sup>96</sup>. *Sox9*, strongly expressed in RG and glioblasts, is downregulated in OPCs and turned off completely after cells exit the OPC stage; in contrast, *Sox10* is activated at the end of glioblast stage and remains active throughout the developmental process of oligodendrocytes (g).

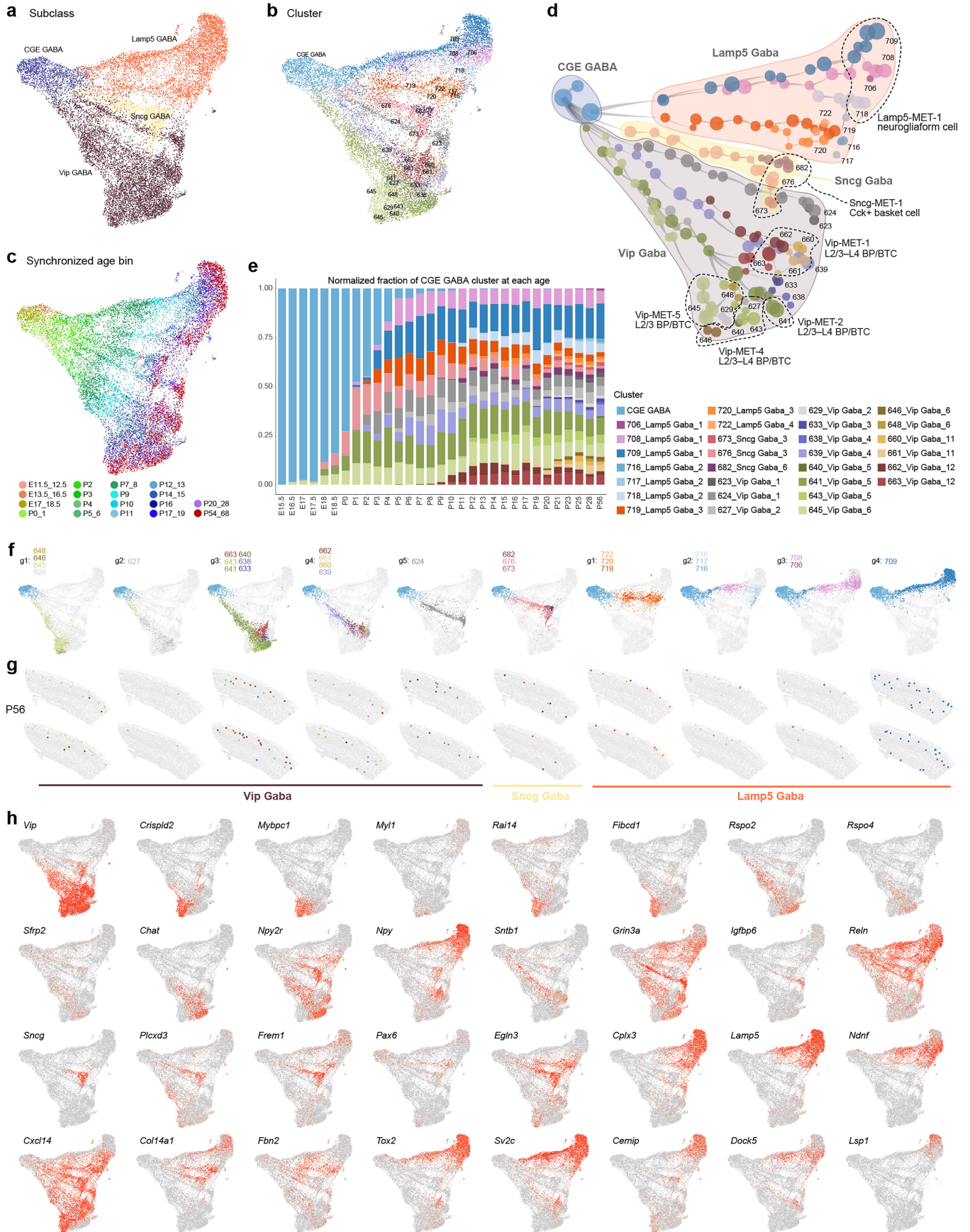


Extended Data Fig. 9 | See next page for caption.

# Article

**Extended Data Fig. 9 | Developmental trajectories of visual cortex MGE GABA cell types.** (a-c) UMAPs for MGE cells colored by subclass (a), cluster (b) and synchronized age (c). (d) Transcriptomic trajectory tree for MGE clusters starting from the common MGE GABA RG antecedent, with corresponding MET types labeled. Nodes are clusters subdivided by synchronized age, and edges represent antecedent-descendant relationship between adjacent nodes, with thicker end at the antecedent node and thinner end at the descendant node. Nodes are grouped by subclass, and adult clusters are labeled. Each MET type is enclosed within a dashed outline among the MGE cell clusters. BC, basket cells. MC, Martinotti cells. (e) Clusters are grouped together based on similar trajectories. Within each cluster group, all cells along their trajectories, including all antecedent nodes, are shown and are colored by cluster membership. (f) Spatial distribution of MGE subclasses and clusters within each subclass in P56 MERFISH data, based on the ABC-WMB Atlas<sup>14</sup>. (g) Marker genes illustrating cell type diversification along trajectories. (h) Cluster composition of all MGE cells at each age. Specifically, *Ascl1* and *Tead2* are strongly enriched in progenitor stage, followed by activation of *Lhx6*, *Nkx2-1* and *Lhx8*, which are key regulators of development of MGE-derived GABAergic neurons<sup>54,97,98</sup>. *Nkx2-1* and *Lhx8* are transiently expressed, while *Lhx6* persists to adulthood. We also observed expression of *Nfib* and *Sp9* in early stages of MGE cells, which slowly decrease and maintain low level expression in some adult

cell types. *Nfib*, *Sp9* and *Nkx2-1* are enriched in Pvalb chandelier and Lamp5 Lhx6 subclasses even in adulthood, while they are downregulated during development in most other MGE cell types. While *Sst* is expressed early in embryonic stages, *Pvalb* is not expressed until after eye opening. Within the Pvalb subclass, group 1 with clusters 736 and 754 and group 2 with cluster 741 both correspond to the Pvalb MET 3 type (in L5). Cluster 736 emerges from 754 at P19. Group 3 contains clusters 742 and 752, with 742 corresponding to Pvalb MET 4 (in L2/3), and 752 diverging from 742 at P17 and corresponding to Sst MET 2. Group 4 with clusters 743, 744 and 747 (split at P11) corresponds to Pvalb MET 2 (in L6). The *Th<sup>+</sup>* Pvalb cluster 735 corresponds to Pvalb MET 1 (in L6). However, the developmental trajectory of cluster 735 (emerging at P1) appears highly ambiguous, with its closest antecedent being Sst cluster 758. Many Sst clusters emerge relatively late within each group, with late activation of key genes. For example, *Crh* is activated around P5 and *Crhr2* around P10. Trajectory analysis suggests that *Crhr2<sup>+</sup>* clusters 811, 814, 818, 819 and 820 diverge from *Crh<sup>+</sup>* cluster 758 around P5, with further divergence occurring after P19. In group 1, 757, 758 and 761 split at P20, 811 and 814 split at P12, and 818, 819 and 820 split at P21. In group 2, 795 is born around P14, while 797 and 806 diverge from 803 around P17–19. In group 3, all 5 clusters diverge from 792 at P19–21. In group 5, 777 splits from 780 at P19.

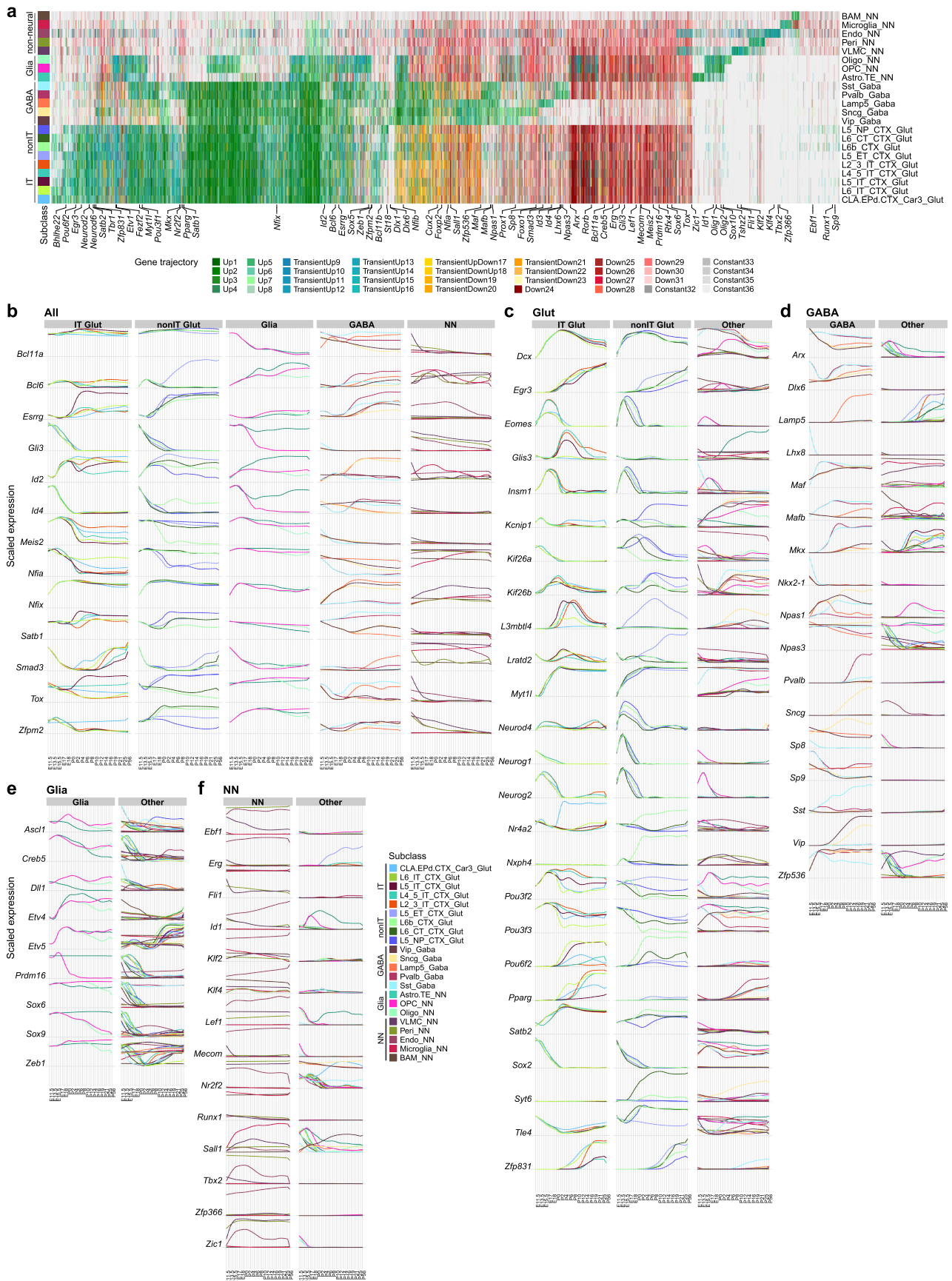


Extended Data Fig. 10 | See next page for caption.

# Article

**Extended Data Fig. 10 | Developmental trajectories of visual cortex CGE GABA cell types.** (a-c) UMAPs for CGE cells colored by subclass (a), cluster (b) and synchronized age (c). (d) Transcriptomic trajectory tree for CGE clusters starting from the common CGE GABA antecedent, with corresponding MET types labeled. Nodes are clusters subdivided by synchronized age, and edges represent antecedent-descendant relationship between adjacent nodes, with thicker end at the antecedent node and thinner end at the descendant node. Nodes are grouped by subclass, and adult clusters are labeled. Each MET type is enclosed within a dashed outline among the CGE cell clusters. BP/BTC, bipolar/bitufted cells. (e) Cluster composition of all CGE cells at each age. (f) Clusters are grouped together based on similar trajectories. Within each cluster group, all cells along their trajectories, including all antecedent nodes, are shown and are colored by cluster membership. (g) Spatial distribution of CGE subclasses and clusters within each subclass in P56 MERFISH data, based on the ABC-WMB Atlas<sup>14</sup>. (h) Marker genes illustrating cell type diversification along trajectories. Specifically, in the Vip subclass, group 1 (*Crispld2* and *Mybpc1*) contains

clusters 645, 646, 648 and 629 that are split from 645 around P21. Group 2 (*Rspo2* and *Rspo4*) contains cluster 627. Group 3 (*Chat* and *Npy2r*) contains the root cluster 641, plus 643 and 663 emerging at P11, 633 at P19, 638 at P23, and 640 at P56. Group 4 (*Sntb1*) contains clusters 662, 661, 660 and 639, with 639 emerging at P2, 662 at P9, 661 at P15, and 660 at P21. Group 5 (*Grin3a* and *Igf1bp6*) contains two clusters, with 623 split from 624 at P23. The Sncg subclass has one main trajectory marked by *Plcxd3*, *Frem1*, *Egln3* and *Sncg*, with *Sncg* expressed the latest. Among the 3 Sncg clusters, cluster 676 gives rise to 682 and 673 at P11 and P20, respectively. In the Lamp5 subclass, group 1 (*Egln3*, *Col14a1* and *Fbn2*) contains clusters 719 (emerging at P2), 720 (P12) and 722 (P21). Group 2 (clusters 716, 717 and 718, split from 718 at P25) and group 3 (clusters 706 and 708, split at P28) are very similar, marked by shared expression of *Dock5* and *Ndnf*, with 708 as the root cluster and 718 split from 708 at P11. Group 4, containing cluster 709 and enriched in *Lsp1* and *Cemip*, shares expression of *Tox2* and *Sv2c* with groups 2 and 3 and emerges at P1 along with cluster 708.



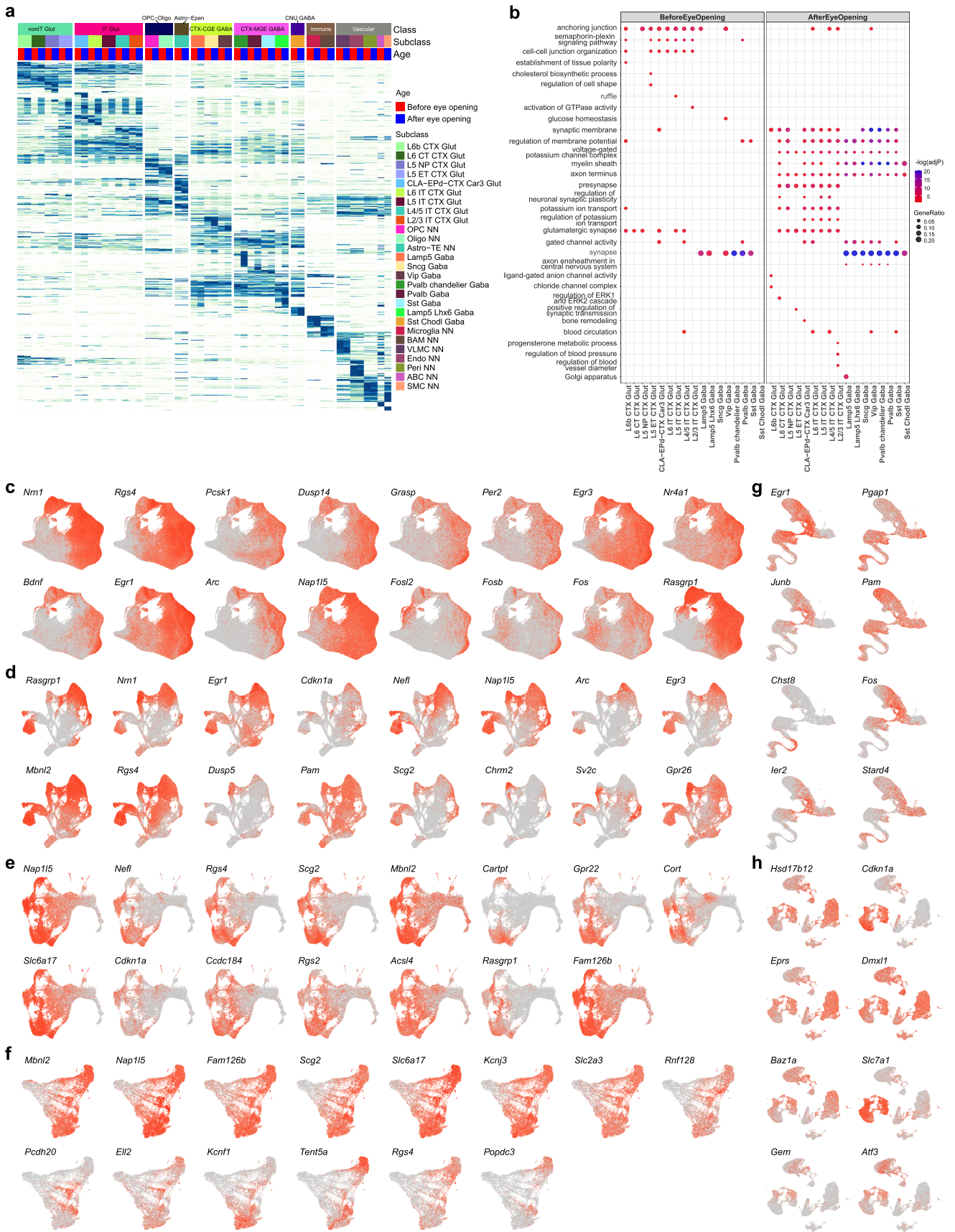
Extended Data Fig. 11 | See next page for caption.

## Article

---

**Extended Data Fig. 11 | Gene expression trajectories across cell types and ages during development.** (a) Heatmap showing the trajectory types of all DE genes, with each column representing a gene and each row corresponding to a subclass. Colors represent different trajectory types, with green representing gene trajectories with increased expression over time, blue representing transiently upregulated gene trajectories, yellow representing transiently downregulated gene trajectories, red representing gene trajectories with

decreased expression over time, and grey representing gene trajectories with no change over time. (b-f) Unique gene expression trajectories of individual genes in different subclasses. The trajectories are shown for genes that are expressed in all subclasses (b), genes specific to glutamatergic subclasses (c), GABAergic subclasses (d), glial subclasses (e), or other non-neuronal (NN) subclasses (f). The “Other” label in each panel refers to other subclasses not highlighted in that panel.

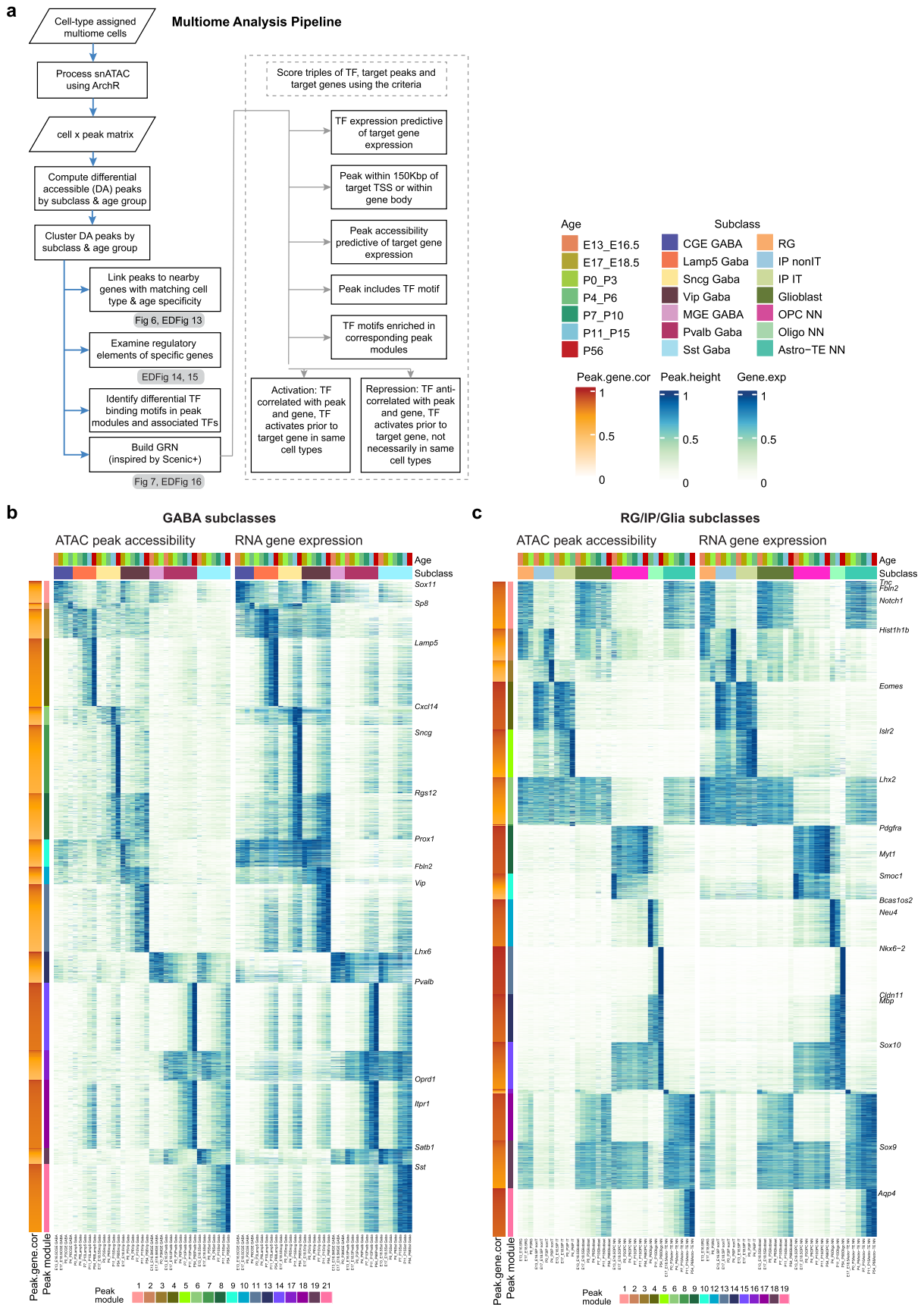


# Article

## **Extended Data Fig. 12 | Gene expression changes before and after eye opening.**

**(a)** Heatmap showing the expression of specific DE genes in each subclass before and after eye opening. **(b)** GO enrichment dot plot showing example significant top GO terms before or after eye opening in each neuronal subclass. Dot size and color indicate gene ratio (the percentage of genes that are present in a GO term compared to the total number of genes in that category) and significance ( $-\log_{10}$  adjusted p-value), respectively. A Benjamini-Hochberg (BH)

adjusted p-value cutoff of 0.01 was used to determine significant GO terms. Max gene ratio was set to 0.2 and max significance was set to 20. **(c-h)** UMAP representations showing expression changes of representative immediate early genes (IEGs) in IT glutamatergic (c), nonIT glutamatergic (d), MGE GABAergic (e), CGE GABAergic (f), glial (g) and immune and vascular (h) cell types.



Extended Data Fig. 13 | See next page for caption.

# Article

---

**Extended Data Fig. 13 | Correspondence of chromatin accessibility and gene expression across GABAergic and glial cell types and ages during development.** (a) Multiome data analysis flowchart. (b,c) Heatmap representations of corresponding peak accessibility and gene expression in GABA subclasses (b) and glia subclasses (c). In each panel, each row corresponds to a peak-gene pair, ordered by peak module and peak-gene correlation, and each column

corresponds to a subclass-by-age group. The left-hand heatmap shows the average peak accessibility in each subclass-by-age group. The right-hand heatmap shows the average gene expression in each subclass-by-age group. Accessibility and expression values are normalized, with maximum value of 1 per peak or gene and 0 indicating no accessibility or expression.



# Article

## Extended Data Fig. 14 | Differential accessibility peaks associated with the *Cux2* gene in different cell types or different developmental ages.

**(a)** Heatmap representation of accessibility of differentially accessible peaks located in *Cux2* gene body and 50 kb upstream. Each row corresponds to a peak, ordered by peak module, and each column corresponds to a cell category defined by subclass and age group. The *Cux2* gene expression level is shown in purple at the top. The heatmap color represents the average peak accessibility (height) in each subclass-by-age group, normalized with 1 indicating the maximum value for each peak and 0 indicating no accessibility. The peak module and maximum peak height are shown for each peak to the right. Specific peaks are numbered and labeled. **(b)** The accessibility tracks per subclass surrounding the *Cux2* gene, along with the genomic locations of labeled peaks in (a). TSS, transcription start site. **(c)** UMAP representation of Multiome nuclei, colored by *Cux2* expression and accessibility of a subset of peaks labeled in (a). TF gene *Cux2* exhibits markedly distinct temporal patterns across different cell types. We extracted all the accessibility peaks located within the *Cux2* gene body (193 kb) and 50 kb upstream of *Cux2*'s main TSS. We observed strikingly complex accessibility patterns of different peaks in

different subclasses and ages (a). There are distinct peak modules specific to IP (module 2), IP IT and IMN IT (module 3), IMN IT and upper layer IT cells (module 5), Car3 cells (module 6), shared by L2/3 IT, L4/5 IT and Car3 cells (module 7), specific to early L2/3 (module 9), shared by OPC and MGE (module 14), shared by IP and MGE (module 1), or specific to MGE (modules 10,11,13). We labeled specific peaks with distinct patterns in both the heatmap and the cell-type genomic tracks (a and b). Most of the peaks present in early-stage RG and IP populations disappear in adulthood, except those that are present near the promoter or widely accessible. The accessibility of peaks in the promoter area overall shows strong consistency with RNA expression across all the cell types under study, while the peaks in more distal areas show accessibility in a highly cell-type and temporally specific manner. To study the subtler temporal progression, we examined the expression of *Cux2* gene and accessibility of specific peaks at the single cell level (c). Peak 1 is specific to IP, Peak 10 to IMN IT and L2-4, Peak 8 to MGE (decreasing over time), Peak 5 to L2-4 (increasing over time), and Peak 15 specific to Car3 and surprisingly in microglia (although expression of *Cux2* gene in microglia was not observed).



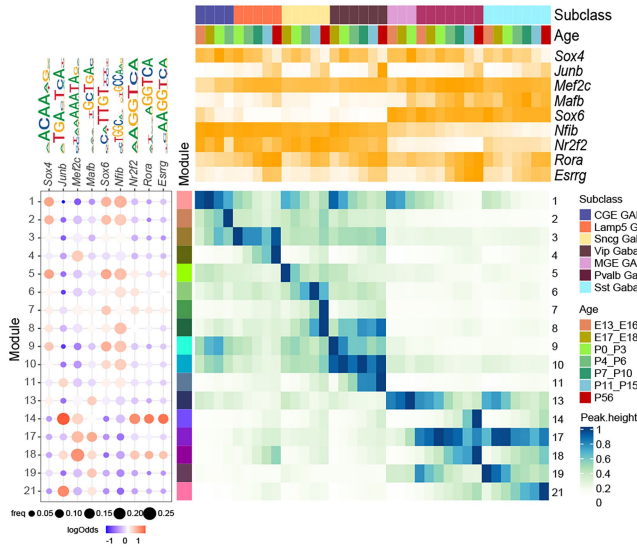
# Article

## Extended Data Fig. 15 | Differential accessibility peaks associated with the *Grik1* gene in different cell types or different developmental ages.

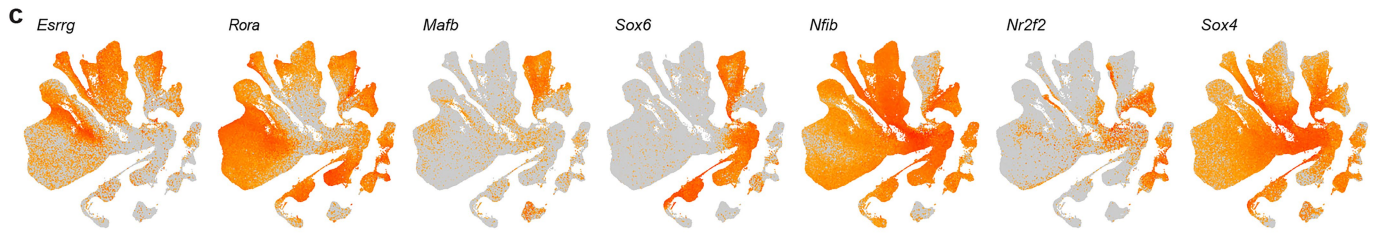
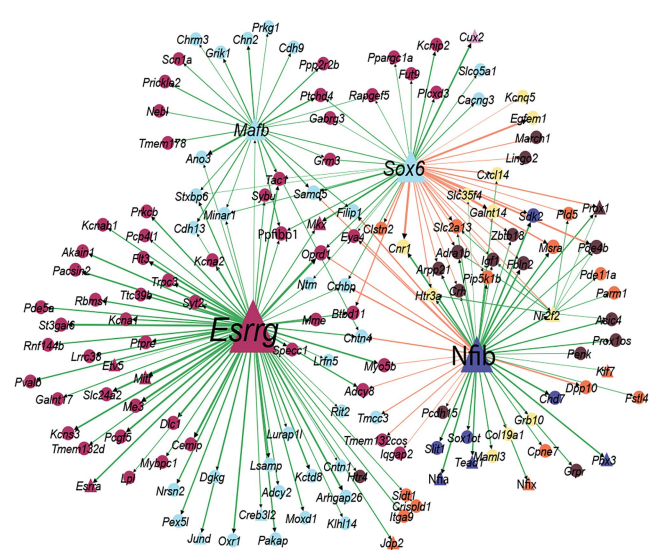
(a) Heatmap representation of accessibility of differentially accessible peaks located in *Grik1* gene body and 50 kb upstream. Each row corresponds to a peak, ordered by peak module, and each column corresponds to a cell category defined by subclass and age group. The *Grik1* gene expression level is shown in purple at the top. The heatmap color represents the average peak accessibility (height) in each subclass-by-age group, normalized with 1 indicating the maximum value for each peak and 0 indicating no accessibility. The peak

module and maximum peak height are shown for each peak to the right. Specific peaks are numbered and labeled. (b) The accessibility tracks per subclass surrounding the *Grik1* gene, along with the genomic locations of labeled peaks in (a). TSS, transcription start site. (c) UMAP representation of Multiome nuclei, colored by *Grik1* expression and accessibility of a subset of peaks labeled in (a). Ion channel gene *Grik1* is activated postnatally in L4/5IT, L5NP, OPC, MGE and CGE, and its 394 kb gene body is associated with highly distinct peaks in each case, allowing fine-tuning of gene regulation specific to each cell type.

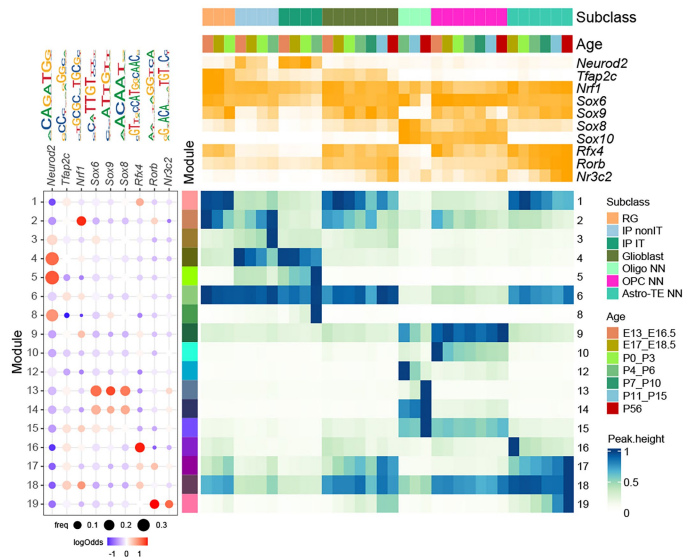
**a** TF regulators of GABA subclasses



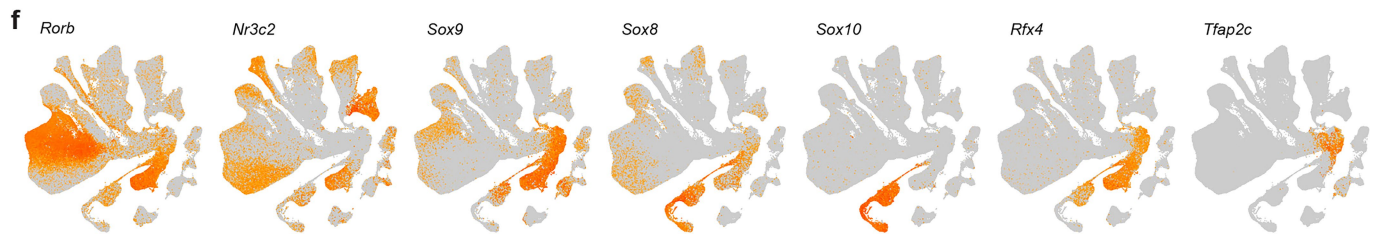
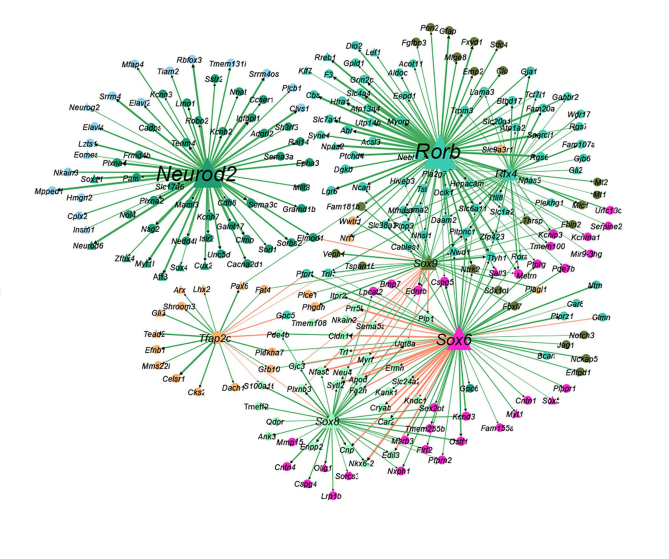
**b** Gene regulatory network of GABA subclasses



**d** TF regulators of Glia subclasses



**e** Gene regulatory network of Glia subclasses

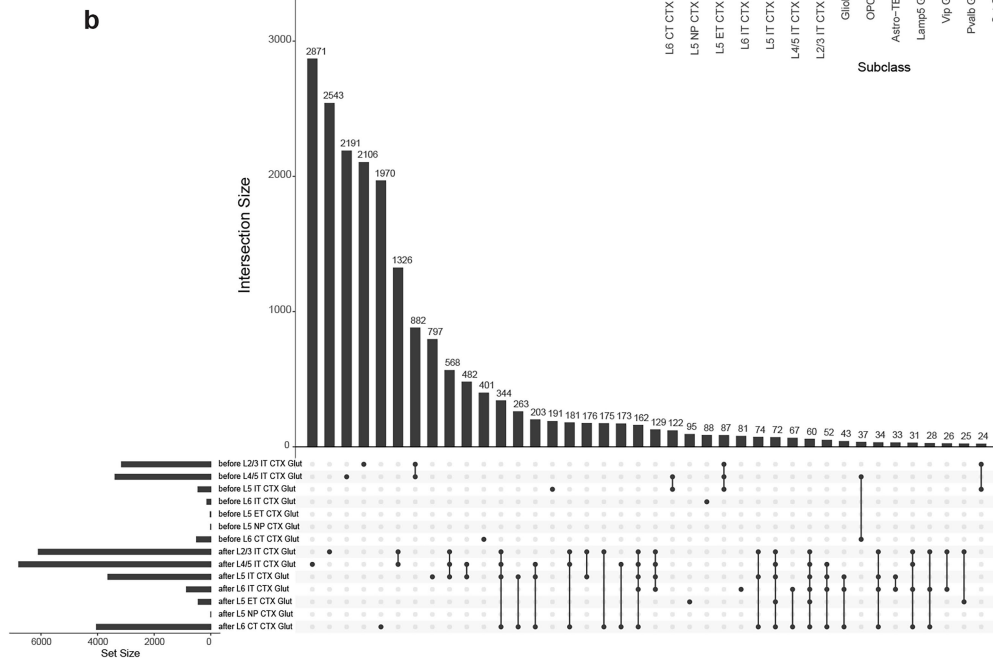
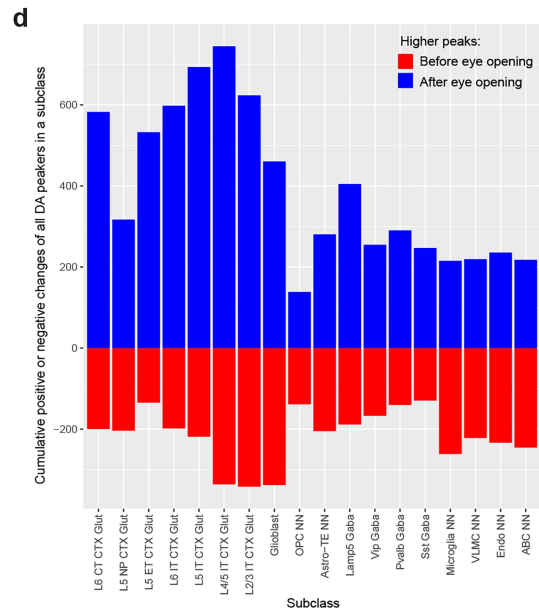
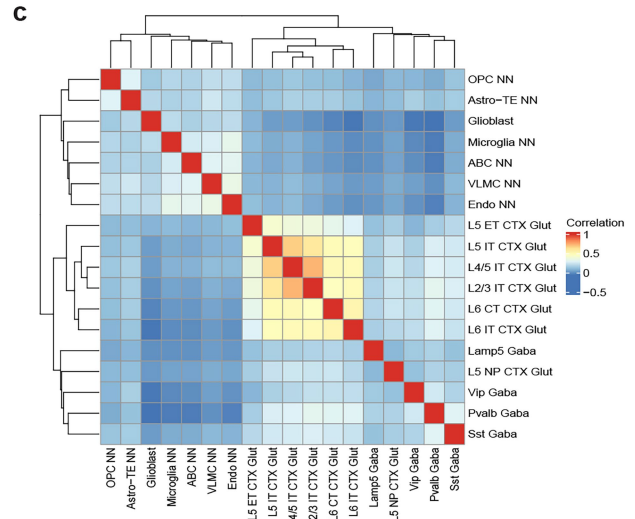
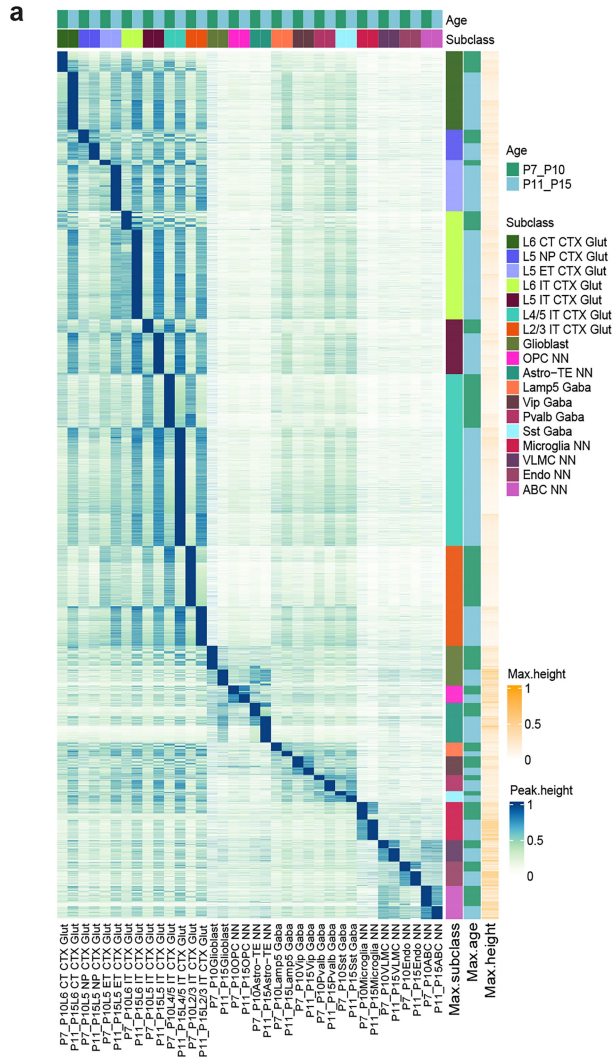


Extended Data Fig.16 | See next page for caption.

# Article

**Extended Data Fig. 16 | Gene regulatory networks for GABAergic and Glial cell types.** (a,d) TF DNA-binding motif enrichment for chromatin accessibility peak modules with different cell type and temporal specificities in GABAergic (a) and Glial (d) cell types. Within each panel, the dot plot to the left shows the average motif frequency for each peak module, dot size indicates the frequency, and color corresponds to the log-odds of motif occurrence in each module relative to random chance. The large heatmap at the bottom shows the average accessibility for each peak module (in rows) across each subclass-by-age group (in columns). The heatmap at the top shows the average expression of specific TFs corresponding to the motif across each subclass-by-age group. The values in the heatmaps are normalized per peak module or gene with 1 indicating the maximum value, and 0 indicating no accessibility or expression. (b,e) Gene regulatory networks for GABAergic (b) and Glial (e) cell types. Nodes represent genes, with triangles denoting TFs and circles denoting other genes. Each node is colored according to the subclass in which the gene is most highly expressed. Activation interactions are colored in green, repression in orange, and edge widths reflect interaction strengths. (c,f) Expression of TF regulators on UMAPs for GABAergic (c) and Glial (f) cell types. Here we also provide additional notes regarding specific TF regulators for all cell types. *Nr4a2* is identified as a key regulator of the CLA-EPd-CTX Car3 subclass, with its motif enriched in both early and late Car3 subclass-specific peak modules, targeting marker genes including *Car3*, *Oprk1*, and *Nr2f2* (Fig. 7a,b). *Nr2f2*, activated after *Nr4a2*, is predicted to regulate other late Car3 markers such as *Synpr* and *Col11a1*, suggesting a feed-forward pathway involved in maturation. The bHLH neurogenic motifs, shared by TFs such as *Neurog1/2*, *Neurod1/2/4/6* and *Bhlhe22* (also known as *Bhlhe5*), may exhibit subtle differences depending on other bHLH dimerization partners. These motifs are depleted in peak modules enriched in L2/3 IT and Car3 subclasses (Fig. 7a). While all these TFs are highly expressed in the IP or IMN populations, they are downregulated to various degrees in later stages. *Neurod1* and *Bhlhe22* are downregulated in deep-layer neurons, while *Neurod6* is reduced in upper layers (Fig. 7e). *Prdm8*, a member of histone methyltransferase family, is strongly co-expressed with *Neurod1* and *Bhlhe22*. *Bhlhe22* is known to recruit *Prdm8* to repress target genes, and loss of either leads to similar neuronal mistargeting phenotypes<sup>99</sup>. Our analysis suggests *Bhlhe22* represses deep-layer markers in upper-layer neurons, and when it is downregulated in deep layers, *Neurod6* may activate the same targets (Fig. 7b). This dynamic likely fine-tunes layer-specific gene expression during postnatal IT neuron development. We found POU-III class TFs, *Pou3f1/2/3*, as key regulators for upper layer neurons, consistent with their crucial roles in specifying and maintaining the identity of these neuronal populations<sup>61</sup>, and predicted *Cux2* as a downstream target (Fig. 7a,b). While *Pou3f2* (*Brn2*) is

mostly reported as activators, it has been reported to act as a repressor to downregulate *Cdh13* and *Mitf* in melanoma cells<sup>100,101</sup>, and our analysis suggests that these TFs may repress other deep layer markers, such as *Cobl*. *Rfx3* is predicted as another regulator of upper layer neurons and is predicted to be a downstream target of *Pou3f1*. *Etv1* is predicted to act as an activator in the L5 IT subclass, with subclass-specific targets such as *Myl4* and *Arhgap25*. *Fosl2* is identified as a candidate regulator of the L6 IT subclass; while most IEGs increase expression after eye opening, *Fosl2* is activated early in L6 IT neurons soon after their divergence from IMN IT (Fig. 7b, Extended Data Fig. 12). *Tcf4* and *Sox4* are upregulated in IP stage and gradually decrease after IMN stage. We predicted that *Sox4* simultaneously represses the premature expression of certain neuronal markers, particularly L6 CT markers (Fig. 7c,d), to help maintain cells in the IMN state. It was shown previously that *Sox4* can act as a transcriptional repressor by interfering with the assembly of transcriptional machinery at promoters<sup>102</sup>. We predicted that *Tcf4* supports early neurogenesis and maturation while preventing premature activation of late-stage genes that are typically expressed after eye opening (Fig. 7f,g). It was shown previously that loss of *Tcf4* leads to elevated baseline levels of cFos<sup>103</sup> and profound changes in the structure and excitability of adult neurons. We identified *Mafb* and *Sox6* as major MGE regulators and *Nfib* and *Nr2f2* as CGE regulators. *Mafb* is known to regulate MGE interneuron fate and function<sup>104</sup>. Members of the Nuclear Factor I family, *Nfib*, *Nfia*, and *Nfix*, are known to be co-expressed specifically in CGE-derived interneurons<sup>14</sup>, and our data show that they are activated in this class by E16.5. In the oligodendrocyte trajectory, we observed significant enrichment of SOX motifs, particularly *Sox6*, *Sox8*, and *Sox9*, in oligodendrocyte-specific but not OPC-specific peak modules. The motif for *Sox10*, a close homolog of *Sox8*, was not significantly enriched, possibly due to its low quality in the JASPAR motif database. *Sox8* and *Sox10* are expressed selectively in the OPC-Oligo class. In contrast, *Sox6* and *Sox9* are broadly expressed in RG, glioblasts, astrocytes, and OPCs, but are turned off during oligodendrocyte maturation, *Sox9* in late OPCs and COPs and *Sox6* in NFOLs. Based on these patterns, we propose that *Sox8* and *Sox10* promote oligodendrocyte maturation, while *Sox6* and *Sox9* act as stage-specific repressors of maturation. *Sox8*, *Sox9*, and *Sox10* belong to the SOXE group of TFs, which are known to often function as dimers and regulate diverse developmental processes<sup>105</sup>. *Sox9* may inhibit oligodendrocyte maturation until appropriate developmental signals, which in turn leads to its downregulation. Similarly, *Sox6* has been shown to repress oligodendrocyte maturation in mouse spinal cord<sup>106</sup>. Our results highlight the intricate interplay among SOX family TFs in guiding stage-specific transitions during oligodendrocyte development.



Extended Data Fig. 17 | See next page for caption.

# Article

**Extended Data Fig. 17 | Cell-type specific chromatin accessibility changes before and after eye opening.** **(a)** Heatmap representation of accessibility of DA peaks before and after eye opening. Each row corresponds to a peak, ordered by the subclass and age group with maximum accessibility. **(b)** Number of DA peaks before and after eye opening shared among different glutamatergic subclasses. Each column corresponds to a combination of different subclasses, and the bar height represents the number of peaks shared by the given combination of subclasses and age group (intersection size). The bar graph to

the left of the subclass labels shows the total number of DA peaks for each subclass before or after eye opening (set size). **(c)** Correlation of the chromatin accessibility changes before and after eye opening among all subclasses. The chromatin accessibility change is measured as the difference of average peak height between the two age groups for the given subclass, based on all the DA peaks defined in (a). **(d)** Cumulative positive and negative changes for each subclass before and after eye opening based on all the DA peaks defined in (a).

## Reporting Summary

Nature Portfolio 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 Portfolio 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 type="checkbox"/>	<input checked="" 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 <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 type="checkbox"/>	<input checked="" 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 checked="" type="checkbox"/>	<input 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 <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 type="checkbox"/>	<input checked="" 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	scRNA-seq data was collected using the 10x Genomics Chromium Single Cell 3' v3 kits. Multiome data was collected using the 10x Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle. Sequencing was performed on the Illumina HiSeq2500 and NovaSeq6000.
Data analysis	<p>Raw scRNA-seq FASTQ files were processed with the Cell Ranger v6.0.1 pipeline using default parameters. Raw snRNA-seq FASTQ and ATAC FASTQ files were processed with the Cell Ranger Arc v2.0.2. Raw MERSCOPE data were decoded using Vizgen software (v231). For cell segmentation, an in-house model to segment cells was used by applying the human-in-the-loop approach introduced in Cellpose 2.0.</p> <p>Doublets were identified using a modified version of the DoubletFinder algorithm which is available in <a href="https://github.com/AllenInstitute/scrattch.hicat">scrattch.hicat</a> (https://github.com/AllenInstitute/scrattch.hicat, v1.0.9).</p> <p>scRNA-seq clustering, snRNA-seq clustering and differential gene expression analysis were performed in R (v4.3.1) using the <a href="https://github.com/AllenInstitute/scrattch.bigcat">scrattch.bigcat</a> package (https://github.com/AllenInstitute/scrattch.bigcat, v0.0.5), which also contains many functions to visualize the data together with the <a href="https://github.com/AllenInstitute/scrattch.vis">scrattch.vis</a> package (https://github.com/AllenInstitute/scrattch.vis, v0.0.210). <a href="https://github.com/AllenInstitute/scrattch.bigcat">Scrattch.bigcat</a> adopted the Parquet file format for storing sparse matrices, which allows for manipulation of matrices that are too large to fit in memory through memory mapping to files on disk. The whole gene count matrices were chunked to smaller Parquet files with a bin size of 10,000 for cells and 500 for genes, which could be loaded efficiently and concurrently using the <a href="https://github.com/apache/arrow">arrow</a> package (v12.0.1, https://github.com/apache/arrow/, https://arrow.apache.org/docs/r/).</p> <p>scRNA-seq synchronized developmental age was estimated in R (v4.3.1) using the KNN algorithms implemented in <a href="https://github.com/LTLA/BiocNeighbors">BiocNeighbors</a> package (v1.16.0, https://github.com/LTLA/BiocNeighbors).</p> <p>scRNA-seq label transfer for adult cells (P56) was conducted using the <a href="https://github.com/AllenInstitute/scrattch.mapping">scrattch-mapping</a> package (v0.55, https://github.com/AllenInstitute/scrattch.mapping). scRNA-seq label transfer for PO-P28 was performed using the <a href="https://github.com/satijalab/seurat">Seurat</a> package (v4.1.4, https://github.com/satijalab/seurat). For mapping to external published data, <a href="#">scVI</a> was used for integration. In the integrated space, <a href="#">Random Forest classifier</a>, implemented in</p>

Python package scikit-learn (v1.3.1, <https://scikit-learn.org/stable/>), was applied to predict the most likely cell type of the scRNA-seq data (E11.5-P4). scRNA-seq trajectory was built using the k-nearest neighbors (KNNs) algorithms implemented in R package BiocNeighbors (v1.16.0, <https://github.com/LTLA/BiocNeighbors>) in the integrated space computed by Seurat package (v4.1.4, <https://github.com/satijalab/seurat>). The trajectory was built based on synchronized developmental age bin for postnatal ages and pseudotime by Monocle3 (<https://github.com/cole-trapnell-lab/monocle3>) for prenatal ages. To compare and validate the trajectory, Python package scVI was applied (v1.0.4, <https://scvi-tools.org/>) using three sets of highly variable genes (HVGs, 1000, 2000, and 3000 genes) between adjacent age bins under three model architectures with varying complexity: default (n\_hidden = 128, n\_layer = 1, n\_latent = 10), medium (n\_hidden = 128, n\_layer = 2, n\_latent = 32), and large (n\_hidden = 256, n\_layer = 3, n\_latent = 32). scRNA-seq gene modules were computed using Louvain clustering implemented in the igraph package (v1.5.1.9002, <https://r.igraph.org/>). To build the gene trajectory, Generalized Additive Model (GAM), which was implemented via the “gam” function from the mgcv R package (v1.9-3, <https://CRAN.R-project.org/package=mgcv>), was fit to predict each DE gene’s normalized expression trajectory within each subclass, with synchronized age as the predictor. The fitted gene trends were hierarchically clustered using fastcluster R package (v1.3.0, <https://CRAN.R-project.org/package=fastcluster>) with a Ward linkage and Euclidean distance metric. The gene trend clusters were further merged using k-means. Gene ontology enrichment analysis was conducted using R packages clusterProfiler (v4.0.5, <https://github.com/YuLab-SMU/clusterProfiler>) and gprofiler (v0.2.2, <https://CRAN.R-project.org/package=gprofiler2>). Integration of scRNA-seq data and Multiome snRNA-seq data was conducted via Python package scVI (v1.0.4, <https://scvi-tools.org/>). In the integrated space, Random Forest classifier, implemented in Python package scikit-learn (v1.3.1, <https://scikit-learn.org/stable/>), was applied to predict the most likely cell type of the snRNA-seq data. Multiome peak calling was conducted using ArchR (v1.0.2, <https://github.com/GreenleafLab/ArchR>), and differentially accessible peaks were identified using the scatch.bigcat package (<https://github.com/AllenInstitute/scatch.bigcat>). Peak modules were computed using the Jaccard-Leiden clustering algorithm. The KNN algorithm was applied to each peak and Jaccard similarity graph (v1.5.1.9002, <https://r.igraph.org/>) was computed based on number of shared nearest neighbors between each pair of peaks. Leiden clustering algorithm (v1.1.3, <https://github.com/kharchenkolab/leidenAlg>) was applied to the Jaccard similarity graph. Differential motif analysis was performed using the chi-square test implemented in scatch.bigcat. SCENIC+ (v1.0a2, <https://github.com/aertslab/scenicplus>) was reimplemented with modifications tailored to developmental datasets and to better integration with the ArchR pipeline. Multiome data (snRNA-seq and snATAC-seq) was integrated with motif analysis to identify regulatory relationships through triplet scores linking TFs, accessible chromatin peaks, and target genes. Solo was implemented to detect MERFISH doublets. MERFISH cells were mapped to PO VIS (E18.5, PO and P1) reference using scatch.mapping package at cluster level. Selected MERFISH and PO VIS reference cells labeled as IMN and IP were integrated using scVI (n\_hidden = 256, n\_layer = 3, n\_latent = 32), and a Random Forest classifier was trained on the scRNA-seq sub-clusters. To visualize the scRNA-seq data, Multiome data and MERFISH data, we used the single-cell visualization tool CirroCumulus (v1.1.56, <https://cirroCumulus.readthedocs.io/en/latest/>).

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The identifier containing links to all primary data is <https://assets.nemoarchive.org/dat-5y9mf0h>. The 10x scRNA-seq dataset is available at <https://assets.nemoarchive.org/dat-0oyried>, the Multiome snRNA-seq is available at <https://assets.nemoarchive.org/dat-bbqchppq>, and the Multiome snATAC-seq is available at <https://assets.nemoarchive.org/dat-5ke3d8i>. In addition, the processed 10x scRNA-seq dataset is available at [https://allen-developmental-mouse-atlas.s3.amazonaws.com/scRNA/DevVIS\\_scRNA\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.amazonaws.com/scRNA/DevVIS_scRNA_processed.h5ad), the processed Multiome snRNA-seq dataset is available at [https://allen-developmental-mouse-atlas.s3.us-west-2.amazonaws.com/Multiome/DevVIS\\_multiome\\_snRNA\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.us-west-2.amazonaws.com/Multiome/DevVIS_multiome_snRNA_processed.h5ad), and the processed Multiome snATAC-seq dataset is available at [https://allen-developmental-mouse-atlas.s3.amazonaws.com/Multiome/DevVIS\\_multiome\\_snATAC\\_processed.h5ad](https://allen-developmental-mouse-atlas.s3.amazonaws.com/Multiome/DevVIS_multiome_snATAC_processed.h5ad).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="This study does not involve human research participants."/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="NA"/>
Population characteristics	<input type="text" value="NA"/>
Recruitment	<input type="text" value="NA"/>
Ethics oversight	<input type="text" value="NA"/>

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

## 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](https://www.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	Sample size calculations were not performed. For scRNA-seq data and Multiome data, sample size was determined based on extensive prior experience using these technologies in brain. Post-hoc analysis of clustering across developmental time points shows that it is robust to downsampling, which indicates that the sample size is sufficient to distinguish cell types as described in the manuscript.
Data exclusions	Cells from scRNA-seq underwent stringent QC and cell quality was assessed based on gene detection, qc score, and doublet score. Nuclei from Multiome data underwent stringent QC and nucleus quality was assessed based on gene detection, doublet score, signal-to-background ratio (TSS) and number of unique nuclear fragments (nFragments). The qc score was calculated by summing the log transformed expression of a set of genes whose expression level is decreased significantly in poor quality cells. We used this qc score to quantify the integrity of cytoplasmic mRNA content, which tended to show bimodal distribution. Cells at the low-end were very similar to single nuclei, which we removed for downstream analysis. Doublets were identified using a modified version of the DoubletFinder algorithm (available in <a href="https://github.com/AllenInstitute/scratch.hicat">scratch.hicat</a> , <a href="https://github.com/AllenInstitute/scratch.hicat">https://github.com/AllenInstitute/scratch.hicat</a> , v1.0.9) and removed when doublet score > 0.3. Post-clustering we excluded clusters outside cortex by comparing the expression of Foxg1, Emx1 or Emx2 for scRNA-seq and Multiome data. MERFISH cell quality was assessed based on detected genes, detected mRNA transcripts, and the percentage of blank barcodes per cell. Additional clusters of Multiome data outside cortex were excluded by PO MERFISH data.
Replication	Most of the scRNA-seq experiments were carried out at least twice independently and at least 2 mice were used. Not all attempts at replication were successful. In unique cases library construction fails due to technical issues which warrant another attempt to replicate library construction. For MERFISH data, we evenly sampled a mouse brain to capture most brain regions and get an even distribution of cell types throughout the brain. For MERFISH data, we evenly sampled a mouse brain to capture most brain regions. Most brain regions are captured across multiple sections and replication was successful.
Randomization	Randomization of animals to different groups is not relevant to our study design. There were no experimental vs. control groups.
Blinding	Prior to clustering, single cell transcriptomes were analyzed for previously known marker genes and were segregated into large groups: non-neuronal, glutamatergic and GABAergic. Clustering was then performed blind to the cell source or any other metadata that could reveal sample identity.

## 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 & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input 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"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved 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

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

### Laboratory animals

For scRNA-seq data, animals were euthanized at E11.5-E18.5 (n=13), P0-P5 (n=6), P6-P10 (n=7), P11-P28 (n=28). Each animal's unique ID, sex, age, and genotype are listed in Supplementary Table 1. The donor information for the adult mice can also be found in Yao et al., 2023 (<https://doi.org/10.1038/s41586-023-06812-z>) Supplemental Table 2.  
For Multiome data, animals were euthanized at E13.5-E18 (n=16), P0-P5 (n=10), P8-P9 (n=2), P11 (n=1), P14 (n=3), P56 (n=5). Each animal's unique ID, sex, age, and genotype is listed in Supplementary Table 1. In some cases, transgenic mice were used for

fluorescence-positive cell isolation by fluorescence-activated cell sorting (FACS). To enrich neurons profiled by scRNA-seq, cells were isolated from the pan-neuronal Snap25-IRES2-Cre line (RRID: IMSR\_JAX:023525) crossed to the Ai14-tdTomato reporter (RRID: IMSR\_JAX:007914).

Mice had ad libitum access to food and water and were group-housed within a temperature (21-22°C), humidity (40-51%), and light (14/10hr light/dark cycle) controlled room within the vivariums of the Allen Institute for Brain Science.

#### Wild animals

This study did not involve wild animals.

#### Reporting on sex

From ages P0 to P56, male mice were used to collect scRNA-seq data (P12 and P56 have both male and female). For the embryonic ages E11.5 to E18 we collected brains and performed post-hoc sex identification. Male embryo's were used to collect majority samples of the embryonic ages.

For Multiome data, male mice were used to collect data (P0, P4, P14 and P56 have both male and female). For the embryonic ages E11.5 to E18, we collected brains and performed post-hoc sex identification. Male embryo's were used to collect majority samples of the embryonic ages.

For MERFISH data, male mouse was used to collect data (P0).

#### Field-collected samples

This study did not involve field-collected samples.

#### Ethics oversight

All experimental procedures related to the use of mice were approved by the Institutional Animal Care and Use Committee of the Allen Institute for Brain Science, in accordance with NIH guidelines.

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

## Plants

#### Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

#### Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

#### Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.*

## 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

Sample preparation was done according to protocols: Allen Institute for Brain Science 2021. Slice Preparation with Tissue Dissociation - Mouse Protocol. protocols.io <https://dx.doi.org/10.17504/protocols.io.bq6wmzfe> and Allen Institute for Brain Science 2020. FACS Single Cell Sorting. protocols.io <https://dx.doi.org/10.17504/protocols.io.be4cjgsw>.

Single cells were isolated by adapting previously described procedures. The brain was dissected, submerged in ACSF, embedded in 2% agarose, and sliced into 350- $\mu$ m coronal sections on a compresstome (Precisionary Instruments). Block-face images were captured during slicing. Regions of interest (ROIs) were then microdissected from the slices and dissociated into single cells as previously described. Fluorescent images of each slice before and after ROI dissection were taken at the dissection microscope. These images were used to document the precise location of the ROIs using annotated coronal plates of CCFv3 as reference.

Dissected tissue pieces were digested with 30 U/ml papain (Worthington PAP2) in ACSF for 30 minutes at 30°C. Due to the short incubation period in a dry oven, we set the oven temperature to 35°C to compensate for the indirect heat exchange, with a target solution temperature of 30°C. Enzymatic digestion was quenched by exchanging the papain solution three times with quenching buffer (ACSF with 1% FBS and 0.2% BSA). Samples were incubated on ice for 5 minutes before trituration. The tissue pieces in the quenching buffer were triturated through a fire-polished pipette with 600- $\mu$ m diameter opening approximately 20 times. The tissue pieces were allowed to settle and the supernatant, which now contained suspended single cells, was transferred to a new tube. Fresh quenching buffer was added to the settled tissue pieces, and trituration and supernatant transfer were repeated using 300- $\mu$ m and 150- $\mu$ m fire polished pipettes. The single cell suspension was passed through a 70- $\mu$ m filter into a 15-ml conical tube with 500  $\mu$ l of high BSA buffer (ACSF with 1% FBS and 1% BSA) at the bottom

to help cushion the cells during centrifugation at 100 x g in a swinging bucket centrifuge for 10 minutes. The supernatant was discarded, and the cell pellet was resuspended in the quenching buffer. We collected 1,508,284 cells without performing FACS. The concentration of the resuspended cells was quantified, and cells were immediately loaded onto the 10x Genomics Chromium controller.

To enrich for neurons or live cells, cells were collected by fluorescence-activated cell sorting (FACS, BD Aria II running FACSDiva v8) using a 130- $\mu$ m nozzle. Neuronal enrichment was achieved by sorting RFP+ cells from Snap25-IRES2-Cre/wt;Ai14/wt transgenic mice counterstained with DAPI to exclude dead or dying cells. To enrich for live cells, samples were stained and sorted for Hoechst+ or Calcein+/Hoechst+ signal. Cells were prepared for sorting by passing the suspension through a 70- $\mu$ m filter. Sorting strategy was as previously described (Tasic et al 2018). 30,000 cells were sorted within 10 minutes into a tube containing 500  $\mu$ l of quenching buffer. We found that sorting more cells into one tube diluted the ACSF in the collection buffer, causing cell death. We also observed decreased cell viability for longer sorts.

Instrument

FACSAria II or FACSAria Fusion

Software

FACSDiva v8

Cell population abundance

Abundance of sorted cell populations for 10x genomics scRNA-seq were determined by hemocytometer post-FACS.

Gating strategy

The morphology gate (SSC-A vs FSC-A) here includes all events that pass FSC threshold to allow profiling of all possible RFP+ cells. SC-FSC and SC-SSC are used to exclude doublets, and RFP+ cells are sorted from the rest of the cell based on the RFP+ DAPI- phenotype. Gating strategy for RFP+ mouse neurons is described in more detail here: Allen Institute for Brain Science 2020. FACS Single Cell Sorting. protocols.io <https://dx.doi.org/10.17504/protocols.io.be4cjgsw>.

Live cells were sorted using a similar strategy. The morphology gate includes all events that pass FSC threshold to allow profiling of all possible live cells. SC-FSC and SC-SSC are used to exclude doublets, and live cells are sorted from the rest of the cell population based on the Hoechst+ or Calcein+/Hoechst+ phenotype.

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