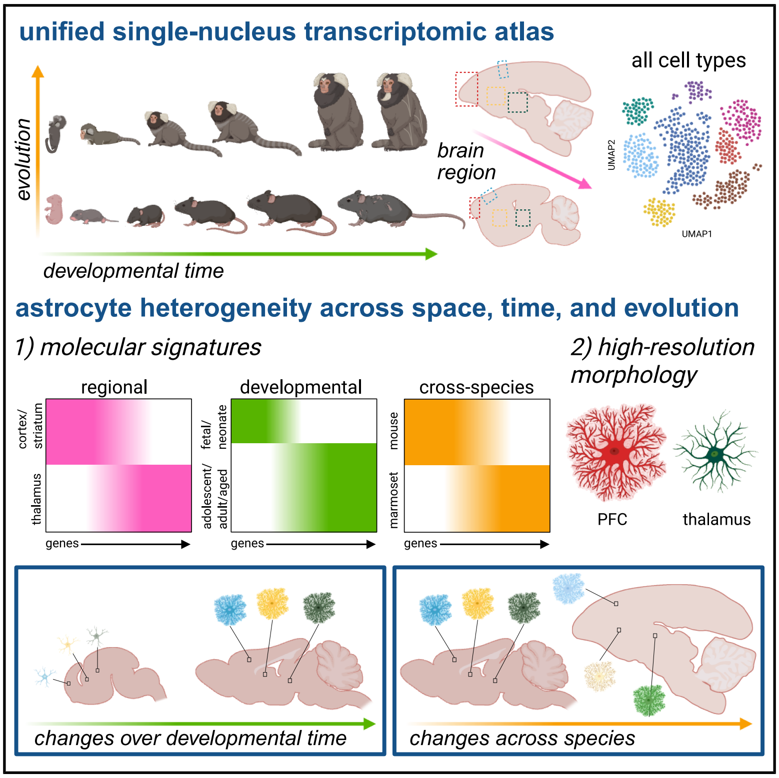
Neuron

NeuroResource

A transcriptomic atlas of astrocyte heterogeneity across space and time in mouse and marmoset

Graphical abstract



Highlights

- A transcriptomic atlas across brain regions and development in mouse and marmoset
- Astrocyte regional heterogeneity evolves over postnatal development
- Astrocyte transcriptomes are broadly conserved across species with divergent signatures
- Expansion microscopy reveals regional distinctions in astrocyte morphology

Authors

Margaret E. Schroeder,
Dana M. McCormack,
Lukas R. Metzner, ...,
Edward S. Boyden, Fenna M. Krienen,
Guoping Feng

Correspondence

fengg@mit.edu

In brief

In this NeuroResource, Schroeder et al. present a transcriptomic atlas across brain regions and developmental time points in mouse and marmoset. Detailed analysis focused on astrocytes revealed that their embryonically patterned regional heterogeneity changes significantly over the course of postnatal development, with both species conservation and divergence.

Schroeder et al., 2025, Neuron *113*, 3942–3965
December 3, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.neuron.2025.09.011

CellPress

CellPress
OPEN ACCESS

Neuron

NeuroResource

A transcriptomic atlas of astrocyte heterogeneity across space and time in mouse and marmoset

Margaret E. Schroeder,[1,2] Dana M. McCormack,[1] Lukas R. Metzner,[1] Jinyoung Kang,[1,2] Katelyn X. Li,[1] Eunah Yu,[1] Lisa Melamed,[1] Kirsten M. Levandowski,[1,4] Heather Zaniewski,[1] Qiangge Zhang,[1,4] Edward S. Boyden,[1,2,3,5,6,7,8,9] Fenna M. Krienen,[10,11] and Guoping Feng[1,2,3,4,11,12,*]
[1]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA
[2]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA
[3]Yang Tan Collective, MIT, Cambridge, MA 02139, USA
[4]The Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA
[5]Center for Neurobiological Engineering and K. Lisa Yang Center for Bionics, MIT, Cambridge, MA 02139, USA
[6]Department of Biological Engineering, MIT, Cambridge, MA 02139, USA
[7]Koch Institute, MIT, Cambridge, MA 02139, USA
[8]Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[9]Media Arts and Sciences, MIT, Cambridge, MA 02139, USA
[10]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: fengg@mit.edu
https://doi.org/10.1016/j.neuron.2025.09.011

SUMMARY

How astrocyte regionalization unfolds over development is not fully understood. We used single-nucleus RNA sequencing to characterize the molecular diversity of brain cells across six developmental stages and four brain regions in the mouse and marmoset brain. Our analysis revealed striking regional heterogeneity among astrocytes, particularly between telencephalic and diencephalic regions in both species. Most of the region patterning was private to astrocytes and not shared with neurons or other glial types. Though astrocytes were already regionally patterned in late embryonic stages, this region-specific astrocyte gene expression signature changed significantly over postnatal development, and its composition suggests that regional astrocytes further specialize postnatally to support their local neuronal circuits. Across mouse and marmoset, we found hundreds of species-differentially expressed genes and divergence in the expression of astrocytic region- and age-differentially expressed genes. Finally, we used expansion microscopy to show that astrocyte morphology is also regionally specialized.

INTRODUCTION

The mammalian brain is composed of thousands of heterogeneous, molecularly defined cell types.[1,2] This heterogeneity is prominent between cells from different anatomical regions that arise from distinct developmental compartments. This regional specialization is critical for circuit formation and proper brain function. In recent years, this heterogeneity has been cataloged through large-scale single-cell and single-nucleus RNA sequencing (scRNA-seq and snRNA-seq, respectively), which enables molecular profiling in unprecedented detail and scale.[3] The past decade has seen the publication of multiple brain cell type transcriptomic atlases, including those of the entire adult mouse brain,[1,2] the adult human brain,[4] the developing mouse[5] and human[6] brains, and the adult marmoset brain.[7,8] Most, but not all, of these atlases have focused primarily on characterizing neurons, long considered the brain's principal cell type. Indeed, several studies used cell sorting methods to enrich for neurons.[9]

Astrocytes, an abundant class of glia, play critical roles in neuronal circuit assembly and function in healthy and pathological states.[10–14] Although their morphological heterogeneity has long been appreciated,[15,16] their molecular heterogeneity, particularly across brain regions, was only revealed more recently by microarray[17] and bulk RNA-seq studies[18–20] and later by scRNA-seq studies.[8,21–25] Early lineage-tracing studies in the mouse spinal cord and brain revealed that astrocyte precursors from different embryonic domains are molecularly distinct.[26,27] Adult mouse astrocytes maintain epigenetic marks from their region-restricted radial glia ancestors,[28] which may contribute to the significant heterogeneity of adult astrocyte populations. There is also abundant evidence supporting the role of extrinsic cues in astrocyte regionalization, including the formation of various cortical morphological subtypes from a shared astrocyte progenitor[29]; the up- or downregulation of ion channels, transporters, and receptors in response to neuronal inputs[30]; and the molecular and morphological adaptation of distinct, developmentally patterned septal astrocyte subtypes after cross-region heterotopic transplant.[31]

As has been found with neurons, it is likely that transcriptionally defined astrocyte populations are developmentally influenced by their respective microenvironments and perform distinct functions. Yet, the developmental time course of astrocyte regional patterning, the composition of astrocyte subtypes over development, and the conservation of these features between rodents and primates remain unclear.[32] To address this knowledge gap, we applied snRNA-seq to characterize astrocyte molecular diversity across six developmental stages and four brain regions in mouse and marmoset. To complement the transcriptomic studies, we characterized complex astrocyte morphology and protein localization at high resolution across brain regions using expansion microscopy.

We used single-nucleus sequencing to generate a dataset of 1.4 million brain cell nuclei across multiple stages and brain regions in mouse and marmoset. A unified study, with data generated from a single lab using highly consistent methodology, has the advantage of reduced technical variation compared with datasets integrated across research groups, nuclei isolation protocols, and sequencing platforms, whose variation is difficult to remove *in silico*.[33,34]

Our analysis shows that astrocytes are regionally patterned before birth and at all subsequent time points. Importantly, we found dramatic changes in the transcriptional signatures underlying astrocyte regional identity between birth and early adolescence in both species, highlighting the importance of postnatal regional cues in shaping astrocyte identity. We explored the functional implications of genes differentially expressed between astrocytes from different brain regions and between astrocytes at different developmental time points. Furthermore, we identified both region-shared and region-divergent developmental transcriptional signatures in astrocytes.

Many of the region-, age-, and species-differentially expressed genes (rDEGs, aDEGs, and sDEGs, respectively) in astrocytes implicated morphogenesis pathways. Indeed, astrocyte morphology, which is highly ramified and complex, including sub-micron-scale processes that contact synapses and blood vessels, is essential for their many functions.[35,36] Therefore, to assess whether astrocyte morphology is also regionally specialized, we used a new variant of expansion microscopy, expansion revealing (ExR),[37] to characterize virally labeled astrocyte morphology and nanoscale protein expression with enhanced resolution. We found that gray matter thalamic astrocytes in mice were significantly smaller and less complex than their striatal and cortical counterparts and showed differences in protein expression.

RESULTS

A multi-region transcriptomic atlas of the developing mouse and marmoset brain

To create the cross-region, cross-species, cross-development snRNA-seq atlas, we dissected prefrontal cortex (PFC), motor cortex (MO), striatum, and thalamus from freshly harvested mouse and marmoset brains at late embryonic, neonatal, early adolescent, late adolescent, young adult, and aged time points and snap froze the tissue (Figure 1A). We collected tissues from 2 marmoset donors, one male and one female, at gestational day (GD)135, neonate, 7 months, 14 months, 30 months (4 donors, previously collected data in the lab), and 11+ years. For mouse, we collected tissue from 3 mouse biological replicates, at least one female, at embryonic day (E)18.5, postnatal day (P)4, P14, P32, P90, and 90 weeks (see Table S1 for mapping of donors to biological replicates; see STAR Methods). We generated single-nuclei suspensions from the snap-frozen tissue, without enriching for any particular cell type, and generated single-nucleus transcriptomes using 10× Genomics Chromium v3.1 chemistry (see Figure S1 for sequencing coverage statistics). Though the adult (4 donors aged 29–32 months, together labeled 30 months) marmoset snRNA-seq data were generated using a different nuclei isolation protocol and reference genome,[7] the data integrated very well across studies (Figure S2A). The data were also well integrated across biological sex (Figures S2 and S3A). Quantitative measures of integration quality[38,39] suggest that our integration is well mixed across biological replicates while preserving true biological variability (donor mixing = 0.8981 for mouse and 0.9155 for marmoset; neighbor consistency = 0.6680 for mouse and 0.4694 for marmoset; and average silhouette width = 0.6437 for mouse and 0.6634 for marmoset; see STAR Methods).

After rigorous quality control, including removal of ambient RNA, low-quality nuclei, and doublets (see STAR Methods), we obtained 597,668 mouse nuclei and 881,832 marmoset nuclei, which were composed of 12 broad cell classes (Figures 1Bi, 1Ci, S2, and S3). We annotated more granular cell type (Leiden[40]-determined) clusters within each cell class (Figure S4; Tables S2, S3, and S4). We employed the Allen Brain Cell Atlas's (ABCA) MapMyCells portal (https://portal.brain-map.org/atlases-and-data/bkp/mapmycells) to refine our annotations of neuronal subtypes and to help correct for modest cross-region contamination resulting from dissection error (Figures S5 and S6; see STAR Methods). Throughout the paper, "dissected" brain region refers to the original region label for the sample in which the nucleus was processed, while "region" or "assigned region" refers to the brain region assigned post hoc in the case of cells informatically predicted to arise from neighboring structures.

The total neuron-to-astrocyte ratio (across regions and developmental time points) was modestly higher in mouse (6.28) than marmoset (5.24). To more quantitatively assess cell type composition differences across dissected region (cortex, striatum, or thalamus), age, and sex, we used single-cell compositional data analysis (scCODA),[41] which implements a Bayesian model of cell type counts to address the issue of low sample sizes in snRNA-seq data. scCODA confirmed many significant differences in cell type proportion between regions and ages for each species, including the expected low numbers of excitatory neurons in striatum, increasing oligodendrocyte abundance with age in both species, and minimal sex differences in cell type composition (Figures S2C and S3C; Tables S5, S6, S7, and S8).

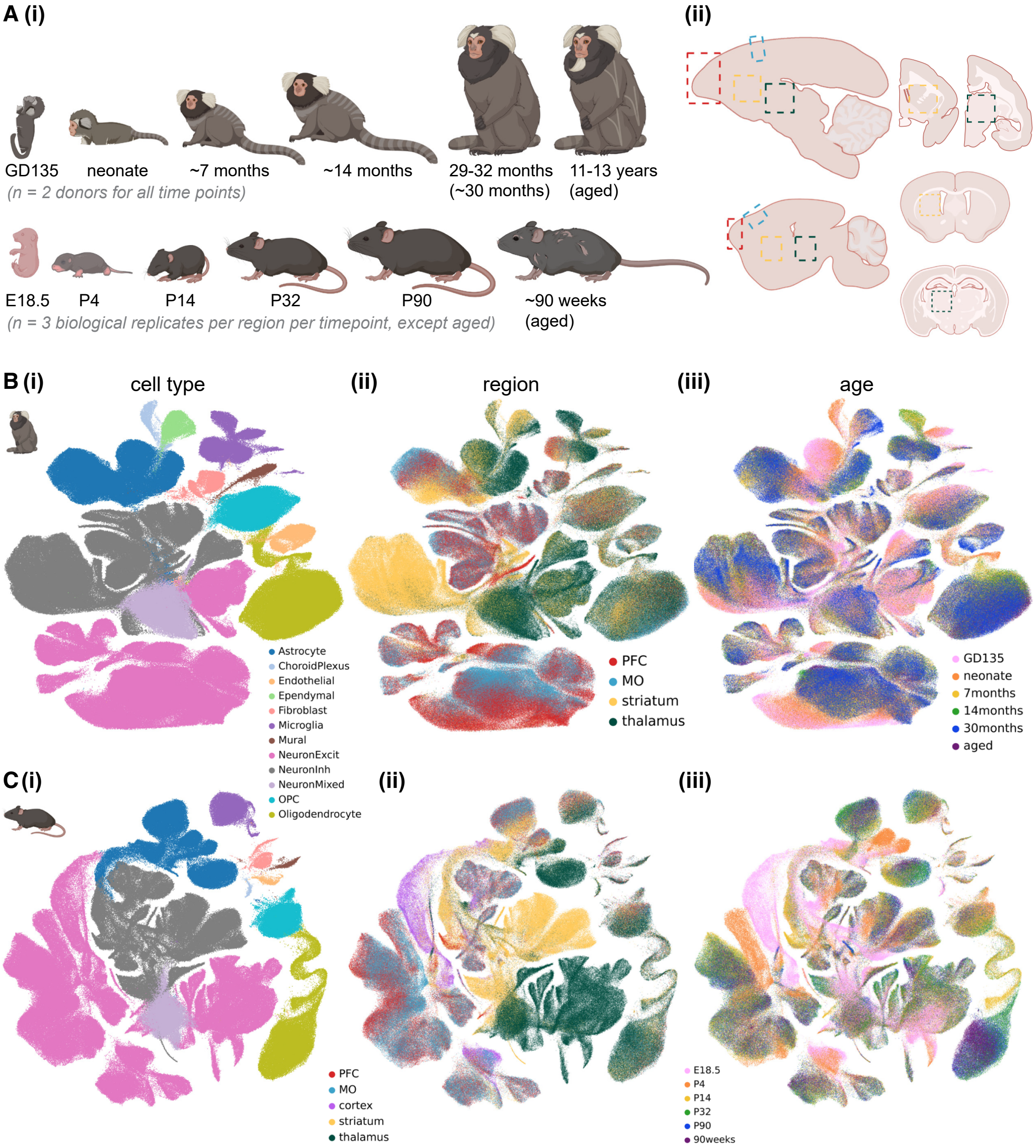


Figure 1. A multi-region transcriptomic atlas of brain cell diversity across postnatal development in marmoset and mouse

(A) Cross-development and cross-region sampling strategy in marmoset (top row) and mouse (bottom row). (i) Developmental time points profiled (some approximate, see STAR Methods and Table S1): GD, gestational day; E, embryonic day; P, postnatal day. (ii) Brain regions profiled, including prefrontal cortex (PFC, red dashed boxes), motor cortex (MO, blue dashed boxes), striatum (yellow dashed boxes), and thalamus (green dashed boxes), shown in either sagittal (left) or coronal (right, for subcortical regions only) slices. Schematics generated using BioRender.com.

(B and C) Integrated UMAP embedding of marmoset (B, 881,832 nuclei) or mouse (C, 597,668 nuclei) nuclei from PFC, MO, striatum, and thalamus across all developmental time points assayed and a randomly downsampled portion of adult nuclei from our previous study,[7] colored by (i) assigned cell type, (ii) dissected brain region, or (iii) developmental time point. Legend for (B)–(Ci) is shared.

See also Figures S1–S6.

We found several cell type clusters that were enriched or depleted in developing (late embryonic or neonate) brains (Figures S4C and S4F). For example, in both species, there were immature cortical excitatory neuron, microglia, and astrocyte clusters composed mostly of nuclei from late embryonic and fetal donors. The committed oligodendrocyte precursor (COP) and newly formed oligodendrocyte (NFOL) cluster was primarily composed of nuclei from the neonate time point in marmoset, with some nuclei even coming from late embryonic donors, but was primarily composed of nuclei from early

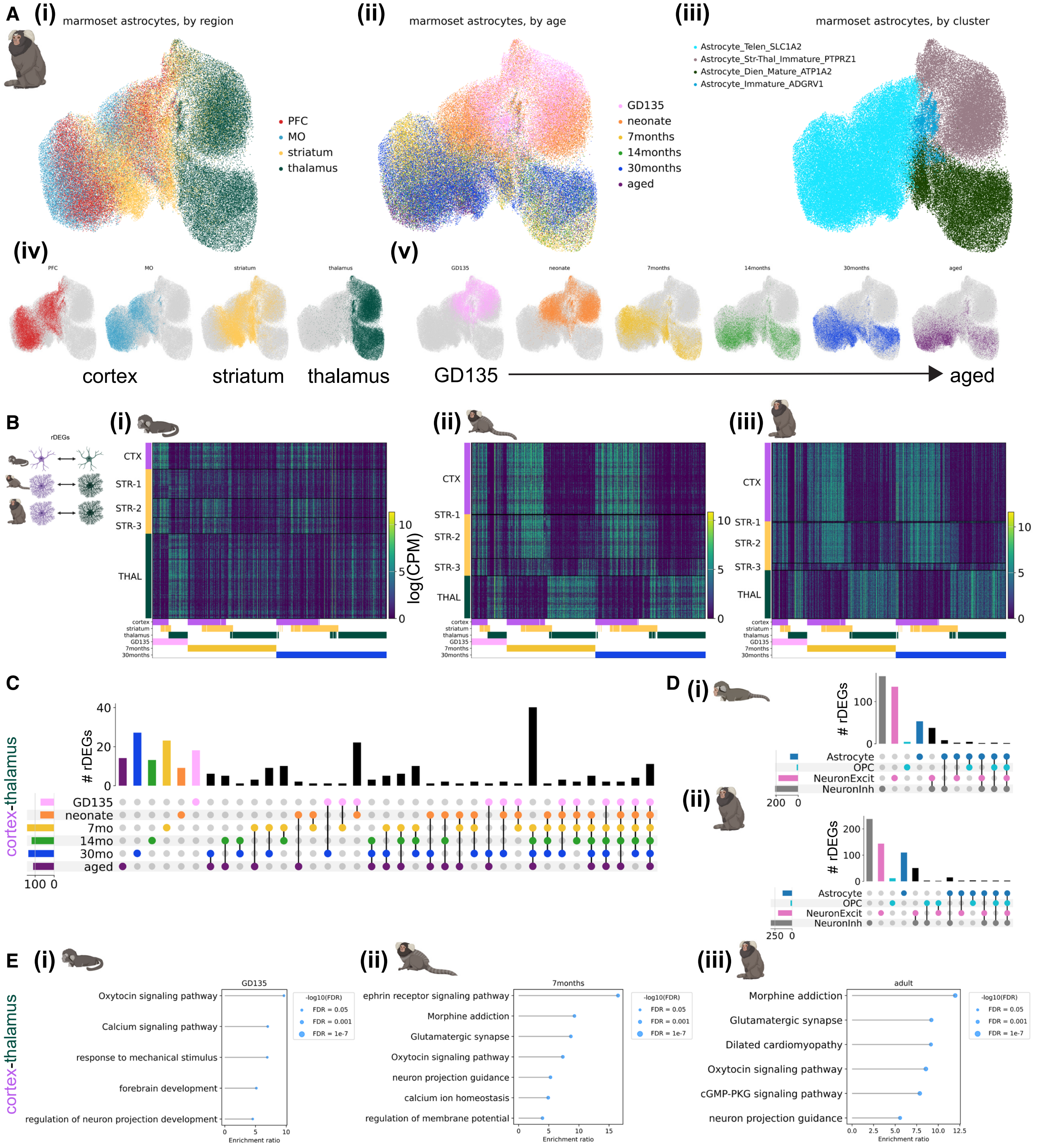


Figure 2. Developmental changes and cell-type specificity of astrocyte regional heterogeneity across postnatal development in the marmoset

(A) Integrated UMAP embeddings of 103,009 marmoset astrocytes colored by (i) assigned brain region (one-hot color encoded in iv), (ii) developmental time point (one-hot color encoded in v), and (iii) Leiden cluster assignment.

(B) Expression heatmaps (rows are cells and columns are genes) of regional differentially expressed genes (rDEGs) between astrocytes from cortex, striatum, and thalamus at (i) GD135, (ii) 7 months, and (iii) in 30-month-old marmosets in log counts per million (logCPM). The raster plots beneath each heatmap indicate the time point and region(s) of upregulation for each rDEG. rDEGs are ordered (same order in i and iii) first by time point and then by region of highest expression, and they are plotted more than once if they are present at more than one time point. Striatal astrocytes are ordered by subtype identity (see STAR Methods).

(legend continued on next page)

adolescent (P14) donors in mouse, with no COP/NFOLs coming from E18.5 mouse, indicating earlier oligodendrocyte maturation in marmoset.

Astrocyte regional heterogeneity is embryonically patterned and unfolds over postnatal development

We observed striking regional heterogeneity among astrocytes at all developmental time points sampled in both species, particularly between astrocytes of diencephalic (thalamus) and telencephalic (cortex and striatum) origin (Figures 2A and 3A). This is in line with multiple studies demonstrating embryonic regional patterning of astrocytes.[6,8,27] These regional populations further divided into an immature population, primarily composed of nuclei from late embryonic and neonatal time points, and a mature population, composed of nuclei from late adolescent time points onward. These separate populations suggest that embryonically patterned regional astrocyte populations undergo significant changes from the time of birth (neonate or P4) to early adolescence (7 months or P14). Notably, abundant populations of immature astrocytes remained present in the mouse striatum through adulthood (P90).

Marmoset

Transcription factors and morphogen gradients set up initial boundaries between developmental compartments, such as the telencephalon and the diencephalon.[42,43] It could be that such early influences are present only transiently at initial astrocyte specification or that later stages retain initial molecular distinctions and accumulate others over development. A combination of the two is also possible, where some genes follow one pattern (developmentally transient expression) or the other (sustained expression throughout the lifespan). We calculated rDEGs at each developmental time point from metacells, one-dimensional vectors of averaged normalized expression across all cells in a given grouping (see STAR Methods), of each region. For marmoset, where each donor (biological replicate) was represented in each brain region, we calculated rDEGs separately for each donor and required that rDEGs be above threshold (minimum expression and log fold change) requirements in both donors. We found 70 rDEGs whose expression differed between fetal cortical and thalamic astrocytes, 142 of such rDEGs by early adolescence (7 months), and 134 in adulthood (30 months, see Figures 2Bi–2Biii for the expression pattern of the union of these rDEGs at each time point). Focusing on cortex-thalamus rDEGs (which were most numerous, Table S9), we found that relatively few persisted across all developmental time points (Figure 2C). 50% were shared between fetal and neonate, and 51% between late adolescent and aged, but only 4% continued to act as regional patterning signatures throughout their lifespan. We found that many more rDEGs were shared between late adolescent, young adult, and aged time points (61) than between fetal, neonate, and 7-month time points (23) and neonate, 7-month, and 14-month time points (25, Figure 2C). These data suggest that regional astrocyte gene expression signatures emerge in the embryonic brain, change drastically over the course of early postnatal development, and stabilize during adolescence into adulthood.

If early telencephalic and diencephalic patterning persists in astrocytes, cortex and striatum should retain common rDEGs compared with thalamus. To assess the degree of pairwise astrocyte rDEG sharing across the 3 brain structures, we correlated the log fold-change differences in astrocyte regional gene expression between different region pairs (e.g., cortex-striatum vs. cortex-thalamus) for both rDEGs (log fold change > 0.5) and non-rDEGs at fetal, early adolescent, and adult time points. We found that cortex-striatum vs. cortex-thalamus fold changes exhibited a high degree of correlation in fetal marmoset (Pearson's r = 0.78), which decreased over developmental time (Pearson's r = 0.47 in adult marmoset, Figures S7A–S7Ci). At all 3 time points, striatum-thalamus vs. striatum-cortex fold changes were negatively correlated or uncorrelated (r = −0.43, −0.26, and 0.04 at GD135, 7 months, and 30 months, respectively, Figures S7A–S7Cii). Finally, thalamus-striatum vs. thalamus-cortex fold changes were highly correlated (r > 0.80) at all 3 time points (Figures S7A–S7Ciii). Together, these results suggest that cortical and striatal astrocytes share transcriptional divergence from thalamic astrocytes at all ages but become more transcriptionally similar later in development. At the same time, cortical and thalamic astrocytes both diverge from striatal astrocytes but in distinct ways, as indicated by the negative correlation between striatum-thalamus and striatum-cortex fold changes at GD135 and 7 months and the lack of correlation at 30 months.

Dorsal radial glia populate the neocortex in a stereotyped progression, giving rise first to glutamatergic neurons, then to astrocytes, and finally to oligodendrocytes.[44–46] In the developing thalamus, radial glial progenitors likely follow the same cell type sequence.[46,47] As we showed previously for variable genes across the adult neocortex,[7] far more adult cortex-thalamus rDEGs are private to astrocytes than are shared with neurons or OPCs, despite their shared lineage[44] (Figure 2D). Surprisingly, this remained true even at the earliest stages we sampled (GD135 and neonate). Together, these observations suggest that astrocytes gain regional identity early in their maturation but that their continued regional identity is facilitated by distinct genes across their lifespan (Figure 2C).

Many rDEGs nominate core cellular functions that may be further regionally specialized in astrocytes. For example, ephrins

---

(C) UpSet plot showing the number of unique and overlapping cortex-thalamus rDEGs between developmental time points. The colored dots below each vertical bar indicate which age(s) share that set of rDEGs, whereas the colored horizontal bars indicate the total number of cortex-thalamus rDEGs for each age. Overlap categories with 0 rDEGs are not shown.

(D) UpSet plot (as in C) showing the number of overlapping cortex-thalamus rDEGs between OPCs, astrocytes, excitatory neurons, and inhibitory neurons for (i) neonate and (ii) adult marmoset.

(E) Gene Ontology (GO) and pathway analysis on cortex-thalamus (enriched in either region) astrocyte rDEGs via WebGestalt 2024 in (i) GD135, (ii) 7-month, and (iii) 30-month marmoset astrocytes. Lollipop plots show the enrichment ratio of GO biological process and KEGG pathways from an over-representation analysis, with tip size inversely proportional to the false discovery rate (FDR).

See also Figures S7–S9 and S12.

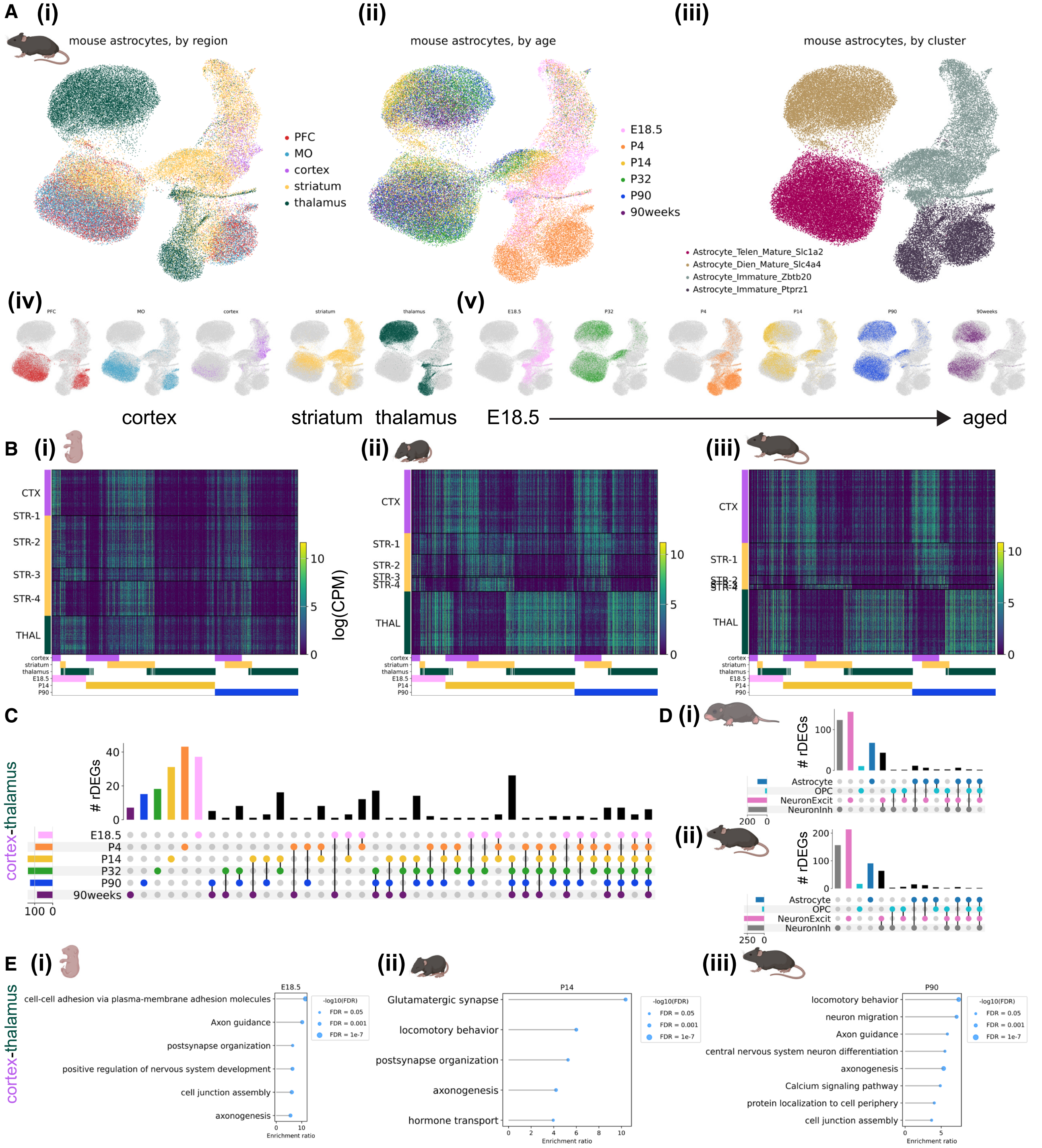


Figure 3. Developmental changes and cell-type specificity of astrocyte regional heterogeneity over postnatal development in the mouse

(A) Integrated UMAP embeddings of 68,485 mouse astrocytes colored by (i) brain region (one-hot color encoded in iv), (ii) developmental time point (one-hot color encoded in v), and (iii) Leiden cluster assignment.

(B) Expression heatmap (rows are cells and columns are genes) of regional differentially expressed genes (rDEGs) between astrocytes from cortex, striatum, and thalamus at (i) E18.5, (ii) P14, and (iii) P90 mice in log counts per million (logCPM). The raster plots beneath each heatmap indicate the time point and region(s) of upregulation for each rDEG. Genes are ordered as in Figure 2B. Striatal astrocytes are ordered by subtype identity (see STAR Methods).

(C) UpSet plot showing the number of overlapping cortex-thalamus rDEGs between developmental time points, as in Figure 2C.

(legend continued on next page)

such as *EFNB2* (identified as an rDEG in 30-month marmoset astrocytes) and *EFNA5* (7, 14, and 30 months and aged) are upregulated in cortical astrocytes, and the ephrin receptor *EPHB1* (7, 14, and 30 months and aged) is upregulated in cortical and striatal astrocytes. Neuron-astrocyte signaling via ephrin ligands and receptors regulates axon guidance and synaptogenesis.[48] Thus, neuron-astrocyte ephrin signaling may be specialized in the telencephalon. Cyclic-AMP (cAMP)-related signaling molecules *ADCY1* (30 months and aged) and *ADCY8* (identified at all ages) are upregulated in thalamic astrocytes. As in neurons, astrocytic cAMP is an important second messenger following G-protein-coupled receptor (GPCR) activation[49] and modulates synaptic plasticity.[50] *ITPR1* (7, 14, and 30 months and aged), a calcium channel that controls calcium release from the endoplasmic reticulum, an important source of intracellular calcium during astrocyte signaling,[51] is also upregulated in thalamic astrocytes. These rDEGs suggest that thalamic astrocytes may have developed specialized pathways for calcium and cAMP signaling, potentially in response to the release of upstream GPCR ligands by thalamus-projecting and thalamic neurons. Additionally, astrocyte rDEGs included ion channels (e.g., *TRPM3* [GD135; 7, 14, and 30 months; and aged], a non-selective Ca^{2+}-permeable ion channel and thalamic rDEG); synapse-related proteins (e.g., *SPARC*, a thalamic rDEG, at 7, 14, and 30 months and aged, which regulates synaptogenesis[52]); neurotransmitter transporters and receptors (e.g., *SLC6A11*/GAT3 [7, 14, and 30 months and aged], a thalamic rDEG and γ-aminobutyric acid [GABA] transporter; *SLC1A3*/GLAST [at GD135 and neonate], a glutamate transporter higher in the cortex and commonly used astrocyte marker gene; and *GRM3*/mGluR3 [all ages except GD135], a telencephalic rDEG and metabotropic glutamate receptor); and a thyroid hormone receptor (*SLCO1C1* [GD135, neonate, and 30 months], a cortical and later striatal rDEG). These rDEGS point more directly to astrocyte adaptation to the local synaptic and neuronal niche.

To characterize astrocyte rDEG pathways in a more unbiased manner, we used WebGestalt 2024[53,54] over-representation analysis to test for enrichment of cortex-thalamus rDEGs (bidirectionally, i.e., upregulated in either cortex or thalamus) in Gene Ontology (GO) biological process and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Enriched pathways implicated oxytocin and calcium signaling and neuronal projection development for GD135 astrocyte rDEGs; ephrin signaling, synaptic transmission, and calcium ion homeostasis for 7-month astrocyte rDEGs; and glutamatergic synaptic transmission, oxytocin signaling, and cGMP-protein kinase G (cGMP-PKG) signaling for adult marmoset astrocyte rDEGs (Figure 2E). A summary of WebGestalt results for cortex-thalamus astrocyte rDEGs at each age is provided in Table S10. To complement this pathway analysis and facilitate exploration of rDEG functions, we queried UniProt[55] for each rDEG (see STAR Methods) to return its full protein name, GO cellular compartment, GO molecular function, and GO biological process annotation(s). These annotations are included in Table S9 for marmoset and Table S10 for mouse. Together, these results suggest that astrocytes are regionally specialized with varied physiological adaptations necessary to support neuronal transmission and activity in their local environment.

Compared with cortex-thalamus expression differences, there were many fewer cortex-striatum rDEGs (0 in fetal marmoset, 12 in neonate, 8 in 7-month, 9 in 14-month, 25 in 30-month, and 12 in aged marmoset). At neonate as well as 7- and 14-month time points, at least half of these cortex-striatum rDEGs overlapped with cortex-thalamus rDEGs. At 14 months, these overlapped genes included *MYO16*, an unconventional myosin protein implicated in neurodevelopment[56] (higher in striatum and thalamus); *UNC5C*, a netrin receptor family member involved in axon guidance[57,58] (higher in cortex); *MAPK10*, a mitogen-activated protein kinase (higher in striatum and thalamus); *STXBP6*, a syntaxin-binding protein that is part of the soluble NSF attachment protein receptor (SNARE) complex in neurons (higher in striatum and thalamus); *GRIK2*, a kainate-type ionotropic glutamate receptor subunit (higher in striatum and thalamus); *EYA2*, a transcriptional coactivator and phosphatase (higher in striatum and thalamus); *DYNC1I1*, a member of the cytoplasmic dynein 1 complex involved in intracellular transport (higher in striatum and thalamus); and *PTPRE*, a protein tyrosine phosphatase family member involved in cell signaling with various downstream consequences (higher in cortex). Each of these genes points to a biological process, such as glutamate sensing, phosphorylation, and exocytosis, for which striatal and thalamic astrocytes may be differentially invested.

Mouse

We calculated mouse astrocyte rDEGs using the previously described metacell method on expression data pooled across all biological replicates (see STAR Methods). As with marmoset, mouse astrocyte gene expression varied across developmental time points, and most astrocyte rDEGs were not shared with other cell types (Figures 3A–3D; Table S11). E18.5 astrocyte cortex-thalamus rDEGs (79 total) included *Cacna2d1*, *Cntn5*, *Nrxn1*, *Creb5*, *Slco1c1*, and *Slc6a11*, and together they were enriched for biological processes including cell-cell adhesion, axon guidance, and postsynaptic organization (Figure 3Ei; Table S12). P14 astrocyte cortex-thalamus rDEGs included 26 of the rDEGs present at E18.5 (19% of total P14 rDEGs), in addition to rDEGs that only emerged at P14. These included voltage-gated calcium channel subunit *Cacna1a*, the glutamate-gated kainate receptor *Grik4*, the N-glycoprotein *Thsd7a*, the cholesterol transporter *Gramd1b*, and the inward-rectifying potassium channel *Kcnj6*. Together, P14 cortex-thalamus astrocyte rDEGs were enriched in glutamatergic synapse, hormone transport, and postsynaptic organization pathways (Figure 3Eii; Table S12). There were 124 cortex-thalamus astrocyte rDEGs at P90, which included many of the rDEGs present at earlier time points (17% of P90 rDEGs

---

(D) UpSet plot (as in C) showing the number of overlapping cortex-thalamus rDEGs between OPCs, astrocytes, inhibitory neurons, and excitatory neurons for (i) P4 and (ii) P90 mouse. Overlap categories with 0 rDEGs are not shown.

(E) Gene Ontology (GO) and pathway analysis on cortex-thalamus (enriched in either region) astrocyte rDEGs via WebGestalt 2024 in (i) E18.5, (ii) P32, and (iii) P90 mouse astrocytes. Lollipop plots were generated as in Figure 2E.

See also Figures S7, S10, S11, and S13.

were present at E18.5 and 52% were present at P14), and they were enriched in neuron migration, axon guidance, calcium signaling, and cell junction pathways (Figure 3Eiii; Table S12). Compared with marmoset, mouse astrocytes had more cortex-striatum astrocyte rDEGs throughout development, especially at P14 (13 at E18.5, 24 at P4, 63 at P14, 15 at P90, and 33 at 90 weeks). However, at most 15 of these (at P14) overlapped with cortex-thalamus rDEGs, suggesting a more distinct transcriptional niche for mouse striatal astrocytes compared with cortex, as explored below.

As we did with marmoset, we assessed in mouse the degree to which regional imprinting of astrocyte gene expression persists across development by correlating log fold-change gene expression differences between region pairs at E18.5, P14, and P90. We found that cortex-striatum vs. cortex-thalamus fold changes exhibited a high degree of correlation at E18.5 (Pearson's r = 0.85), which decreased dramatically at P14 (r = 0.26) and increased again at P90 (r = 0.40, Figures S7D–S7Fi). Compared with marmoset, striatum-cortex vs. striatum-thalamus rDEGs were more positively correlated at juvenile and adult stages (P14 [r = 0.48] and P90 [r = 0.19, Figures S7D–S7Fii]). Consistent with marmoset, thalamus-striatum vs. thalamus-cortex fold changes were highly correlated ($r > 0.70$) at all 3 time points (Figures S7D–S7Fiii). Together, these results suggest that whereas cortical and striatal astrocytes are similarly divergent from thalamic astrocytes (likely reflecting the their distinct telencephalic-diencephalic origins), mouse striatal astrocytes develop and maintain a unique transcriptional signature distinct from cortex and thalamus.

To validate the existence of these regional astrocyte populations and the differential expression of selected rDEGs *in situ*, we conducted multiplexed RNA fluorescence *in situ* hybridization (FISH) using the RNAscope platform (Advanced Cell Diagnostics, see STAR Methods) in neonate and adult animals of both species. We used CellProfiler 4.2.5[59,60] to quantify the fraction of astrocytes positive for each target gene in each region and the fraction of each astrocyte nucleus covered by the probe for each target gene in each region (referred to as mean intensity, see STAR Methods). Most rDEGs followed the expected regional and developmental expression pattern in marmoset astrocytes, including *SPARC*, which was enriched in diencephalic astrocytes (more so in adulthood); *FOXG1*, which marked telencephalic astrocytes; *GFAP*, which was elevated in thalamus in adult but not neonate; and *KCNH7*, which was a telencephalic rDEG in neonate but not adult (Figures S8 and S9; see STAR Methods for a discussion of the few rDEGs whose *in situ* expression differed from snRNA-seq predictions).

Similarly, we found that most mouse astrocyte rDEGs followed the expected regional and developmental expression pattern in P4 and P90 mouse astrocytes, including *Clmn*, *Slco1c1*, *Csmd1*, *Sparc* (in adult mouse), and *Kcnd2* (Figures S10 and S11; see Table S13 for source data and statistics). For additional validation, we analyzed the differential expression of our selected mouse rDEGs in the Allen Mouse Brain Cell Atlas whole-brain MERSCOPE v1 dataset[1] and found it to be largely consistent with our snRNA-seq data (Figure S11B; see STAR Methods). This was also true for the whole list of P90 mouse rDEGs, which also showed differential expression across a wider selection of brain regions (Figure S11C).

To assess the extent of astrocyte intra-regional heterogeneity, we performed subclustering on cortical, striatal, and thalamic astrocytes, from all developmental time points, separately for each species (Figures S12 and S13; see STAR Methods). We found at least 4 astrocyte subclusters within each region, which primarily distinguished protoplasmic and fibrous/interlaminar subtypes (the latter being identified by *GFAP*, *AQP4*, and/or *ID3* expression[8,21,61,62]) and immature and mature astrocytes. In both species, the majority of astrocytes in the cortex and striatum were protoplasmic. In the marmoset thalamus, a larger proportion of astrocytes were *GFAP*+, *AQP4*+, or *ID3*+ (Figure S12C; Table S14), suggesting a higher proportion of fibrous astrocytes, consistent with the greater abundance of white matter in this region (Figures S2B and S2C). Nevertheless, it is unclear the extent to which the definitions of protoplasmic, fibrous, and intralaminar apply outside of the cortex. The mouse striatum had the most intra-regional heterogeneity, with 12 subclusters (Figure S13B), in large part due to immature populations, including *Top2a*+ rostral migratory stream progenitors. As in previous studies, we found that *CRYM*/*Crym* marks a subset of striatal astrocytes[19] and *SPARC*/*Sparc* marks thalamic astrocytes in both species.[18] Several of our subclusters mapped specifically (>70% of cells in the subcluster) to a single adult mouse ABCA cluster, though many mapped to several ABCA clusters, especially immature and mixed fibrous/protoplasmic subclusters (Table S15). Taken together with our FISH data, which was obtained in gray matter regions, our subclustering analysis suggests that most of the astrocytes in our study, and therefore likely most of the resulting rDEGs, arise from protoplasmic or gray matter astrocytes.

Shared and subtype-specific predicted mechanisms of neuron-astrocyte communication

Many of the astrocyte rDEGs implicated neuron-astrocyte communication, suggesting that the regional molecular identity of astrocytes may arise in part from customized interactions with the vast diversity of specialized neuronal types across the mammalian brain.[1,4,63] Our previous analysis showed that rDEGs are not substantially shared across neurons and glia (Figures 2D and 3D), which rules out the influence of pan-cell type regional patterning. Neurons and astrocytes communicate via myriad signaling pathways. We assessed whether neuron and astrocyte cluster pairs sampled from the same region were over-enriched for known ligand-receptor (L-R) interactions by analyzing the data with CellPhoneDB.[64] To increase the specificity of our predicted L-R results, we restricted CellPhoneDB analysis to neurons and astrocytes only (see STAR Methods).

Across most brain regions and ages in the marmoset, we found neurexin and neuroligin (NRXN/NLGN) family members, contactin (CNTN) family members, fibroblast growth factor and receptor (FGF/FGFR) family members, and neural cell adhesion molecule (NCAM) family members to be the most enriched predicted neuron-astrocyte (N-A) and astrocyte-neuron (A-N) L-R molecules (Table S16). Despite the commonality of these L-R pairs between astrocytes and all neuronal subtypes, for each neuronal subtype in a given region, we found unique or

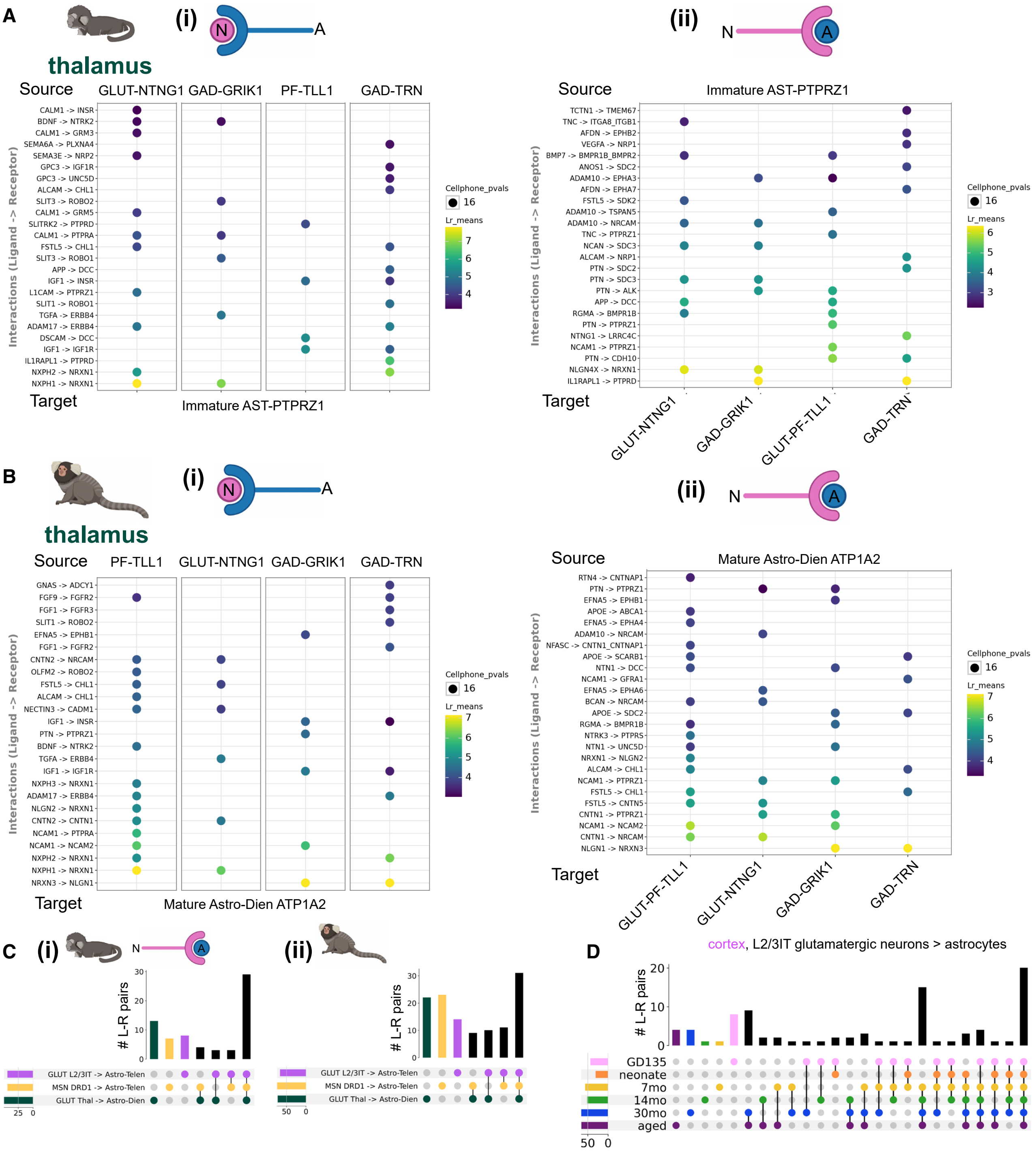


Figure 4. Cell-cell communication analysis for neuron-astrocyte and astrocyte-neuron-predicted ligand-receptor pairs across regions and developmental time points in marmoset

(A) Dot plot showing magnitude and specificity of the top 25 near-unique (shared with at most one other neuronal cluster) CellPhoneDB-predicted (i) neuron-astrocyte (N-A) and (ii) astrocyte-neuron (A-N) ligand-receptor (L-R) pairs for the most abundant astrocyte and neuronal Leiden clusters in the fetal marmoset thalamus. The source cell expresses the ligand (left side of arrow on the row labels), whereas the target cell expresses the receptor (right side of arrow). The color of the dot indicates L-R expression magnitude ("Lr_means," see STAR Methods), while the size of the dot is inversely related to the *p* value on L-R expression sensitivity ($-\log10(p)$).

(B) Same as (A) for the 14-month marmoset thalamus.

(legend continued on next page)

near-unique L-R and R-L pairs with astrocytes. For example, in the fetal marmoset thalamus, *SLT3→ROBO2* is specific to midbrain-derived *GRIK1*+ thalamic inhibitory neurons and immature thalamic astrocytes, whereas *AFDN→EPHA7* signaling is specific to immature astrocytes and thalamic reticular nucleus (TRN) GABAergic neurons, compared with the other neuronal subtypes examined (Figure 4A). Later in development, at 14 months, many of the same A-N and A-N L-R combinations were present, while some new pairs, such as *EFNA5→EPHB1* for thalamic astrocytes to *GRIK1*+ midbrain-derived GABAergic neurons, emerged (Figure 4B; see Table S17 for quantification of the fraction of L-R pairs shared across 3 or more ages for all neuronal clusters within a region for both species).

To summarize the shared and divergent expression of predicted L-R pairs underlying neuron-astrocyte communication across regions, we examined the overlap of these pairs for the most abundant neuronal cluster and the most abundant astrocyte cluster in cortex, striatum, and thalamus (Figure 4C). We found that many L-R pairs were shared across regions at both GD135 and 14 months (43% and 29% of total L-R pairs, respectively), wheras the thalamus (at GD135), and later the striatum (at 14 months), had the most L-R pairs not shared with other regions for the clusters examined. To exclude the possibility that the regional specificity of N-A L-R pairs is due solely to neuronal heterogeneity, we performed analyses examining the magnitude and specificity of L-R pairs between local (non-projecting) neurons and regional astrocyte populations, the overlap of N-A and neuron-OPC L-R pairs within a region, and the proportion of astrocyte subtype DEGs overlapping with ligand (when source) or receptor (when target) lists compared with neurons. We found, based on transcriptomic analysis, that the regional specificity of N-A L-R pairs is not solely explained by neuronal heterogeneity, reflecting a contribution of astrocyte regional heterogeneity (for detailed results, please see the STAR Methods section and associated notebooks in our GitHub repository). We note that experimental manipulation, such as neuronal ablation and/or reprogramming, is required to definitively support this conclusion. In several cases, different members of the same family were used as region-specific N-A/A-N L-R pairs in different regions. For example, in 14-month marmoset, *EFNA5→EPHA5* was unique to cortical astrocytes→cortical layer 2/3 intratelencephalic (L2/3IT) glutamatergic neurons, whereas *EFNA5→EPHA7* was unique to striatal astrocytes→*DRD1*+ medium spiny neurons, and *EFNA5→EPHA6* was shared across all three regional A→N subtype pairs.

To assess how the expression of L-R pairs underlying N-A communication changes over the course of development, we examined the concordance of L-R pairs of a single neuronal cluster (cortical glutamatergic L2/3IT neuron) and cortical astrocyte at different developmental time points. In contrast to the expression of rDEGs, a larger proportion (20/90, 22%) of L-R pairs (all L-R pairs meeting minimum expression criteria, including pairs shared with other neuronal and astrocyte clusters, were included in this analysis) were shared between all time points (Figure 4D), suggesting that these putative mediators of N-A communication emerge early and are maintained throughout development. However, 15/90 (17%) L-R pairs emerged at 7 months and were maintained throughout adulthood. At later developmental time points, many more L-R pairs were shared between ages than were unique (Figure 4D). This, along with the increased proportion of rDEGs shared across later time points (Figure 2C), suggests that the expression of molecules underlying N-A communication stabilizes postnatally in marmosets at some point between 0 and 7 months.

Overall, patterns of predicted neuron-astrocyte and astrocyte-neuron communication in the mouse were similar to the marmoset, including the involvement of the NRXN/NLGN, CNTN, FGF/FGFR, and NCAM families (Figures S14A and S14B; Table S16). As in marmoset, mouse thalamus had more unique N-A L-R pairs than cortex or striatum at P4 and P90 (Figure S14C). Unlike in marmoset, mice had more age-specific L-R pairs (from cortical L2/3IT glutamatergic neurons to cortical astrocytes) at earlier time points (particularly at P4, with 17/84 or 20% of all L-R pairs unique at this time point), before stabilizing with more shared L-R pairs at later time points. This suggests that mediators of mature N-A interactions emerge relatively later in mouse (Figure S14D). Only 11% of L2/3IT→astrocyte L-R pairs were shared across the lifespan. Taken together, these results suggest that many L-R pairs potentially underlying neuron-astrocyte communication are shared across developmental time points and regions in both species. However, more neuron-astrocyte-predicted L-R pairs emerged later in development.

Age-dependent refinement of astrocyte identity

In mouse, initiation of gliogenesis in the diencephalon precedes that in the telencephalon by approximately 1 GD (E13.5 vs. E14.5[65,66]). To determine whether relative immaturity of telencephalic glia compared with diencephalic glia could explain the robust regional expression differences we observed at each sampled time point (Figures 2 and 3), we examined the developmental trajectory of astrocytes in pseudotime—a prediction of position along a low-dimensional developmental trajectory based on RNA expression only—using Palantir.[67] Palantir recovered the known developmental trajectory of the oligodendrocyte lineage[68] in both species (Figures S15A and S15B). Furthermore, it underscored the precocious myelination in the marmoset brain compared with mouse, as evidenced by a faster rate of pseudotime progression toward maturity and a larger proportion of newly formed and myelinating

---

(C) UpSet plot showing the number of overlapping N-A-predicted L-R pairs between regions, from the most abundant neuronal and astrocyte subtypes in each region for (i) fetal and (ii) late adolescent marmoset. For cortex, glutamatergic L2/3IT neurons to cortical astrocytes; striatum, *DRD1*+ medium spiny neurons to striatal telencephalic astrocytes; and thalamus, thalamic glutamatergic neurons to thalamic astrocytes. The colored dots below each vertical bar indicate which regional N-A subtypes share that set of L-R pairs, whereas the colored horizontal bars indicate the total number of L-R pairs for each N-A subtype. All L-R pairs meeting minimum expression criteria, including pairs shared with other neuronal and astrocytic clusters, were included here and in (D).

(D) UpSet plot (as in C) showing the number of overlapping cortical glutamatergic L2/3IT neurons to cortical-astrocyte-predicted L-R pairs between ages. See also Figure S14.

oligodendrocytes at earlier time points in marmoset (Figures S15C and S15D).

In astrocytes from both species, pseudotime analysis with separate terminal states for mature telencephalic ("AST-TE" branch) and diencephalic ("AST-DI" branch) astrocytes revealed a transcriptional developmental trajectory within astrocytes that aligned with actual age and annotation of mature and immature Leiden clusters (Figures 5A, 5B, S15E, and S15F). Pseudotime values were slightly higher in mature diencephalic versus mature telencephalic astrocytes, and higher in cortical astrocytes than striatal astrocytes, in both species at mature time points, suggesting arrival at distinct terminal states, though these need not be more or less mature than one another (Figures 5A, 5B, S15E, and S15F). The rate of maturation (that is, distribution of pseudotime values during development relative to those in adulthood) also differed slightly across regions. To identify genes potentially driving pseudotime transitions, we used Mellon[69] to calculate gene change scores, a measure of expression change in regions of low cell-state density (see STAR Methods), for each pseudotime trajectory branch. Many of the top 25 change-scoring genes in both branches were cortex-thalamus rDEGs (Figures S15C and S15D; Table S18), suggesting that the expression of region-enriched genes is correlated with, and potentially drives, astrocyte maturation.

Next, for each region, we binned marmoset astrocytes by pseudotime quintile in the appropriate trajectory branch rather than by actual age and recomputed rDEGs. As with rDEGs grouped by actual age, the number of rDEGs increased from pseudotime bin 1 (PT1) to pseudotime bin 5 (PT5): 54 rDEGs at PT1, 108 at PT2, 117 at PT3, 177 at PT4, and 181 at PT5. We found that matching by predicted maturational stage largely recapitulated the original rDEGs calculated from actual age: 56% of PT1 cortex-thalamus rDEGs overlapped with GD135 rDEGs, 81% of PT2 rDEGs overlapped with neonate rDEGs, 39% of PT3 rDEGs overlapped with 7-month rDEGs, 85% of PT4 rDEGs overlapped with 14-month rDEGs, 81% of PT5 rDEGs overlapped with adult rDEGs, and 85% of PT5 rDEGs overlapped with aged rDEGs. This suggests that regional imprinting of astrocytes is not simply driven by relative differences in the birth timing of cells across the different brain structures.

We next sought to determine the sequence of molecular changes that unfold in astrocytes within a given region over time. We calculated aDEGs within each brain region from metacells of each age (see STAR Methods). In each brain region, there were over 100 unique aDEGs (unique after pooling pairwise age combinations). In marmoset, the largest fraction of aDEGs distinguished GD135/neonate from 7-month and older astrocytes (Figure 5C). aDEGs enriched in the 30-month dataset could conceivably arise from the different sample preparation and reference genome used in our previous study[7]; for this reason, we used the 14-month time point to further assess age-related changes across regions.

Examining the overlap of marmoset astrocyte GD135 vs. 14-month aDEGs between brain regions (409 aDEGs in total), we found that ~19% were shared between cortex, striatum, and thalamus (Figure 5D). The striatum had a modest number of GD135 vs. 14-month aDEGs not shared with other regions (21/409), whereas the cortex had 3-fold more (63/409), and the thalamus had the most (156/409)—as expected, given the stark regional heterogeneity between telencephalon and diencephalon (Figure 2). Additionally, we found that few to no GD135 vs. 14-month aDEGs were shared between astrocytes, OPCs, and excitatory neurons—or astrocytes, OPCs, and GABAergic neurons—in the cortex (Figure 5E).

We found similar results in mice, where we calculated P4 vs. P90 rDEGs, as mouse E18.5 astrocytes were transcriptionally immature relative to marmoset GD135 astrocytes, and GD135-P4 time points appear to have better correspondence (as discussed in the next section). One notable difference from marmoset was that early adolescent (P14) astrocytes in mice expressed many aDEGs shared with embryonic and neonate time points, particularly in striatum (Figure S16A). As in marmoset, thalamic astrocytes had more unique aDEGs than their cortical and striatal counterparts (Figure S16B), and most astrocyte aDEGs were not shared with other radial glia-derived cell types (Figure S16C).

We found very few (3 or less) astrocyte aDEGs that were cell type agnostic and region specific (i.e., that were also aDEGs in neurons and OPCs for a given brain region, see STAR Methods). In contrast, there were 74 (marmoset GD135 vs. 14-month) and 56 (mouse P4 vs. P90) aDEGs that were astrocyte specific and region agnostic, reflecting more universal aspects of astrocyte transcriptional maturation, regardless of brain region. We found a similar number of astrocyte-specific, region-specific aDEGs (20 in striatum, 51 in cortex, and 135 in thalamus for GD135 vs. 14-month marmoset), which reflects the brain region's influence on the maturation of astrocytes *only* in a given brain region. In both species, the developmental pattern of selected astrocyte-specific, cortex-specific aDEGs was similar—but not identical—in the striatum, being more dissimilar with the thalamus (Figures 5F and S16D).

Conservation and divergence of astrocyte patterning in mouse and marmoset

Hundreds of differentially expressed genes distinguish adult human and mouse astrocytes,[70] and engrafting human glial progenitors into mouse brain results in mature astrocytes that retain certain human-specific astrocyte characteristics.[71] This suggests that aspects of an astrocyte's developmental program are cell intrinsic and are shaped by its species-specific genomic features. We therefore aimed to compare transcriptional signatures of telencephalic and diencephalic regional astrocyte populations between marmoset and mouse. We integrated a randomly downsampled subset (100,000 nuclei each) of mouse and marmoset nuclei (all cell types included). To do so, we used 547 highly variable 1:1 ortholog genes selected from top differentially expressed genes of superclusters (related groups of Leiden clusters) shared across species and employed the semi-supervised variational auto-encoder scANVI[72] (see STAR Methods). The resulting integrated uniform manifold approximation and projection (UMAP) plot showed broad conservation of superclusters between mouse and marmoset, despite differences in cell type proportions across development (Figures S17A and S17B). Another method called SATURN,[73] which avoids 1:1 mapping of genes, had largely concordant cross-species integration results (Figures S17C and S17D).

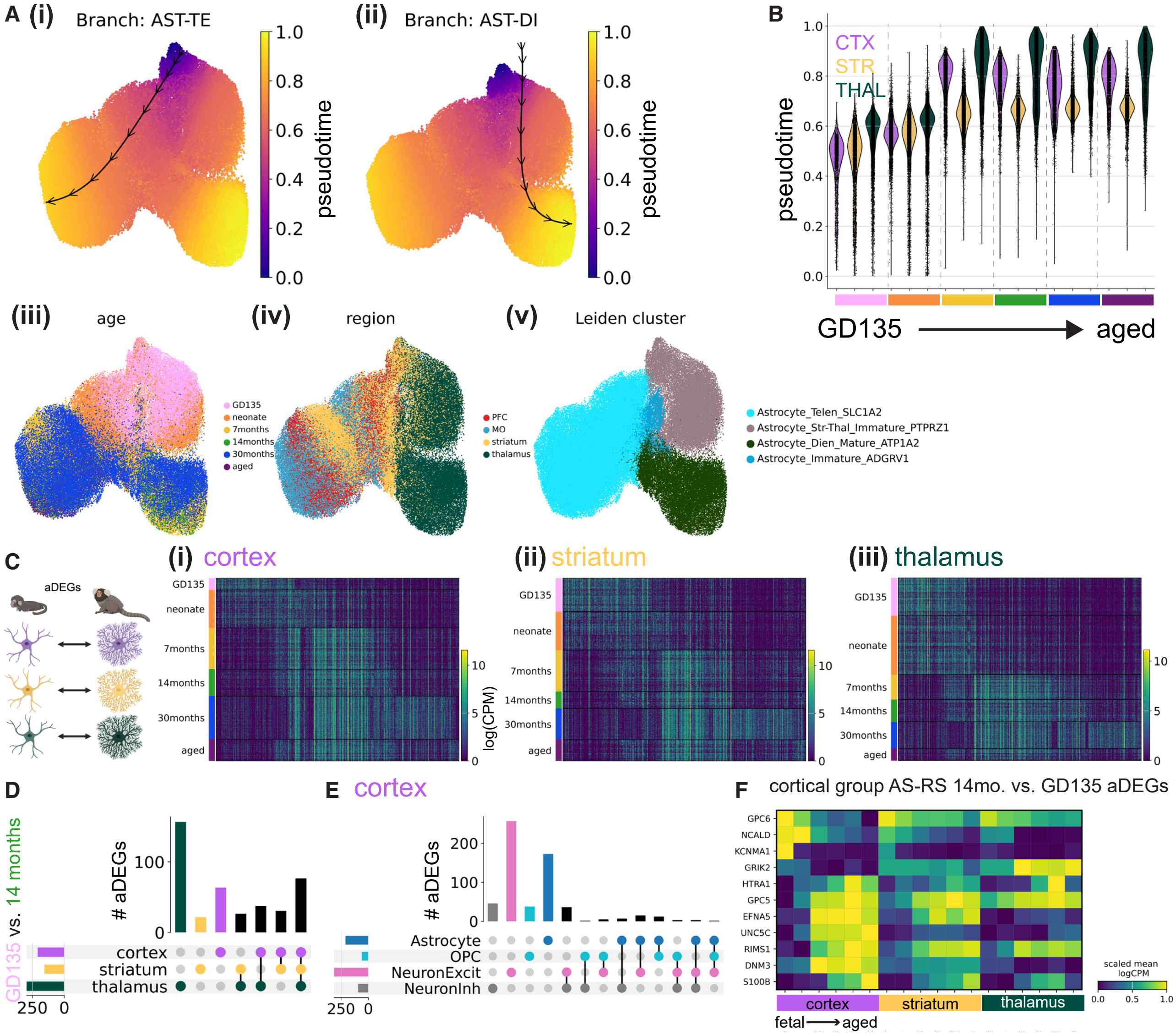


Figure 5. The postnatal developmental specification of marmoset astrocytes within and across brain regions

(A) Integrated UMAP embeddings of 103,009 marmoset astrocytes colored by (i and ii) Palantir-predicted pseudotime, (iii) developmental time point, (iv) brain region (assigned), and (v) Leiden cluster assignment. Trajectory path (black lines and arrows) are overlaid for (i) telencephalic astrocyte (AST-TE) and (ii) diencephalic astrocyte (AST-DI) branches.

(B) Violin plot (scanpy's default) showing the estimated distribution of pseudotime values for the astrocytes in (A) grouped by region within each developmental time point (color code as in Ai). Vertical dashed lines indicate separation between time points.

(C) Heatmaps (rows corresponding to nuclei and columns to gene) showing expression in logCPM of astrocyte age-differentially expressed genes (aDEGs) in astrocytes from (i) cortex, (ii) striatum, and (iii) thalamus, grouped by developmental time point, as indicated on the left of the heatmap. The strategy for calculating aDEGs is schematized on the left.

(D) UpSet plot showing the number of overlapping GD135 vs. 14-month astrocyte aDEGs between cortex, striatum, and thalamus. The colored dots below each vertical bar indicate which region(s) share that set of aDEGs, while the colored horizontal bars indicate the total number of cortex-thalamus aDEGs for each region. Overlap categories with 0 aDEGs are not shown.

(E) UpSet plot (as in D) showing the number of overlapping GD135 vs. 14-month cortical astrocyte aDEGs between OPCs, astrocytes, excitatory neurons, and inhibitory neurons.

(F) Matrix plot showing mean expression of selected cortex group astrocyte-specific, region-specific (AS-RS) aDEGs (rows) in marmoset astrocytes grouped by region and developmental time point (columns, blocked by region first and then by increasing age within each region block). Expression was standardized between 0 and 1 by subtracting the minimum and dividing by the maximum for each trait.

See also Figures S15 and S16.

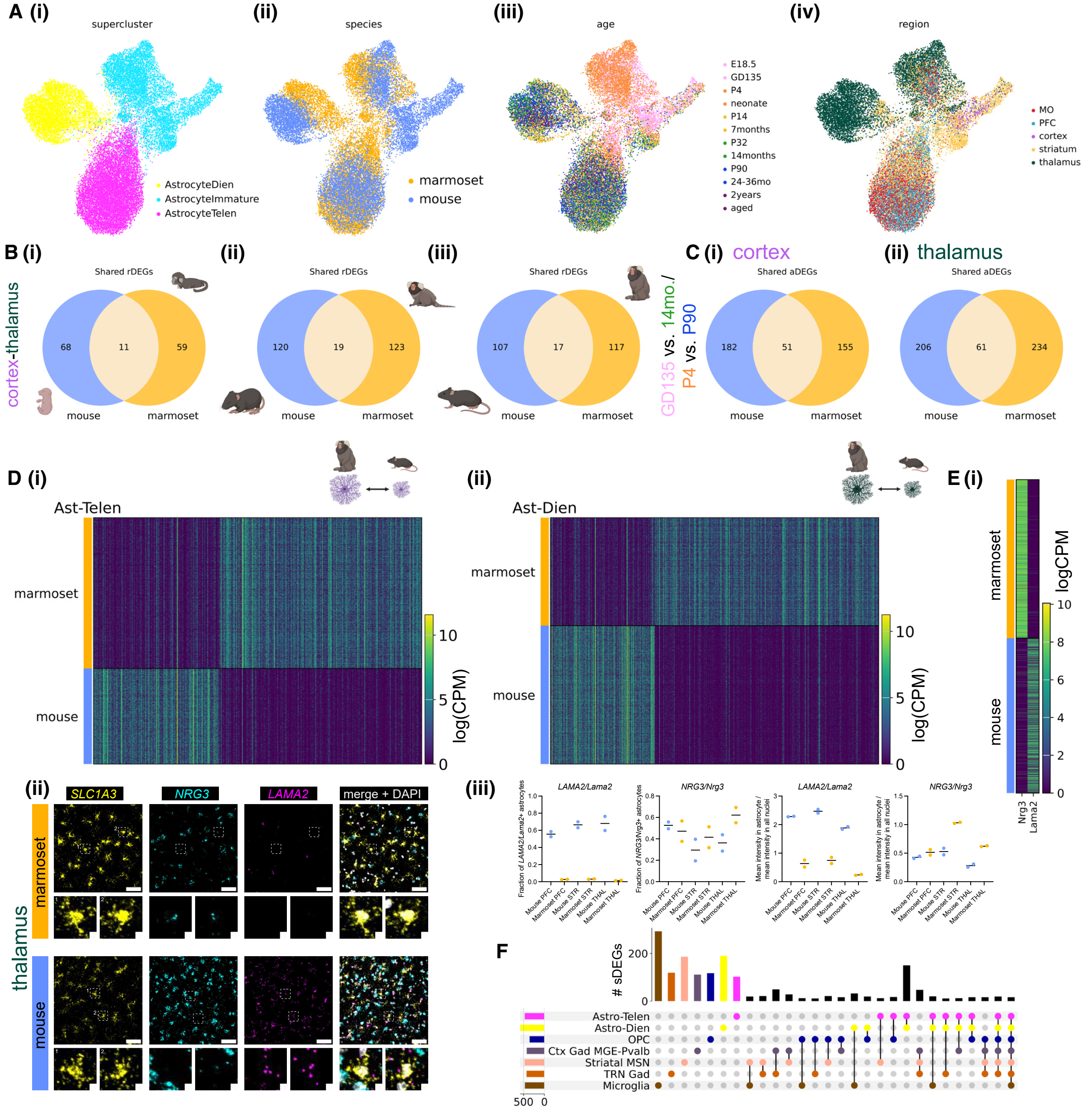


Figure 6. Conservation and divergence of the development of astrocyte heterogeneity in mouse and marmoset

(A) scANVI-integrated UMAP embeddings of marmoset and mouse astrocytes, colored by (i) supercluster, (ii) species, (iii) age, and (iv) region.

(B) Venn diagrams showing regional differentially expressed genes (rDEGs) between cortex and thalamus astrocytes shared across mouse and marmoset at (i) fetal, (ii) early adolescent, and (iii) young adult time points.

(C) Venn diagram showing age-differentially expressed genes (aDEGs) shared between mouse and marmoset astrocytes within the (i) cortex and (ii) thalamus.

(D) Heatmaps showing expression in logCPM of species-differentially expressed genes (sDEGs) between marmoset and mouse within (i) telencephalic astrocytes and (ii) diencephalic astrocytes.

(E) *In situ* validation of selected sDEGs in marmoset and mouse tissue with the RNAscope v2 assay. (i) Heatmap (rows are nuclei and columns are genes) of sDEG *NRG3/Nrg3* and *LAMA2/Lama2* expression in marmoset and mouse astrocytes. (ii) Top row: maximum intensity projections of cropped fields of view in the marmoset (top row) and mouse (bottom row) thalamus stained via RNAscope v2 FISH for astrocyte markers *SLC1A3/Slc1a3*, *NRG3/Nrg3*, and mouse *LAMA2/Lama2*. Scale bar, 50 μm. Bottom row: high-magnification images of the boxed astrocytes in the top row. Scale bar, 5 μm. (iii) CellProfiler quantification of sDEG abundance in both species from RNAscope v2 data (*n* = 2 donors per species, see STAR Methods). From left to right: fraction of *LAMA2/Lama2*-positive

(legend continued on next page)

Species-integrated astrocytes partitioned into three superclusters that segregated by developmental stage and by brain structure (diencephalon vs. telencephalon) (Figure 6A). This indicates that, at the level of broad cephalic domains, region patterning is conserved between the two species. Mature telencephalic astrocytes showed better species integration than diencephalic or immature astrocytes, and immature mouse astrocytes formed a distinct cluster (Figure 6A). This finding implies that, at birth, mouse astrocyte maturity lags behind that of marmoset, in line with our findings about oligodendrocyte maturation (Figures S13A, S13B, S13G, and S13H). Interestingly, marmoset astrocyte aDEGs had more discrete expression boundaries across time (Figure 5C), whereas mouse astrocyte aDEGs had more continuous temporal expression, especially in the striatum (Figure S16). Additionally, marmoset rDEGs shared across ages were largely divided into younger (GD135 and neonate) and older (7 months and older) groups (Figure 2C). In contrast, temporally overlapping mouse rDEGs were more evenly distributed across individual ages and smaller groups of ages (Figure 3C), suggesting that developmental changes occur more slowly over the sampled time points in mouse.

In mouse, but not marmoset, we observed immature astrocyte clusters composed of nuclei from all time points, suggesting continued generation of new astrocytes throughout the lifespan. These included the *Top2a+* immature astrocyte population seen in the neurogenic subventricular zone throughout the lifespan (Figures S17D and S17E), which forms part of the rostral migratory stream.[74] All marmoset astrocytes and cortical and thalamic mouse astrocytes exhibited separate embryonic and neonatal subclusters from early adolescent and older cells (Figures S12, S13A, and S13C). In contrast, in the mouse striatum, there were several more immature clusters (9 total), including some composed of astrocytes from mature time points (Figure S13B, “Str_Ast2,” “Str_Ast6,” and “Str_Ast12”). However, the relative immaturity of mouse astrocytes after adolescence is not driven solely by the persistence of *Top2a+* cells in the SVZ, as we found an immature population of *Top2a-* striatal astrocytes in mouse, but not marmoset, that may reside outside of the SVZ (Figure S13B, portions of “Str_Ast2” and “Str_Ast12”). We note that the presence of this immature cluster in the striatum does not suggest that all mouse astrocytes are less mature than their marmoset counterparts in adolescence and adulthood.

We next tested whether genes that best distinguished astrocytes from a given brain region in one species were more likely than chance to be rDEGs in the other species. Focusing on astrocyte cortex-thalamus rDEGs at each developmental time point, we found that the majority of rDEGs were not shared across species and that the proportion of shared rDEGs decreased only slightly from fetal to early adolescence time points, from ∼14%–16% to ∼13%–14% in both species (Figure 6B; full list of species-overlapping and species-unique cortex-thalamus astrocyte rDEGs at each developmental time point in Table S21). The proportion of overlapping rDEGs was not greater than chance (see STAR Methods, *p* value from a Fisher’s exact test > 0.05 at all developmental time points).

Similarly, the majority of astrocyte aDEGs in cortex and thalamus were not shared between species. 51 cortical aDEGs (22% of mouse cortical P4–P90 aDEGs and 25% of marmoset GD135–14-month aDEGs) and 61 thalamic aDEGs (23% of mouse thalamic P4–P90 aDEGs and 21% of marmoset thalamic GD135–14-month aDEGs) were shared between species (Figure 6C, *p* value from a Fisher’s exact test on the proportion overlapped = 0.025 for cortex and 0.454 for marmoset). Lists of species-shared and species-unique aDEGs by region are provided in Table S22.

Next, we directly tested for differential expression of 1:1 orthologs between species within shared superclusters. We calculated sDEGs based on 1:1 orthologs between species within each integrated supercluster using our metacell method (see STAR Methods). We found hundreds of sDEGs in both telencephalic (464 total) and diencephalic (579 total) astrocytes, whose expression could clearly distinguish between marmoset- and mouse-derived populations (Figure 6D). Astrocyte sDEGs encoded both cytosolic and membrane-bound proteins with varied cellular functions (Table S24). For example, telencephalic astrocyte sDEGs higher in marmoset included *NALCN*/*Nalcn*, a non-selected sodium leak channel; the RNA-binding protein *RBFOX2*/*Rbfox2*; *KCNT2*/*Kcnt2*, a sodium-activated potassium channel subunit; *FABP7*/*Fabp7*, a fatty-acid-binding protein with established roles in neurogenesis; and *DNM3*/*Dnm3*, a multi-domain GTPase involved in membrane remodeling. Even this short list of genes suggests that important cellular functions, such as ion buffering, RNA processing, fatty-acid binding, and membrane remodeling, may differ between astrocytes of different species. Furthermore, 63/286 telencephalic and 71/395 diencephalic sDEGs were SFARI[75] 3.0 autism spectrum disorder (ASD)-related genes (see STAR Methods). These overlaps are significantly higher than chance (assuming 20,000 protein-coding genes in the human genome, *p* values from Fisher’s exact tests $< 10^{-15}$ for both telencephalic and diencephalic astrocyte sDEGs). Complete lists of sDEGs for each supercluster analyzed, and GO annotations for cellular compartment, biological process, and molecular function, are provided in Table S23.

We validated the differential expression of 2 astrocyte sDEGs, *NRG3*/*Nrg3* (present in neurons in both species but higher in marmoset astrocytes) and *LAMA2*/*Lama2* (higher in mouse astrocytes), *in situ* using RNAscope (Figure 6E). 50% of the genes that distinguish diencephalic astrocytes between species were

astrocytes, fraction of *NRG3*/*Nrg3*-positive astrocytes, normalized mean intensity of *LAMA2*/*Lama2* (mean intensity in expanded astrocyte nuclei divided by mean intensity in all nuclei), and normalized mean intensity of *NRG3*/*Nrg3*. Data points are from individual biological replicates, with 2 slices averaged for mouse, and the horizontal black line denotes the median.

(F) UpSet plot showing shared sDEGs across superclusters. The colored dots below each vertical bar indicate which supercluster(s) share that set of sDEGs, whereas the colored horizontal bars indicate the total number of mouse-marmoset sDEGs for each supercluster. For simplicity, only supercluster combinations with 10 or more shared sDEGs are shown.

See also Figure S17.

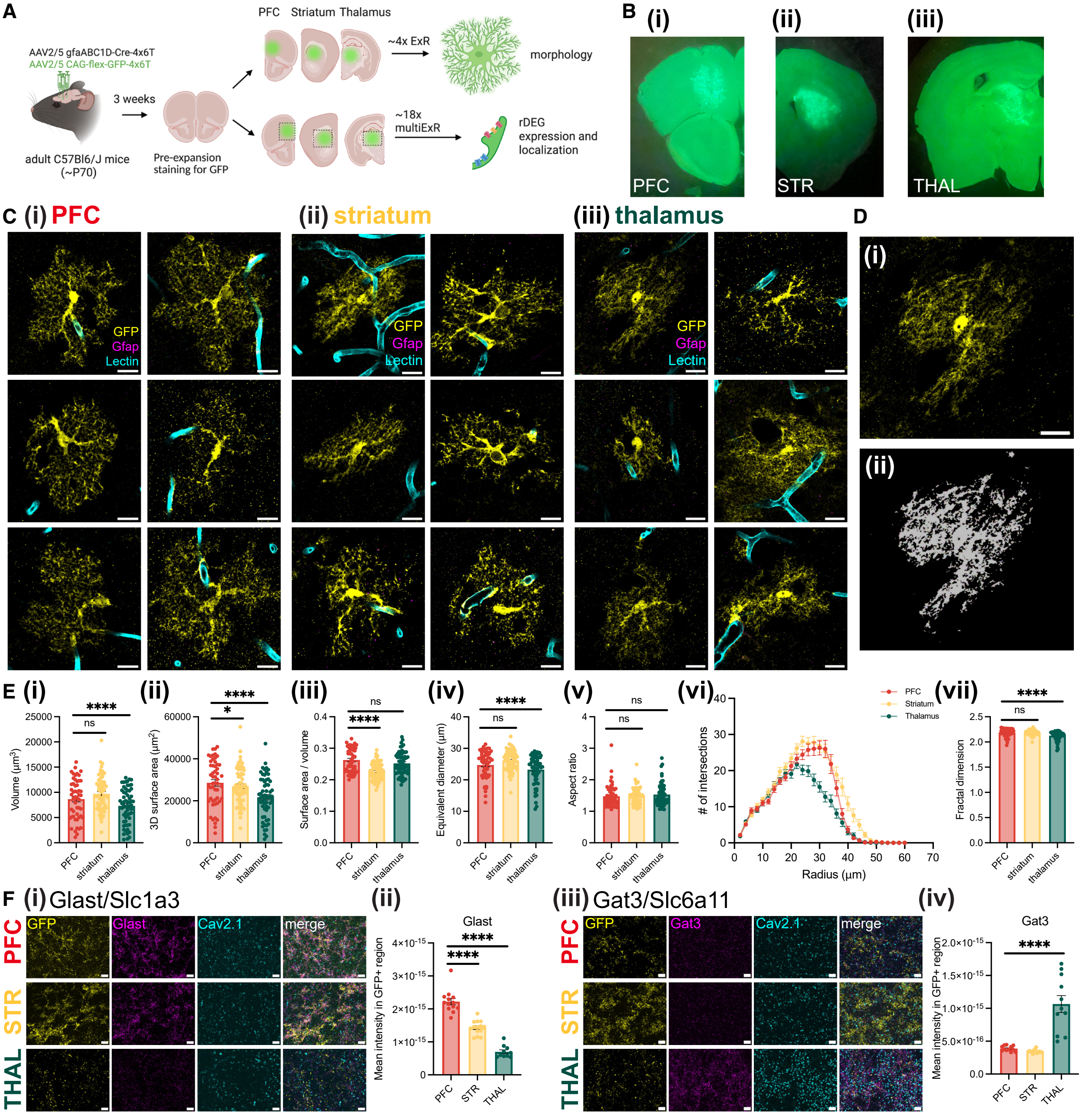


Figure 7. Expansion microscopy of virally labeled astrocytes in the mouse and marmoset brain

(A) Viral labeling approach for mouse astrocytes (see STAR Methods), created using BioRender.com.

(B) Brain slice hemispheres containing regions of interest in (i) PFC, (ii) striatum, and (iii) thalamus after pre-expansion staining for GFP, viewed from a dissecting microscope under blue light illumination.

(C) Single z-slices of background-subtracted images of ~3.5× expanded astrocytes in the (i) PFC, (ii) striatum, and (iii) thalamus, co-stained with GFP, GFAP, and the blood vessel marker Lectin. Shown are 6 examples of medium- to high-GFP-expressing astrocytes from 6 separate mice. Scale bar, 10 μm in biological units. Contrast was adjusted to 35% saturation (Fiji's[82,83] "auto") in most cases and was further increased for dimmer astrocytes to aid visibility; therefore, astrocytes are not equally contrast adjusted. See Videos S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, and S36.

(D) (i) Single z-slice of one of the thalamic astrocytes in (Ciii) and (ii) its corresponding segmentation. Scale bar, 10 μm in biological units.

(E) Bar plots showing quantified morphological properties for mouse astrocytes from PFC, striatum, and thalamus (*n* = 52–60 astrocytes from 3 female and 5 male mice for each region), with statistical significance determined using a linear mixed effects model with "animal" as the random effect group variable; see Table S25): (i) volume, (ii) surface area, (iii) surface area/volume, (iv) equivalent diameter, (v) aspect ratio, (vi) Sholl analysis (number of intersections with

(legend continued on next page)

shared with telencephalic astrocytes (Table S24). These telencephalic-diencephalic astrocyte-shared sDEGs made up a larger fraction (62%) of telencephalic astrocyte sDEGs. This suggests that evolution has acted on the astrocyte class as a whole, while also shaping divergent regional astrocyte programs between species. Additionally, as with astrocyte rDEGs and aDEGs, we found that most astrocyte sDEGs were not shared with other cell types (superclusters), including OPCs, cortical MGE-derived *PVALB+* interneurons, GABAergic TRN neurons, striatal MSNs, and microglia (Figure 6F). This result underscores that the evolutionary divergence of a cell type's transcriptome unfolds at different rates across cell types.[61,76,77] Taken together, these findings support both conservation and divergence of postnatal astrocyte regional specialization in mouse and marmoset.

Astrocytes have regionally divergent morphology and protein expression

Many of the genes we found to vary in astrocytes by region, age, and species implicate processes involved in morphological specification. Indeed, astrocyte morphology, which is highly ramified and complex, is essential for their specialized functions: endfeet contact blood vessels to help form the blood-brain barrier and shuttle water and nutrients, while terminal processes closely appose synapses to uptake ions and neurotransmitters[35] and regulate synapse development and function.[78] Because many of these morphological features exist at the sub-micron scale, conventional light microscopy is not sufficient to visualize the full morphological complexity of astrocytes.[36] We wondered whether nanoscale astrocyte morphology might also be regionally specialized between gray matter regions in cortex, striatum, and thalamus, as recently demonstrated for several CNS regions using diffraction-limited approaches.[25] Thus, we used ExR, a new variant of protein-decrowding expansion microscopy,[37] to visualize astrocyte morphology with enhanced resolution and to compare morphological properties across the PFC, striatum, and thalamus, as these regional populations represented the major molecular subpopulations in our snRNA-seq data (Figures 2 and 3).

We used a viral approach to label astrocytes for expansion (Figures 7A and 7B; see STAR Methods) and created 3D binary segmentations to quantitatively assess morphological differences across regions (Figure 7D; Videos S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, and S36). We calculated the volume, surface area, and equivalent diameter (measures of size); surface-area-to-volume ratio (S:V, a measure of shape, inversely proportional to size); aspect ratio (a measure of elongation); fractal dimension (FD)[79] (a measure of complexity and self-similarity); and branching complexity via Sholl analysis,[80] with most of these parameters having been used in previous studies to characterize astrocyte morphology.[36,81]

After correcting for multiple comparisons across the 6 univariate measures, we found significant differences in size (volume, surface area, and equivalent diameter), shape (surface-area-to-volume ratio), and morphological complexity (FD) between astrocytes from different brain regions, particularly between cortical/striatal and thalamic astrocytes (Figure 7E, $n = 52$–60 astrocytes from 3 female and 5 male mice for each region, with statistical significance determined using a linear mixed effects model with "animal" as the random effect group variable, see Table S25). Specifically, thalamic astrocytes were smaller and less complex compared with cortical and striatal astrocytes, while striatal astrocytes had smaller surface area and surface-area-to-volume ratios compared with cortical astrocytes. To exclude the possibility that proximity to under-digested blood vessels or incomplete capture in the axial dimension impacted these differences, we repeated this analysis on a subset of astrocytes meeting additional criteria and found largely similar results (Figure S18A; see STAR Methods). Similarly, Sholl analysis revealed fewer intersections at radii larger than ~25 μm for thalamic compared with cortical and striatal astrocytes (Figure 7Evi). Taken together, these results suggest that thalamic astrocytes are smaller and less morphologically complex than their cortical and striatal counterparts, in line with prior work using conventional microscopy.[25]

We next probed rDEG protein product expression level and localization in astrocytes at the nanoscale using ExR. We processed tissue from 2 of the virally labeled adult C57BL/6J mice used for morphology analysis and proceeded with staining for Glast (encoded by the telencephalic rDEG *Slc1a3*) and Gat3 (encoded by the thalamic rDEG *Slc6a11*), alongside GFP (astrocyte processes) and Cav2.1 (a presynaptic protein), in gels prepared from PFC, striatum, and thalamus. We found that the expression of Glast and Gat3 at the protein level in GFP-labeled astrocytes agreed with snRNA-seq predictions: Glast expression was highest in the cortex, lower in the striatum, and lowest in the thalamus, whereas Gat3 expression was higher in the thalamus than the cortex or striatum (Figures 7F and S18B). Both proteins localized to astrocyte processes near synapses. Taken together, these results support the differential expression of rDEG protein products between telencephalic and diencephalic astrocytes and reveal the localization of rDEG protein products on and near astrocyte processes, in close proximity to neuronal synapses.

DISCUSSION

Although astrocyte regional molecular heterogeneity has been evident for some time,[84,85] the source of this regional

---

concentric shells as a function of radius), and (vii) FD by the box counting method (see STAR Methods). ns, non significant, $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.

(F) Maximum intensity projection composites of cropped (in x and y and across 51 z-slices) regions of GFP-labeled astrocyte processes co-stained with either (i) Glast (telencephalic rDEG) or (iii) Gat3 (thalamic rDEG) and the synaptic protein Cav2.1, alongside GFP generated using the ExR protocol (~18× expansion factor). Shown are processes from astrocytes in the PFC, striatum, and thalamus. Scale bar, 0.5 μm. Contrast was manually adjusted by setting equal minimum and maximum intensity values for Cav2.1 and rDEG targets and using Fiji's "auto" for GFP. (ii and iv) Bar plots showing quantified mean intensity of either Glast or Gat3 within masked GFP+ regions (astrocytes) in the whole ExR image volume ($n = 10$–16 fields of view from 2 mice, with statistical significance determined using a linear mixed effect model as described in the STAR Methods, see Table S26). Asterisks and p as in (E). In (E) and (F), error bars indicate standard error of the mean.

See also Figure S18.

heterogeneity—in particular, the relative contributions of embryonic patterning versus response to environmental cues after birth—is not well understood.[86] To help bridge this knowledge gap, we generated a unified, multi-region, postnatal developmental snRNA-seq atlas of mouse and marmoset brain cells. Because our dataset contains all brain cell types, we anticipate that this atlas will be a valuable resource for the field. As such, we have made both raw and processed data publicly available on the Gene Expression Omnibus (GEO) and Neuroscience Multi-omic Data Archive (NeMO) or the Broad Single Cell Portal, respectively (see data and code availability). The latter is useful for exploring cell type clusters and querying the expression pattern of genes of interest across ages and regions, and it does not require coding expertise.

We found that astrocytes were regionally patterned before birth in both species, which was not unexpected, given the prevalence of homeobox patterning genes among astrocyte rDEGs,[8] evidence from a recent study showing regionally patterned glioblasts in the first-trimester human brain,[6] and older lineage tracing showing regional allocation of astrocytes based on the region of their originating radial glia.[27] Less predictably, we discovered dramatic changes in astrocyte regional identity between birth and early adolescence, in line with their maturation during this period. This period also coincides with peak synaptogenesis, pruning, and myelination,[87] consistent with the notion that astrocyte specialization depends on the activity of neighboring cells.[88]

The functions of embryonically patterned and postnatally acquired astrocyte rDEGs were varied, but they implicate known astrocyte processes, including supporting synaptic transmission, ion transport, neurotransmitter uptake, cell-cell adhesion, and morphological specification. The function of some rDEGs, including *SLC6A11* (GAT-3) and *SPARC*, has been studied in astrocytes and shown to be important in modulating the effects of brain injury[89] and controlling synaptogenesis,[52] respectively. We anticipate that future mechanistic studies of other astrocyte rDEGs will reveal yet more essential functions.

Despite the striking regional heterogeneity of astrocytes, many predicted neuron-astrocyte L-R pairs were shared across regions. Even those not shared across regions were functionally similar, suggesting that neurons and astrocytes have developed a common language of molecular communication across the forebrain. Indeed, some of our rDEGs were members of the same protein family or functional class, pointing to variations on a common theme of neuron-astrocyte crosstalk across brain regions. Many of the top predicted neuron-astrocyte L-R pairs, such as NRXNs/NLGNs, are more traditionally associated with neuron-neuron contact at the synapse.[90] However, adhesion molecules such as ephrins, NRXNs/NLGNs, and NrCAMs have been shown to play important roles in neuron-astrocyte communication.[48]

In both species, we found hundreds of aDEGs; many of these were astrocyte specific but region agnostic, some were astrocyte specific and region specific, and a small number were cell type agnostic but region specific. Our astrocyte-specific, region-agnostic aDEGs can be interpreted as a core forebrain astrocyte developmental program and were more likely to be shared across the species. For example, in marmoset, this included *NTRK2*, which encodes the brain-derived neurotrophic factor (BDNF) receptor TrkB, the short isoform of which has been shown to be essential for astrocyte morphogenesis.[91] Perhaps unsurprisingly, several of our region-specific aDEGs were also rDEGs and/or had a high degree of functional overlap with rDEGs. Although pseudotime approaches have limitations[92] and may not fully capture how maturational states differ across brain regions, they can provide information about the progression of change along a trajectory that is correlated with actual age. For this reason, we used pseudotime to compare relative maturation differences across regions. This analysis suggested that intrinsic maturation rates are relatively low drivers of regional differences in gene expression.

In both species, most astrocyte rDEGs, aDEGs, and sDEGs were not shared with OPCs or neurons, suggesting that astrocyte region and age specializations are unique rather than general to all radial-glia-derived cell types in the same region, developmental time point, or species. This suggests either that regional gene expression signatures change throughout neuro- and glio-genesis or that the downstream transcriptional effects of this early regional patterning depend on the daughter cell's fate. Evidence for both exists in the cortex.[93] Why neurons and astrocytes exhibit stark regional patterning in adulthood, albeit in different ways, while the oligodendrocyte lineage does not, is an outstanding question for future study.

The present study characterized two mammalian neuroscience model species, mouse and marmoset. Although mice and humans have a high degree of genetic conservation,[94] mice have certain limitations as a model for studying the human brain, including the lack of a well-developed PFC and complex social behaviors, as well as their poor visual acuity. In light of these limitations, non-human primates (NHPs), with whom we share much closer genetic ancestors, are considered as more translationally relevant models of brain function and dysfunction. The common marmoset has become an increasingly popular NHP model in neuroscience studies due to its faster generation time for genetic engineering, shorter lifespan than other larger primates for developmental and late-onset disease studies, and complex social behaviors.[95]

The astrocyte sDEGs that we found encode proteins involved in key cellular functions that may have undergone evolutionary selection, and a significant portion have been associated with ASD. Future studies exploring the function of these sDEGs within astrocytes may reveal how primate astrocytes have evolved to suit the unique anatomy and physiology of the primate brain. Our findings suggest that each species may have evolved by recruiting different sets of genes that facilitate postnatal regional specialization of astrocytes. We found that many cell types in the marmoset brain are transcriptionally more mature at time of birth than their mouse counterparts, in line with previously documented precocious development in early postnatal marmosets.[96] This species divergence in transcriptional maturity at time of birth suggests that researchers should use caution when comparing early postnatal time points between rodents and NHPs, especially in light of differences in developmental tempo between species.[97]

Our quantitative profiling of astrocyte morphology using expansion microscopy in the mouse brain shows that astrocyte size, shape, and complexity do vary across brain regions and are most distinct between thalamus and cortex. Prior studies have also found appreciable morphological differences in mouse astrocytes across brain regions.[19,25,98] Although other higher-resolution approaches such as electron microscopy may reveal additional differences in aspects of astrocyte morphology, we demonstrate that expansion microscopy, particularly ExR, offers an inexpensive and accessible alternative to other super-resolution approaches for characterizing astrocyte morphology with enhanced resolution. We anticipate that this approach could also be used to study morphological changes in astrocytes after manipulation and/or in disease contexts.

There are several notable limitations to the current study, only some of which we discuss here. The use of nuclei instead of whole cells prevents our detection of RNAs in the cytoplasm, including those locally translated in distant processes of both neurons[99] and astrocytes,[100] which are likely relevant in establishing cell type and state identity. However, others have found similar cell type discrimination capabilities for scRNA-seq and snRNA-seq in the mouse cortex, despite lower RNA content (20%–50% of total cellular mRNA) in single nuclei.[101] The second significant limitation is the relatively small sample size, especially for marmosets, due to practical limitations, including cost, which limits our ability to compare between sexes.

Third, we relied in part on pathway analysis to summarize patterns and deduce functional implications arising from sets of rDEGs. Our use of WebGestalt did not incorporate any fold-change or p value information for genes, treating each DEG equally regardless of its differential expression level, which may skew results. Furthermore, pathway analysis is only as accurate as the underlying annotations, which can be lacking for glial biology. Finally, many genes are involved in several pathways. For these reasons, we encourage interested readers to directly examine our DEG lists provided in Tables S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, and S32.

Fourth, there are significant challenges in integrating snRNA-seq data across species. Even before data analysis, read alignment will differ across species, varying with the quality and content of reference genome annotation. During data analysis, approaches requiring direct merging of the cell × gene count matrices result in the loss of biological information because only roughly 50%–60% of total genes detected in either species were mapped as one-to-one orthologs. For this reason, we also integrated the data across species with an orthogonal approach that does not rely on one-to-one ortholog mapping. Both approaches are further limited by the need for *a priori* cell type annotation, which may bias toward or against integration of shared and unshared superclusters, respectively. Therefore, we are most confident in sDEGs, which are all one-to-one orthologs and calculated within shared superclusters, as a measure of species divergence.

Fifth, our approach for labeling astrocytes for morphological analysis relies on viral infection and manual identification of astrocytes meeting a minimum brightness level for imaging and segmentation, which may be biased toward a certain astrocyte subtype. Finally, we relied on RNAscope HiPlex to assess rDEG mRNA levels *in situ*. Any multiplexing technique that involves repeated stripping and restaining suffers from some level of reduced fluorescence intensity in later rounds. Therefore, any researcher interested in following up on a gene or protein of interest should perform additional confirmatory studies with a single round of imaging.

Taken together, our data support a model of astrocyte regional specialization that includes both embryonic patterning and postnatal specialization in response to local environmental cues, including synapse formation and neuronal activity, as has been previously suggested.[25,85] However, transcriptomic data alone are insufficient to support a definitive role (or lack thereof) of the neuronal microenvironment on the development of astrocyte regional identity. To determine whether or not early transcriptional patterning is required for proper postnatal astrocyte specialization, a cross-regional astrocyte heterotopic transplant would be illuminating. That is, would a thalamic-born astrocyte be able to acquire the transcriptional and morphological profile of a cortical astrocyte if transplanted in early postnatal life? Evidence from such an experiment in septal astrocyte populations suggests the answer is yes.[31] Alternatively, but not mutually exclusively, early developmental regional patterning may "prime" astrocytes to receive and react appropriately to the signals they receive in their local niches later in development, as a recent study has shown in the context of GABA-induced morphogenesis.[102] We anticipate that the current study will be a useful starting point for hypotheses such as these.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to, and will be fulfilled by, the lead contact, Guoping Feng (fengg@mit.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw data (sequencing reads in fastq format) for all 10× Chromium snRNA-seq samples and CellBender-cleaned aligned counts matrices (in .h5 format) are publicly available for download on the Neuroscience Multi-omic Data Archive (NeMO) at: https://data.nemoarchive.org/biccn/grant/u01_feng/feng/transcriptome/sncell/10x_v3.1/ (NeMO: https://assets.nemoarchive.org/dat-4j0ndn0) and on the Gene Expression Omnibus (GEO) at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE309702 (GEO: GSE309702)
- Pre-processed, clustered, and annotated data (in .h5ad format) are publicly available for download, exploration, and gene search on the Broad Single Cell Portal at: https://singlecell.broadinstitute.org/single_cell/study/SCP2719/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-mouse-brain (mouse) and https://singlecell.broadinstitute.org/single_cell/study/SCP2706/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-marmoset-brain (marmoset).
- Registered RNAscope HiPlex/v2 FISH image stacks, raw and background-subtracted expansion microscopy image volumes, and binarized 3D tracings (both in .tif format) are publicly available for download on BossDB at https://bossdb.org/project/schroeder2025 (BossDB: https://doi.org/10.60533/boss-2024-nqrj).
- Custom scripts used to analyze data are publicly available at https://github.com/Feng-Lab-MIT/AstrocyteHeterogeneity (Zenodo: https://doi.org/10.5281/zenodo.16911362).

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

Pictograms were generated using BioRender.com. M.E.S. was supported by the MathWorks Science Fellowship at MIT, the Collamore-Rogers Fellowship at MIT, the NSF Graduate Research Fellowship Program #1745302, and NIH 1F31MH133329-01. G.F. was supported by the National Institute of Mental Health and BRAIN Initiative (U01MH114819), Hock E. Tan and K. Lisa Yang Center for Autism Research of the Yang Tan Collective at MIT, the Poitras Center for Psychiatric Disorders Research at MIT, and Stanley Center for Psychiatric Research at Broad Institute of MIT and Harvard. This work was supported by BRAIN Initiative grant UM1MH130981 to G.F. and F.M.K. E.S.B. was supported by HHMI, Lisa Yang, NIH R01AG087374, NIH 1R01EB024261, Good Ventures, Tom Stocky and Avni Shah, Kathleen Octavio, NIH 1R01AG070831, and European Research Council (ERC) SYNERGY grant no. 835102. We thank Eric Nyase, Andrew Harrahill, and In-Hye Kang for AAV packaging and preparation; Michael Debarbardine for generating mitochondrial gene annotations for mCalJa1.2.Pat.X; Yanay Rosen for generating the marmoset protein embedding for SATURN; Menglong Zeng for guidance, particularly on molecular cloning, and for sharing many reagents; Morgan Fleishman for animal colony maintenance and support; Victoria Beja-Glasser for guidance and the gift of several wild-type mice; Ryan Kast for guidance, particularly on RNA FISH, and for sharing many reagents; Samia Silva de Castro, Jitendra Sharma, and Yefei Chen for performing necropsies and harvesting marmoset tissue; Alyssa Lutservitz and Xian Adikonis for tips on the 10× snRNA-seq protocol; 10× Genomics Technical Support; and Stephen Yu for assistance with marmoset PCR-based sex genotyping. We thank all members of the Feng and Boyden labs for insight and suggestions throughout the project. We gratefully acknowledge the MIT BioMicroCenter, the Broad Institute Genomics Platform, and MIT Department of Comparative Medicine for their support and assistance of this work.

AUTHOR CONTRIBUTIONS

M.E.S., K.X.L., G.F., and F.M.K. conceptualized the project. M.E.S. harvested and microdissected mouse brain tissue and microdissected marmoset brain tissue with assistance from K.X.L. and D.M.M. Q.Z. oversaw marmoset breeding and donor selection and performed brain tissue harvesting. M.E.S., L.R.M., and K.X.L. conducted snRNA-seq experiments. M.E.S., D.M.M., and K.X.L. analyzed the snRNA-seq data, with guidance from F.M.K. M.E.S. and L.R.M. performed mouse colony management. L.M. performed stereotaxic surgery for AAV injection in adult mice and performed transcardial perfusions alongside D.M.M. E.Y. and J.K. optimized collagenase treatment for blood vessel preservation and made ExR gels. M.E.S. dissected, pre-stained, stained, and imaged low-expansion ExR gels and analyzed images. J.K. stained and imaged high-expansion ExR gels. H.Z. and K.M.L. imaged marmoset tissue and provided guidance with marmoset tissue sectioning and staining. F.M.K., G.F., and E.S.B. supervised the project. M.E.S., F.M.K., E.S.B., and G.F. acquired funding. M.E.S. wrote and revised the manuscript with D.M.M., K.M.L., F.M.K., G.F., and E.S.B.

DECLARATION OF INTERESTS

J.K. and E.S.B. are co-inventors on a patent application for ExR (US 2020/0271556 A1). M.E.S., J.K., and E.S.B. are co-inventors on a patent for a related technology, multiExR (WO2025/090986). E.S.B. co-founded a company to explore clinical applications of expansion microscopy technologies.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Marmosets
  - Mice
- METHOD DETAILS
  - Marmoset tissue harvest for snRNA-seq
  - Mouse tissue harvest for snRNA-seq
  - Nuclei isolation and single-nucleus RNA sequencing
  - Sex determination in mouse and marmoset fetal and neonate samples
  - RNA fluorescence *in situ* hybridization (FISH)
  - Mouse astrocyte labeling and ExR for mouse tissue
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Read alignment
  - Calculation of sequencing coverage statistics
  - snRNA-seq data preprocessing and quality control
  - Global (species-wide, cross-age, cross-region) integration and annotation
  - Integration quality analysis
  - Alignment to Allen Brain Cell Atlas (ABCA) with MapMyCells
  - Compositional analysis with scCODA
  - Region reassignment of cross-contaminant nuclei in fetal and neonate marmoset
  - Region reassignment of cross-contaminant nuclei in mouse
  - Calculation of rDEGs, aDEGs, and sDEGs using a pseudobulk method
  - Automated UniProt annotation and querying of SFARI Gene 3.0
  - Astrocyte subclustering within each brain region
  - Pathway analysis with WebGestalt
  - Cross-species integration with scANVI
  - Cross-species integration with SATURN
  - Cell-cell communication analysis
  - Pseudotime analysis and calculation of gene change scores
  - FISH quantification with CellProfiler
  - Quantification of mouse astrocyte rDEG expression in the ABCA Whole Mouse Brain atlas
  - Analysis of low-expansion factor ExR morphology images
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2025.09.011.

Received: June 19, 2025
Revised: September 9, 2025
Accepted: September 10, 2025
Published: November 20, 2025

REFERENCES

1. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature *624*, 317–332. https://doi.org/10.1038/s41586-023-06812-z.
2. Langlieb, J., Sachdev, N.S., Balderrama, K.S., Nadaf, N.M., Raj, M., Murray, E., Webber, J.T., Vanderburg, C., Gazestani, V., Tward, D., et al. (2023). The molecular cytoarchitecture of the adult mouse brain. Nature *624*, 333–342. https://doi.org/10.1038/s41586-023-06818-7.
3. Aldridge, S., and Teichmann, S.A. (2020). Single cell transcriptomics comes of age. Nat. Commun. *11*, 4307. https://doi.org/10.1038/s41467-020-18158-5.
4. Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K.W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. Science *382*, eadd7046. https://doi.org/10.1126/science.add7046.

5. La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth, C., Khven, I., Lederer, A.R., Dratva, L.M., et al. (2021). Molecular architecture of the developing mouse brain. Nature *596*, 92–96. https://doi.org/10.1038/s41586-021-03775-x.
6. Braun, E., Danan-Gotthold, M., Borm, L.E., Lee, K.W., Vinsland, E., Lönnerberg, P., Hu, L., Li, X., He, X., Andrusivová, Ž., et al. (2023). Comprehensive cell atlas of the first-trimester developing human brain. Science *382*, eadf1226. https://doi.org/10.1126/science.adf1226.
7. Krienen, F.M., Levandowski, K.M., Zaniewski, H., del Rosario, R.C.H., Schroeder, M.E., Goldman, M., Wienisch, M., Lutservitz, A., Beja-Glasser, V.F., Chen, C., et al. (2023). A marmoset brain cell census reveals regional specialization of cellular identities. Sci. Adv. *9*, eadk3986. https://doi.org/10.1126/sciadv.adk3986.
8. Lin, J.-P., Kelly, H.M., Song, Y., Kawaguchi, R., Geschwind, D.H., Jacobson, S., and Reich, D.S. (2022). Transcriptomic architecture of nuclei in the marmoset CNS. Nat. Commun. *13*, 5531. https://doi.org/10.1038/s41467-022-33140-z.
9. Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.
10. Ben Haim, L., and Rowitch, D.H. (2017). Functional diversity of astrocytes in neural circuit regulation. Nat. Rev. Neurosci. *18*, 31–41. https://doi.org/10.1038/nrn.2016.159.
11. Allen, N.J., and Barres, B.A. (2009). Neuroscience: Glia—more than just brain glue. Nature *457*, 675–677. https://doi.org/10.1038/457675a.
12. Oliveira, J.F., and Araque, A. (2022). Astrocyte regulation of neural circuit activity and network states. Glia *70*, 1455–1466. https://doi.org/10.1002/glia.24178.
13. Lee, H.G., Wheeler, M.A., and Quintana, F.J. (2022). Function and therapeutic value of astrocytes in neurological diseases. Nat. Rev. Drug Discov. *21*, 339–358. https://doi.org/10.1038/s41573-022-00390-x.
14. Verkhratsky, A., and Nedergaard, M. (2018). Physiology of Astroglia. Physiol. Rev. *98*, 239–389. https://doi.org/10.1152/physrev.00042.2016.
15. Ramón Y Cajal, S., Azoulay, L., swanson, N., and Swanson, W. (1995). Histology of the Nervous System: of Man and Vertebrates (Oxford University Press). https://doi.org/10.1093/oso/9780195074017.001.0001.
16. Oberheim, N.A., Goldman, S.A., and Nedergaard, M. (2012). Heterogeneity of astrocytic form and function. In Methods in Molecular Biology, *814* (Humana Press), pp. 23–45. https://doi.org/10.1007/978-1-61779-452-0_3.
17. Bachoo, R.M., Kim, R.S., Ligon, K.L., Maher, E.A., Brennan, C., Billings, N., Chan, S., Li, C., Rowitch, D.H., Wong, W.H., et al. (2004). Molecular diversity of astrocytes with implications for neurological disorders. Proc. Natl. Acad. Sci. USA *101*, 8384–8389. https://doi.org/10.1073/pnas.0402140101.
18. Morel, L., Chiang, M.S.R., Higashimori, H., Shoneye, T., Iyer, L.K., Yelick, J., Tai, A., and Yang, Y. (2017). Molecular and Functional Properties of Regional Astrocytes in the Adult Brain. J. Neurosci. *37*, 8706–8717. https://doi.org/10.1523/JNEUROSCI.3956-16.2017.
19. Chai, H., Diaz-Castro, B., Shigetomi, E., Monte, E., Octeau, J.C., Yu, X., Cohn, W., Rajendran, P.S., Vondriska, T.M., Whitelegge, J.P., et al. (2017). Neural Circuit-Specialized Astrocytes: Transcriptomic, Proteomic, Morphological, and Functional Evidence. Neuron *95*, 531–549.e9. https://doi.org/10.1016/j.neuron.2017.06.029.
20. Lin, C.C.J., Yu, K., Hatcher, A., Huang, T.W., Lee, H.K., Carlson, J., Weston, M.C., Chen, F., Zhang, Y., Zhu, W., et al. (2017). Identification of diverse astrocyte populations and their malignant analogs. Nat. Neurosci. *20*, 396–405. https://doi.org/10.1038/nn.4493.
21. Bayraktar, O.A., Bartels, T., Holmqvist, S., Kleshchevnikov, V., Martirosyan, A., Polioudakis, D., Ben Haim, L., Young, A.M.H., Batiuk, M.Y., Prakash, K., et al. (2020). Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. Nat. Neurosci. *23*, 500–509. https://doi.org/10.1038/s41593-020-0602-1.
22. Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015–1030.e16. https://doi.org/10.1016/j.cell.2018.07.028.
23. Batiuk, M.Y., Martirosyan, A., Wahis, J., de Vin, F., Marneffe, C., Kusserow, C., Koeppen, J., Viana, J.F., Oliveira, J.F., Voet, T., et al. (2020). Identification of region-specific astrocyte subtypes at single cell resolution. Nat. Commun. *11*, 1220. https://doi.org/10.1038/s41467-019-14198-8.
24. Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999–1014.e22. https://doi.org/10.1016/j.cell.2018.06.021.
25. Endo, F., Kasai, A., Soto, J.S., Yu, X., Qu, Z., Hashimoto, H., Gradinaru, V., Kawaguchi, R., and Khakh, B.S. (2022). Molecular basis of astrocyte diversity and morphology across the CNS in health and disease. Science *378*, eadc9020. https://doi.org/10.1126/science.adc9020.
26. Hochstim, C., Deneen, B., Lukaszewicz, A., Zhou, Q., and Anderson, D.J. (2008). Identification of Positionally Distinct Astrocyte Subtypes whose Identities Are Specified by a Homeodomain Code. Cell *133*, 510–522. https://doi.org/10.1016/J.CELL.2008.02.046.
27. Tsai, H.H., Li, H., Fuentealba, L.C., Molofsky, A.V., Taveira-Marques, R., Zhuang, H., Tenney, A., Murnen, A.T., Fancy, S.P.J., Merkle, F., et al. (2012). Regional astrocyte allocation regulates CNS synaptogenesis and repair. Science *337*, 358–362. https://doi.org/10.1126/science.1222381.
28. Welle, A., Kasakow, C.V., Jungmann, A.M., Gobbo, D., Stopper, L., Nordström, K., Salhab, A., Gasparoni, G., Scheller, A., Kirchhoff, F., et al. (2021). Epigenetic control of region-specific transcriptional programs in mouse cerebellar and cortical astrocytes. Glia *69*, 2160–2177. https://doi.org/10.1002/GLIA.24016.
29. Clavreul, S., Abdeladim, L., Hernández-Garzón, E., Niculescu, D., Durand, J., Ieng, S.H., Barry, R., Bonvento, G., Beaurepaire, E., Livet, J., and Loulier, K. (2019). Cortical astrocytes develop in a plastic manner at both clonal and cellular levels. Nat. Commun. *10*, 4884. https://doi.org/10.1038/s41467-019-12791-5.
30. Farmer, W.T., Abrahamsson, T., Chierzi, S., Lui, C., Zaelzer, C., Jones, E.V., Bally, B.P., Chen, G.G., Théroux, J.-F., Peng, J., et al. (2016). Neurons diversify astrocytes in the adult brain through sonic hedgehog signaling. Science *351*, 849–854. https://doi.org/10.1126/science.aab3103.
31. Xie, Y., Reid, C.M., Turrero García, M., Dale-Huang, F., Granados, A.A., Lu, Y., Li, J., Hanson, S.M., Mancia Leon, W.R., Liu, J., et al. (2025). Astrocyte specification in the mouse septum is shaped by both developmental origin and local signals. Nat. Neurosci. *28*, 1676–1687. https://doi.org/10.1038/s41593-025-02007-z.
32. Molofsky, A.V., and Deneen, B. (2015). Astrocyte development: A Guide for the Perplexed. Glia *63*, 1320–1329. https://doi.org/10.1002/glia.22836.
33. Tung, P.Y., Blischak, J.D., Hsiao, C.J., Knowles, D.A., Burnett, J.E., Pritchard, J.K., and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. Sci. Rep. *7*, 39921. https://doi.org/10.1038/srep39921.
34. Tran, H.T.N., Ang, K.S., Chevrier, M., Zhang, X., Lee, N.Y.S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. *21*, 12. https://doi.org/10.1186/s13059-019-1850-9.
35. Zhou, B., Zuo, Y.-X., and Jiang, R.-T. (2019). Astrocyte morphology: Diversity, plasticity, and role in neurological diseases. CNS Neurosci. Ther. *25*, 665–673. https://doi.org/10.1111/cns.13123.

36. Baldwin, K.T., Murai, K.K., and Khakh, B.S. (2024). Astrocyte morphology. Trends Cell Biol. *34*, 547–565. https://doi.org/10.1016/j.tcb.2023.09.006.

37. Sarkar, D., Kang, J., Wassie, A.T., Schroeder, M.E., Peng, Z., Tarr, T.B., Tang, A.-H., Niederst, E.D., Young, J.Z., Su, H., et al. (2022). Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy. Nat. Biomed. Eng. *6*, 1057–1073. https://doi.org/10.1038/s41551-022-00912-3.

38. Cao, Z.-J., and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat. Biotechnol. *40*, 1458–1466. https://doi.org/10.1038/s41587-022-01284-4.

39. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods *19*, 41–50. https://doi.org/10.1038/s41592-021-01336-8.

40. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

41. Büttner, M., Ostner, J., Müller, C.L., Theis, F.J., and Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. Nat. Commun. *12*, 6876. https://doi.org/10.1038/s41467-021-27150-6.

42. Rubenstein, J.L.R., Shimamura, K., Martinez, S., and Puelles, L. (1998). REGIONALIZATION OF THE PROSENCEPHALIC NEURAL PLATE. Annu. Rev. Neurosci. *21*, 445–477. https://doi.org/10.1146/annurev.neuro.21.1.445.

43. Sasai, Y., and De Robertis, E.M. (1997). Ectodermal Patterning in Vertebrate Embryos. Dev. Biol. *182*, 5–20. https://doi.org/10.1006/dbio.1996.8445.

44. Rowitch, D.H., and Kriegstein, A.R. (2010). Developmental genetics of vertebrate glial–cell specification. Nature *468*, 214–222. https://doi.org/10.1038/nature09611.

45. Di Bella, D.J., Habibi, E., Stickels, R.R., Scalia, G., Brown, J., Yadollahpour, P., Yang, S.M., Abbate, C., Biancalani, T., Macosko, E.Z., et al. (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. Nature *595*, 554–559. https://doi.org/10.1038/s41586-021-03670-5.

46. Anthony, T.E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial Glia Serve as Neuronal Progenitors in All Regions of the Central Nervous System. Neuron *41*, 881–890. https://doi.org/10.1016/S0896-6273(04)00140-0.

47. Frassoni, C., Amadeo, A., Ortino, B., Jaranowska, A., and Spreafico, R. (2000). Organization of radial and non-radial glia in the developing rat thalamus. J. Comp. Neurol. *428*, 527–542. https://doi.org/10.1002/1096-9861(20001218)428:3<527::AID-CNE9>3.0.CO;2-X.

48. Tan, C.X., and Eroglu, C. (2021). Cell adhesion molecules regulating astrocyte–neuron interactions. Curr. Opin. Neurobiol. *69*, 170–177. https://doi.org/10.1016/j.conb.2021.03.015.

49. Sobolczyk, M., and Boczek, T. (2022). Astrocytic Calcium and cAMP in Neurodegenerative Diseases. Front. Cell. Neurosci. *16*, 889939. https://doi.org/10.3389/fncel.2022.889939.

50. Zhou, Z., Okamoto, K., Onodera, J., Hiragi, T., Andoh, M., Ikawa, M., Tanaka, K.F., Ikegaya, Y., and Koyama, R. (2021). Astrocytic cAMP modulates memory via synaptic plasticity. Proc. Natl. Acad. Sci. USA *118*, e2016584118. https://doi.org/10.1073/pnas.2016584118.

51. Bazargani, N., and Attwell, D. (2016). Astrocyte calcium signaling: the third wave. Nat. Neurosci. *19*, 182–189. https://doi.org/10.1038/nn.4201.

52. Kucukdereli, H., Allen, N.J., Lee, A.T., Feng, A., Ozlu, M.I., Conatser, L.M., Chakraborty, C., Workman, G., Weaver, M., Sage, E.H., et al. (2011). Control of excitatory CNS synaptogenesis by astrocyte-secreted proteins hevin and SPARC. Proc. Natl. Acad. Sci. USA *108*, E440–E449. https://doi.org/10.1073/pnas.1104977108.

53. Elizarraras, J.M., Liao, Y., Shi, Z., Zhu, Q., Pico, A.R., and Zhang, B. (2024). WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. Nucleic Acids Res. *52*, W415–W421. https://doi.org/10.1093/nar/gkae456.

54. Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. *33*, W741–W748. https://doi.org/10.1093/nar/gki475.

55. UniProt Consortium (2025). UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Res. *53*, D609–D617. https://doi.org/10.1093/nar/gkae1010.

56. Liu, Y.F., Sowell, S.M., Luo, Y., Chaubey, A., Cameron, R.S., Kim, H.-G., and Srivastava, A.K. (2015). Autism and Intellectual Disability-Associated KIRREL3 Interacts with Neuronal Proteins MAP1B and MYO16 with Potential Roles in Neurodevelopment. PLoS One *10*, e0123106. https://doi.org/10.1371/journal.pone.0123106.

57. Kim, D., and Ackerman, S.L. (2011). The UNC5C Netrin Receptor Regulates Dorsal Guidance of Mouse Hindbrain Axons. J. Neurosci. *31*, 2167–2179. https://doi.org/10.1523/JNEUROSCI.5254-10.2011.

58. Wetzel-Smith, M.K., Hunkapiller, J., Bhangale, T.R., Srinivasan, K., Maloney, J.A., Atwal, J.K., Sa, S.M., Yaylaoglu, M.B., Foreman, O., Ortmann, W., et al. (2014). A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat. Med. *20*, 1452–1457. https://doi.org/10.1038/nm.3736.

59. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100. https://doi.org/10.1186/gb-2006-7-10-r100.

60. Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A., and Goodman, A. (2021). CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics *22*, 433. https://doi.org/10.1186/s12859-021-04344-9.

61. Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature *598*, 111–119. https://doi.org/10.1038/s41586-021-03465-8.

62. Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68. https://doi.org/10.1038/s41586-019-1506-7.

63. Fang, R., Xia, C., Close, J.L., Zhang, M., He, J., Huang, Z., Halpern, A.R., Long, B., Miller, J.A., Lein, E.S., et al. (2022). Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. Science *377*, 56–62. https://doi.org/10.1126/SCIENCE.ABM1741.

64. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. *15*, 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

65. Delaunay, D., Heydon, K., Cumano, A., Schwab, M.H., Thomas, J.-L., Suter, U., Nave, K.-A., Zalc, B., and Spassky, N. (2008). Early Neuronal and Glial Fate Restriction of Embryonic Neural Stem Cells. J. Neurosci. *28*, 2551–2562. https://doi.org/10.1523/JNEUROSCI.5497-07.2008.

66. Hong, W., Gong, P., Pan, X., Ren, Z., Liu, Y., Qi, G., Li, J.-L., Sun, W., Ge, W.-P., Zhang, C.-L., et al. (2024). Temporal-spatial Generation of Astrocytes in the Developing Diencephalon. Neurosci. Bull. *40*, 1–16. https://doi.org/10.1007/s12264-023-01131-9.

67. Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol. *37*, 451–460. https://doi.org/10.1038/s41587-019-0068-4.

68. Marques, S., Zeisel, A., Codeluppi, S., Van Bruggen, D., Mendanha Falcão, A.M., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov,

R.A., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science *352*, 1326–1329. https://doi.org/10.1126/science.aaf6463.

69. Otto, D.J., Jordan, C., Dury, B., Dien, C., and Setty, M. (2024). Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. Nat. Methods *21*, 1185–1195. https://doi.org/10.1038/s41592-024-02302-w.
70. Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron *89*, 37–53. https://doi.org/10.1016/j.neuron.2015.11.013.
71. Han, X., Chen, M., Wang, F., Windrem, M., Wang, S., Shanz, S., Xu, Q., Oberheim, N.A., Bekar, L., Betstadt, S., et al. (2013). Forebrain Engraftment by Human Glial Progenitor Cells Enhances Synaptic Plasticity and Learning in Adult Mice. Cell Stem Cell *12*, 342–353. https://doi.org/10.1016/j.stem.2012.12.015.
72. Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M.I., and Yosef, N. (2021). Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol. Syst. Biol. *17*, e9620. https://doi.org/10.15252/msb.20209620.
73. Rosen, Y., Brbić, M., Roohani, Y., Swanson, K., Li, Z., and Leskovec, J. (2024). Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. Nat. Methods *21*, 1492–1500. https://doi.org/10.1038/s41592-024-02191-z.
74. Sohn, J., Orosco, L., Guo, F., Chung, S.-H., Bannerman, P., Mills Ko, E.M., Zarbalis, K., Deng, W., and Pleasure, D. (2015). The Subventricular Zone Continues to Generate Corpus Callosum and Rostral Migratory Stream Astroglia in Normal Adult Mice. J. Neurosci. *35*, 3756–3763. https://doi.org/10.1523/JNEUROSCI.3454-14.2015.
75. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36. https://doi.org/10.1186/2040-2392-4-36.
76. Jorstad, N.L., Song, J.H.T., Exposito-Alonso, D., Suresh, H., Castro-Pacheco, N., Krienen, F.M., Yanny, A.M., Close, J., Gelfand, E., Long, B., et al. (2023). Comparative transcriptomics reveals human-specific cortical features. Science *382*, eade9516. https://doi.org/10.1126/science.ade9516.
77. Krienen, F.M., Goldman, M., Zhang, Q., C H Del Rosario, R., Florio, M., Machold, R., Saunders, A., Levandowski, K., Zaniewski, H., Schuman, B., et al. (2020). Innovations present in the primate interneuron repertoire. Nature *586*, 262–269. https://doi.org/10.1038/s41586-020-2781-z.
78. Chung, W.-S., Baldwin, K.T., and Allen, N.J. (2024). Astrocyte Regulation of Synapse Formation, Maturation, and Elimination. Cold Spring Harb. Perspect. Biol. *16*, a041352. https://doi.org/10.1101/cshperspect.a041352.
79. Liebovitch, L.S., and Toth, T. (1989). A fast algorithm to determine fractal dimensions by box counting. Phys. Lett. A *141*, 386–390. https://doi.org/10.1016/0375-9601(89)90854-2.
80. Ferreira, T.A., Blackman, A.V., Oyrer, J., Jayabal, S., Chung, A.J., Watt, A.J., Sjöström, P.J., and van Meyel, D.J. (2014). Neuronal morphometry directly from bitmap images. Nat. Methods *11*, 982–984. https://doi.org/10.1038/nmeth.3125.
81. Soto, J.S., Neupane, C., Kaur, M., Pandey, V., Wohlschlegel, J.A., and Khakh, B.S. (2024). Astrocyte Gi-GPCR signaling corrects compulsive-like grooming and anxiety-related behaviors in Sapap3 knockout mice. Neuron *112*, 3412–3423.e6. https://doi.org/10.1016/j.neuron.2024.07.019.
82. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.
83. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
84. Zhang, Y., and Barres, B.A. (2010). Astrocyte heterogeneity: an underappreciated topic in neurobiology. Curr. Opin. Neurobiol. *20*, 588–594. https://doi.org/10.1016/j.conb.2010.06.005.
85. Bayraktar, O.A., Fuentealba, L.C., Alvarez-Buylla, A., and Rowitch, D.H. (2014). Astrocyte Development and Heterogeneity. Cold Spring Harb. Perspect. Biol. *7*, a020362. https://doi.org/10.1101/cshperspect.a020362.
86. Schober, A.L., Wicki-Stordeur, L.E., Murai, K.K., and Swayne, L.A. (2022). Foundations and implications of astrocyte heterogeneity during brain development and disease. Trends Neurosci. *45*, 692–703. https://doi.org/10.1016/J.TINS.2022.06.009.
87. Semple, B.D., Blomgren, K., Gimlin, K., Ferriero, D.M., and Noble-Haeusslein, L.J. (2013). Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. Prog. Neurobiol. *106–107*, 1–16. https://doi.org/10.1016/j.pneurobio.2013.04.001.
88. Freeman, M.R. (2010). Specification and Morphogenesis of Astrocytes. Science *330*, 774–778. https://doi.org/10.1126/science.1190928.
89. Cho, F.S., Vainchtein, I.D., Voskobiynyk, Y., Morningstar, A.R., Aparicio, F., Higashikubo, B., Ciesielska, A., Broekaart, D.W.M., Anink, J.J., van Vliet, E.A., et al. (2022). Enhancing GAT-3 in thalamic astrocytes promotes resilience to brain injury in rodents. Sci. Transl. Med. *14*, 4310. https://doi.org/10.1126/scitranslmed.abj4310.
90. Craig, A.M., and Kang, Y. (2007). Neurexin–neuroligin signaling in synapse development. Curr. Opin. Neurobiol. *17*, 43–52. https://doi.org/10.1016/j.conb.2007.01.011.
91. Holt, L.M., Hernandez, R.D., Pacheco, N.L., Torres Ceja, B., Hossain, M., and Olsen, M.L. (2019). Astrocyte morphogenesis is dependent on BDNF signaling via astrocytic TrkB. eLife *8*, e44667. https://doi.org/10.7554/eLife.44667.
92. Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B., and Saeys, Y. (2021). Recent advances in trajectory inference from single-cell omics data. Curr. Opin. Syst. Biol. *27*, 100344. https://doi.org/10.1016/j.coisb.2021.05.005.
93. Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323. https://doi.org/10.1126/science.aap8809.
94. Breschi, A., Gingeras, T.R., and Guigó, R. (2017). Comparative transcriptomics in human and mouse. Nat. Rev. Genet. *18*, 425–440. https://doi.org/10.1038/nrg.2017.19.
95. Okano, H., Hikishima, K., Iriki, A., and Sasaki, E. (2012). The common marmoset as a novel animal model system for biomedical and neuroscience research applications. Semin. Fetal Neonatal Med. *17*, 336–340. https://doi.org/10.1016/j.siny.2012.07.002.
96. Abbott, D.H., Barnett, D.K., Colman, R.J., Yamamoto, M.E., and Schultz-Darken, N.J. (2003). Aspects of Common Marmoset Basic Biology and Life History Important for Biomedical Research. Comp. Med. *53*, 339–350.
97. Lindhout, F.W., Krienen, F.M., Pollard, K.S., and Lancaster, M.A. (2024). A molecular and cellular perspective on human brain evolution and tempo. Nature *630*, 596–608. https://doi.org/10.1038/s41586-024-07521-x.
98. Khakh, B.S., and Deneen, B. (2019). The Emerging Nature of Astrocyte Diversity. Annu. Rev. Neurosci. *42*, 187–207. https://doi.org/10.1146/annurev-neuro-070918-050443.

99. Holt, C.E., and Schuman, E.M. (2013). The Central Dogma Decentralized: New Perspectives on RNA Function and Local Translation in Neurons. Neuron *80*, 648–657. https://doi.org/10.1016/j.neuron.2013.10.036.

100. Sakers, K., Lake, A.M., Khazanchi, R., Ouwenga, R., Vasek, M.J., Dani, A., and Dougherty, J.D. (2017). Astrocytes locally translate transcripts in their peripheral processes. Proc. Natl. Acad. Sci. USA *114*, E3830–E3838. https://doi.org/10.1073/pnas.1617782114.

101. Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One *13*, e0209648. https://doi.org/10.1371/journal.pone.0209648.

102. Cheng, Y.-T., Luna-Figueroa, E., Woo, J., Chen, H.-C., Lee, Z.-F., Harmanci, A.S., and Deneen, B. (2023). Inhibitory input directs astrocyte morphogenesis through glial GABABR. Nature *617*, 369–376. https://doi.org/10.1038/s41586-023-06010-x.

103. Pool, A.-H., Poldsam, H., Chen, S., Thomson, M., and Oka, Y. (2023). Recovery of missing single-cell RNA-sequencing data with optimized transcriptomic references. Nat. Methods *20*, 1506–1515. https://doi.org/10.1038/s41592-023-02003-w.

104. Takabayashi, S., and Katoh, H. (2011). Sex Identification Using the ZFX and ZFY Genes in Common Marmosets (Callithrix jacchus). Exp. Anim. *60*, 417–420. https://doi.org/10.1538/expanim.60.417.

105. Zargari, M., Sadeghi, M.R., Shahhosseiny, M.H., Kamali, K., Saliminejad, K., Esmaeilzadeh, A., and Khorshid, H.R.K. (2011). Fetal Sex Determination using Non-Invasive Method of Cell-free Fetal DNA in Maternal Plasma of Pregnant Women During 6th–10th Weeks of Gestation. Avicenna J. Med. Biotechnol. *3*, 201–206.

106. Fleming, S.J., Chaffin, M.D., Arduini, A., Akkad, A.-D., Banks, E., Marioni, J.C., Philippakis, A.A., Ellinor, P.T., and Babadi, M. (2023). Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat. Methods *20*, 1323–1335. https://doi.org/10.1038/s41592-023-01943-7.

107. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/S13059-017-1382-0/FIGURES/1.

108. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058. https://doi.org/10.1038/s41592-018-0229-2.

109. Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat. Commun. *13*, 3224. https://doi.org/10.1038/s41467-022-30755-0.

110. Bernstein, N.J., Fong, N.L., Lam, I., Roy, M.A., Hendrickson, D.G., and Kelley, D.R. (2020). Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. Cell Syst. *11*, 95–101.e5. https://doi.org/10.1016/j.cels.2020.05.010.

111. Paxinos, G., Watson, C., Petrides, M., Rosa, M., and Tokuno, H. (2012). The Marmoset Brain in Stereotaxic Coordinates (Elsevier).

112. Gleichman, A.J., Kawaguchi, R., Sofroniew, M.V., and Carmichael, S.T. (2023). A toolbox of astrocyte-specific, serotype-independent adeno-associated viral vectors using microRNA targeting sequences. Nat. Commun. *14*, 7426. https://doi.org/10.1038/s41467-023-42746-w.

113. Chen, F., Tillberg, P.W., and Boyden, E.S. (2015). Optical imaging. Expansion microscopy. Science *347*, 543–548. https://doi.org/10.1126/science.1260088.

114. Challis, R.C., Ravindra Kumar, S., Chan, K.Y., Challis, C., Beadle, K., Jang, M.J., Kim, H.M., Rajendran, P.S., Tompkins, J.D., Shivkumar, K., et al. (2019). Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. *14*, 379–414. https://doi.org/10.1038/s41596-018-0097-3.

115. Lock, M., Alvira, M.R., Chen, S.-J., and Wilson, J.M. (2014). Absolute Determination of Single-Stranded and Self-Complementary Adeno-Associated Viral Vector Genome Titers by Droplet Digital PCR. Hum. Gene Ther. Methods *25*, 115–125. https://doi.org/10.1089/hgtb.2013.131.

116. Valdes, P.A., Yu, C.J., Aronson, J., Ghosh, D., Zhao, Y., An, B., Bernstock, J.D., Bhere, D., Felicella, M.M., Viapiano, M.S., et al. (2024). Improved immunostaining of nanostructures and cells in human brain specimens through expansion-mediated protein decrowding. Sci. Transl. Med. *16*, eabo0049. https://doi.org/10.1126/scitranslmed.abo0049.

117. Kang, J., Schroeder, M.E., Lee, Y., Kapoor, C., Yu, E., Tarr, T.B., Titterton, K., Zeng, M., Park, D., Niederst, E., et al. (2024). Multiplexed expansion revealing for imaging multiprotein nanostructures in healthy and diseased brain. Nat. Commun. *15*, 9722. https://doi.org/10.1038/s41467-024-53729-w.

118. Donath, A., Jühling, F., Al-Arab, M., Bernhart, S.H., Reinhardt, F., Stadler, P.F., Middendorf, M., and Bernt, M. (2019). Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. *47*, 10543–10552. https://doi.org/10.1093/nar/gkz833.

119. Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., Koutrouli, M.; Scverse Community, and Berger, B., et al. (2023). The scverse project provides a computational ecosystem for single-cell omics data analysis. Nat. Biotechnol. *41*, 604–606. https://doi.org/10.1038/s41587-023-01733-8.

120. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44. https://doi.org/10.1038/nbt.4314.

121. Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M.D., Strobl, D.C., Henao, J., Curion, F., et al. (2023). Best practices for single-cell analysis across modalities. Nat. Rev. Genet. *24*, 550–572. https://doi.org/10.1038/s41576-023-00586-w.

122. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. *51*, W207–W212. https://doi.org/10.1093/nar/gkad347.

123. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., et al. (2019). Ensembl 2019. Nucleic Acids Res. *47*, D745–D751. https://doi.org/10.1093/nar/gky1113.

124. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27. https://doi.org/10.1016/j.cell.2018.05.061.

125. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE Trans. Vis. Comput. Graphics *20*, 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248.

126. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

127. Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.

128. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.

129. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of

atomic-level protein structure with a language model. Science *379*, 1123–1130. https://doi.org/10.1126/science.ade2574.

130. Geirsdottir, L., David, E., Keren-Shaul, H., Weiner, A., Bohlen, S.C., Neuber, J., Balic, A., Giladi, A., Sheban, F., Dutertre, C.-A., et al. (2019). Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. Cell *179*, 1609–1622.e16. https://doi.org/10.1016/j.cell.2019.11.010.

131. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

132. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

133. Wang, H., Xu, L., Lai, C., Hou, K., Chen, J., Guo, Y., Sambangi, A., Swaminathan, S., Xie, C., Wu, Z., et al. (2021). Region-specific distribution of Olig2-expressing astrocytes in adult mouse brain and spinal cord. Mol. Brain *14*, 36. https://doi.org/10.1186/s13041-021-00747-0.

134. Martynoga, B., Morrison, H., Price, D.J., and Mason, J.O. (2005). Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. Dev. Biol. *283*, 113–127. https://doi.org/10.1016/j.ydbio.2005.04.005.

135. Kita, Y., Nishibe, H., Wang, Y., Hashikawa, T., Kikuchi, S.S., U, M., Yoshida, A.C., Yoshida, C., Kawase, T., Ishii, S., et al. (2021). Cellular-resolution gene expression profiling in the neonatal marmoset brain reveals dynamic species- and region-specific differences. Proc. Natl. Acad. Sci. USA *118*, e2020125118. https://doi.org/10.1073/pnas.2020125118.

136. Shimogori, T., Abe, A., Go, Y., Hashikawa, T., Kishi, N., Kikuchi, S.S., Kita, Y., Niimi, K., Nishibe, H., Okuno, M., et al. (2018). Digital gene atlas of neonate common marmoset brain. Neurosci. Res. *128*, 1–13. https://doi.org/10.1016/j.neures.2017.10.009.

137. Arshadi, C., Günther, U., Eddison, M., Harrington, K.I.S., and Ferreira, T.A. (2021). SNT: a unifying toolbox for quantification of neuronal anatomy. Nat. Methods *18*, 374–377. https://doi.org/10.1038/s41592-021-01105-7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Chicken-anti-GFP	Abcam	Cat #: ab13790; RRID:AB_2936447
Mouse-anti-GFAP	Millipore	Cat #: MAB360; RRID:AB_11212597
Guinea pig-anti-Cav2.1	Synaptic Systems	Cat #: 152 205; RRID:AB_2619842
Rabbit-anti-EEAT1 (Glast)	Abcam	Cat #: ab181036; RRID:AB_2885103
Rabbit-anti-Gat3	Alomone	Cat #: AGT-003; RRID:AB_2340977
Lycopersicon Esculentum (Tomato) Lectin (LEL, TL), DyLight 649	ThermoFisher	Cat #: L32472
Goat-anti-chicken AF488	ThermoFisher	Cat #: A-11039; RRID:AB_142924
Donkey-anti-goat AF488	Jackson ImmunoResearch	Cat #: 705-545-147; RRID:AB_2336933
Donkey-anti-chicken AF488	ThermoFisher	Cat #: A-78948; RRID:AB_2921070
Donkey-anti-mouse AF555	ThermoFisher	Cat #: A-31570; RRID:AB_2536180
Donkey-anti-goat AF488	ThermoFisher	Cat #: A-32814; RRID:AB_2762838
Donkey-anti-guinea pig CF633	Biotium	Cat #: 20168; RRID:AB_10853143
Donkey-anti-rabbit AF555	ThermoFisher	Cat #: A-31572; RRID:AB_162543
Chemicals, peptides, and recombinant proteins		
Sodium acrylate	Santa Cruz	Cat #: CAS7446-81-3
Acrylamide	Sigma	Cat #: A9099
N,N′-Methylenebisacrylamide (BIS)	Sigma	Cat #: M7279
Ammonium persulfate (APS)	Sigma	Cat #: A3678
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	Sigma	Cat #: T7024
4-Hydroxy-TEMPO (HT)	Sigma	Cat #: 176141
6-((acryloyl)amino)hexanoic Acid, Succinimidyl Ester (AcX)	Thermo Fisher	Cat #: A20770
Paraformaldehyde 32% solution, EM grade	Electron Microscopy Sciences	Cat #: 15714-S
Collagenase from Clostridium histolyticum, Type VII	Millipore Sigma	Cat #: C0773
Acrylamide	Sigma	Cat #: A9099
Protector RNase Inhibitor	Millipore Sigma	Cat #: 3335402001
SPRIselect	Beckman Coulter	Cat #: B23318
ProLong™ Diamond Antifade Mountant	ThermoFisher	Cat #: P36970
TrueBlack Plus	biotium	Cat #: 23014
Critical commercial assays		
Chromium Next GEM Chip G Single Cell Kit, 16 rxns	10× Genomics	Cat #: 1000127
Chromium Next GEM Chip G Single Cell Kit, 48 rxns	10× Genomics	Cat #: 1000120
Chromium Next GEM Single Cell 3’ Kit v3.1, 16 rxns	10× Genomics	Cat #: 1000268
Chromium Next GEM Single Cell 3’ Kit v3.1, 4 rxns	10× Genomics	Cat #: 1000269
Chromium Nuclei Isolation with RNase Inhibitor Kit, 16rxns	10× Genomics	Cat #: 1000494
Dual Index Kit TT Set A 96 rxns	10× Genomics	Cat #: 1000215
RNAscope™ HiPlex12 Reagents Kit (488, 550, 650) v2	Advanced Cell Diagnostics	Cat #: 324419
RNAscope® Protease Plus	Advanced Cell Diagnostics	Cat #: 322331
RNAscope® HiPlex Probes, various	Advanced Cell Diagnostics	Details in Table S30 of this study
RNAscope® Probes, various	Advanced Cell Diagnostics	Details in Table S30 of this study
RNAscope™ Multiplex Fluorescent Detection Kit v2	Advanced Cell Diagnostics	Cat #: 323110

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Fluorescein TSA Fluorescence System Kit	APExBIO	Cat #: K1050
Cy3 TSA Fluorescence System Kit	APExBIO	Cat #: K1051
Cy5 TSA Fluorescence System Kit	APExBIO	Cat #: K1052
Deposited data		
Adult marmoset snRNA-seq data, raw and counts matrices, stored on NeMO portal	Krienen et al.[7]	https://assets.nemoarchive.org/dat-1je0mn3
Adult marmoset snRNA-seq data, analyzed and searchable, stored on CZI	Krienen et al.[7]	RRID:SCR_021059; https://cellxgene.cziscience.com/collections/0fd39ad7-5d2d-41c2-bda0-c55bde614bdb
Marmoset and mouse snRNA-seq data, raw and counts matrices, stored on NeMO portal and the Gene Expression Omnibus (GEO)	This study	https://data.nemoarchive.org/biccn/grant/u01_feng/feng/transcriptome/sncell/10x_v3.1/; NeMO: https://assets.nemoarchive.org/dat-4j0ndn0 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE309702; GEO: GSE309702
Marmoset snRNA-seq data, analyzed and searchable, stored Broad Institute Single Cell Portal	This study	Broad Institute Single Cell Portal: SCP2706; https://singlecell.broadinstitute.org/single_cell/study/SCP2706/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-marmoset-brain
Mouse snRNA-seq data, analyzed and searchable, stored on Broad Institute Single Cell Portal	This study	Broad Institute Single Cell Portal: SCP2719; https://singlecell.broadinstitute.org/single_cell/study/SCP2719/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-mouse-brain
Allen Whole Mouse Brain Transcriptomic Cell Type Atlas	Yao et al.[1]	https://knowledge.brain-map.org/abcatlas
Mouse reference genome, mm10 optimized v2	Pool et al.[103]	https://utsw.app.box.com/s/5tlmnw18tlimb9buc8iol9h9klrnn8wd
Marmoset reference genome, mCalja1.2.pat.X with mitochondrial genes from CM021961.1 annotated using MITOS2	Michael DeBerardine, Princeton Neuroscience Institute	Zenodo: https://doi.org/10.5281/zenodo.16915694
FISH and ExR images, hosted on BossDB	This study	https://bossdb.org/project/schroeder2025; BossDB: https://doi.org/10.60533/boss-2024-nqrj
Experimental models: Organisms/strains		
C57 Bl/6J mice	The Jackson Laboratory	Strain #000664; RRID:IMSR_JAX_000664
Wild-type common marmosets	Feng Lab and New England Primate Resource Center	N/A
Oligonucleotides		
ZfX/Y Forward primer: 5’-CTGTGCA TAACTTTGTTCCTG-3’	Takabayashi et al.[104] modified in this study	N/A
ZfX/Y Reverse primer: 5’-CAGTTGC CTTTGTCATCATC-3’	Takabayashi et al.[104] modified in this study	N/A
SRY Forward primer: 5’-TACAGGC CATGCACAGAGAG-3’	Zargari et al.[105]	N/A
SRY Reverse primer:5’-CTAGCGG GTGTTCCATTGTT-3’	Zargari et al.[105]	N/A
Recombinant DNA		
AAV-CAG-flex-GFP-4x6T	Stanley Thomas Carmichael via Addgene	Addgene plasmid # 196418; RRID:Addgene_196418
AAV-GfaABC1D-Cre-4x6T	Stanley Thomas Carmichael via Addgene	Addgene plasmid # 196410; RRID:Addgene_196410

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRG-pAAV-HELPER-Kan	Charles River Laboratory	RG-pAAV-HELPER-Kan
pAAV2/5	This study	N/A; Capsid gene identical to Addgene #104964; RRID:Addgene_104964
Software and algorithms		
Custom Python, MATLAB, and Fiji scripts; Versions of required packages below are listed in associated environment.yaml files. The version of record for the paper is release v0.1.	This study	https://github.com/Feng-Lab-MIT/AstrocyteHeterogeneity; Zenodo: https://doi.org/10.5281/zenodo.16911362.
MATLAB, version R2024a or later	MathWorks	https://www.mathworks.com/help/install/ug/install-products-with-internet-connection.html
Python, v3.9 or later via pip or conda	PyPI; Anaconda	https://pypi.org/project/pip/; https://www.anaconda.com/docs/getting-started/miniconda/main
Fiji and ImageJ	Schneider et al.[82]; Schindelin et al.[83]	https://imagej.net/software/fiji/downloads
Cellbender v 0.2.0	Fleming et al.[106]	https://github.com/broadinstitute/CellBender
MapMyCells	Allen Institute for Brain Science	https://portal.brain-map.org/atlases-and-data/bkp/mapmycells; RRID:SCR_024672
Scanpy	Wolf et al.[107]	https://scanpy.readthedocs.io/en/stable/
scVI	Lopez et al.[108]	https://scvi-tools.org/
Palantir	Setty et al.[67]	https://github.com/dpeerlab/Palantir
scCODA	Büttner et al.[41]	https://github.com/theislab/scCODA
Mellon	Otto et al.[69]	https://github.com/settylab/Mellon
SATURN	Rosen et al.[73]	https://github.com/snap-stanford/SATURN
scANVI	Xu et al.[72]	https://github.com/scverse/scvi-tools
CellPhoneDB via Liana	Efremova et al.[64]; Dimitrov et al.[109]	https://liana-py.readthedocs.io/en/latest/
Solo	Bernstein et al.[110]	https://docs.scvi-tools.org/en/stable/user_guide/models/solo.html
Other		
Chromium Controller	10× Genomics	Cat #: 1000202 (now obsolete)

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Marmosets

Common marmosets were housed in AAALAC-accredited facilities at MIT, in spacious holding rooms with a 12 h light/dark cycle, temperature 74.0 ± 2.0°F (23.3 ± 1.1°C), relative humidity of 50 ± 20%, and unrestricted access to food and water. Cages contained a variety of perches and enrichment devices. Procedures were conducted with prior approval by the MIT Committee for Animal Care (CAC) and following veterinary guidelines. A list of marmosets used in this study and their ages is provided in Table S1. Marmosets (GD135 - 13+ years old, 10 individuals for snRNA-seq and 4 individuals for FISH), were euthanized and brains harvested as previously described.[7] Marmosets were generated from a total of 14 breeding pairs.

Mice

Animal work was performed in accordance with protocols approved by MIT's Committee on Animal Care and NIH guidelines. All postnatal mice were wild-type C57BL/6J originally obtained from Jackson Laboratories and bred in-house. Timed pregnant C57BL/6J females were either obtained from Jackson Laboratories to arrive between gestation day 11 and 15 or were impregnated in house by setting up overnight mating pairs with females in proestrus or estrus phase. Embryos were harvested at E18.5 (18 days after the plug date). Mice were housed in a facility with a light cycle running from 07:00 to 19:00, temperature 20-22.2°C, humidity 30-70%, and food and water available *ad libitum*. Postnatal mice were not derived from timed pregnant females. Instead, age was determined during regular pup checks by experienced researchers based on the Jax Mice Pup Appearance

Chart(https://oacu.oir.nih.gov/system/files/media/file/2021-02/jaxpupsposter.pdf). Thus, ages are approximate within ±0.5 days for P4 neonates, within ±1 day for P14 early adolescents, within ±3 days for P32 juvenile mice and P90 young adult mice, and within ±1 week for aged mice (90 weeks). Except for the P32 and aged time points, mice were obtained from different litters, and minimal replicate effects were observed in the snRNA-seq data, suggesting adequate matching of developmental time points across replicates. A list of mice used in this study is provided in Table S1.

METHOD DETAILS

Marmoset tissue harvest for snRNA-seq

Briefly, animals were deeply sedated by intramuscular injection of ketamine (20–40 mg/kg) or alfaxalone (5–10 mg/kg), followed by intravenous injection of sodium pentobarbital (10–30 mg/kg). When the pedal withdrawal reflex was eliminated and/or the respiratory rate was diminished, animals were transcardially perfused with ice-cold sterile PBS. Whole brains were rapidly extracted into fresh PBS on ice.

After transporting the brain to the lab on wet ice, the brain was sectioned into coronal blocking cuts (slabs, 2-8 mm in thickness) using a chilled custom-designed marmoset brain matrix.[7] Surgical tools were autoclaved and allowed to cool before use. All tools, the matrix, and the dissecting block were cleaned with RNase Zap™ wipes (ThermoFisher) prior to each dissection. Slabs were transferred to a pre-chilled dissecting block and regions were dissected using a marmoset atlas as reference[111] (Table S27). The areas targeted for prefrontal cortex includes areas 8, 9, 10, 11, 47L, 14R, 46, 47, 13, 32, and 45; areas 6M, 6DC, 4c and 4ab for motor cortex; caudate and putamen for striatum, and all thalamic nuclei except posterior regions of the pulvinar and lateral geniculate nucleus for thalamus. Fetal and neonate brains were not dissected using the brain matrix due to their small size. Instead, the brains were hemisectioned, placed on a cooled dissecting block, and the prefrontal cortex and motor cortex (only for neonate, not dissected at GD135) were scooped from the surface of either hemisphere using anatomical landmarks. Two large (several mm) coronal slabs approximately spanning from the anterior beginning of the temporal lobe to its posterior end were cut using a razor blade, and the striatum and thalamus were dissected from the anterior and posterior slabs respectively (Figures S5A and S5B). For one neonate replicate (21-197), the brain was frozen and stored at −80°C for several months, placed at -20°C overnight prior to the day of dissection, and thawed on ice prior to dissection. Dissected tissue was transferred to chilled 1.5 mL microcentrifuge tubes, snap-frozen in liquid nitrogen, and stored at −80°C until nuclei isolation. Dissections began within 90 min of euthanasia and were performed in a median time of ∼40 min (range 30-80 min).

Mouse tissue harvest for snRNA-seq

Non-neonate animals were acclimated to the lab space for at least 30 min prior to beginning euthanasia. Euthanasia took place between 9am-12pm to minimize circadian rhythm effects, with a maximum of four animals processed per batch. Non-neonate animals were deeply anesthetized with isoflurane and decapitated. Heads were briefly submerged in liquid nitrogen for 3 s. Neonates were anesthetized via hypothermia and decapitated. Surgical tools were autoclaved and allowed to cool before use. All tools, brain mold, and the dissecting block were cleaned with RNase Zap™ wipes prior to each dissection. Brains were harvested and sagittally sectioned at 1 mm thickness for a total of 2 mm from the midline for either hemisphere (total of four ∼1 mm slices), on a brain mold using chilled razor blades. For P4 animals, two ∼2 mm sections from the midline were used. Tissue was dissected from slices on a chilled dissecting block exposed to room air, using a dissecting microscope at 1.6× magnification. Regions of interest were identified using the Allen Institute reference brain atlas at the appropriate time point. Dissected tissue was placed in cooled 1.5 mL microcentrifuge tubes and spun down in a tabletop mini centrifuge prior to snap freezing in liquid nitrogen before storage at -80°C until nuclei isolation. 2-3 neonates were pooled in each tube. Time from decapitation to snap freezing ranged from 7-13 min per animal. Samples from at least 3 mice (at least 1 female) are represented at each developmental time point (except 90 weeks, which is missing a male donor for PFC, see Table S1) and for each brain region. However, due to failures during microdissection, nuclei isolation, and 10× Genomics chip running, not all biological replicates are balanced across brain regions (e.g., some replicates have only one or two brain regions present).

For embryonic brain microdissection, the pregnant dam was deeply anesthetized with an overdose of isoflurane, decapitated, and placed on a cooled dissecting block. The abdomen was opened and placentas were removed from the abdominal cavity. Embryos were harvested from the placenta and rapidly decapitated one-by-one. Heads were frozen on metal disks over dry ice for 5-10 min until frozen solid, stored at -20°C for 1.5 h prior to microdissection. Heads were cut approximately in half using a small mouse brain mold and placed on a dry-ice cooled metal platform. Regions of interest were dissected using a tissue punch (1.27 mm Ted Pella MilTex Biopsy Punch with Plunger, 15110-10) to extract tissue from most medial surface on either hemisphere. 2-3 embryos were pooled in each tube. The other dissection procedures were the same as described above. Dissections were performed in less than 20 min per set of tubes from decapitation to snap freezing.

Nuclei isolation and single-nucleus RNA sequencing

Nuclei were extracted from frozen tissue using the 10× Genomics Chromium Nuclei Isolation Kit (Protocol CG000505, Rev A). Manufacturer instructions were followed with the following notable exceptions: (1) Most samples were dissected and frozen directly in Sample Dissociation Tubes (omitting step e), (2) Total lysis time was decreased to 10-14 total min of incubation in the lysis buffer (longer for marmoset and larger tissue chunks) from when lysis buffer was first added to the first sample (effectively shortening protocol step h) before proceeding to step i; (3) Tissue mass was larger than the 45 mg upper limit recommendation for some marmoset samples; (4) Lysis buffer was supplemented with Roche Protector RNase inhibitor at 0.2 U/μL, and (5) if no pellet was visible following any centrifugation steps, ~200 μL or less of supernatant was retained for samples with a visible pellet or debris. For the final resuspension step (step s), if no pellet was visible, less than ~40-100 μL of supernatant was retained at the bottom of the tube and no additional volume was added prior to nuclei counting. To avoid large clogs, large chunks of marmoset tissue were split in half and processed in parallel for nuclei isolation, and re-pooled prior to 10× Chromium chip loading. For nuclei isolation, a maximum of 4 samples were processed in series by a single researcher in a given preparation (usually 8 samples total, with 2 researchers in parallel). For tissue dissociation (step f), samples with a small amount of tissue (~20 mg or less) were processed (transferred to wet ice, coated with 200 μL lysis buffer, and dissociated with pestle) one at a time. For dissociating larger amounts of tissue (larger than ~20 mg) that required some thawing before pestle dissociation, samples were transferred to wet ice and coated with 200-300 μL of lysis buffer in parallel, and then homogenized with the pestle one at a time.

Nuclei concentration was quantified by staining suspensions with DAPI, loading on a C-Chip hemocytometer, imaging on a fluorescence microscope, and using the Fiji 3D Object Counter (https://imagej.net/plugins/3d-objects-counter) plugin to automatically quantify the number of nuclei within 4 large grid squares (0.8 μL of volume). Debris was assessed by comparing the signal in bright field to the signal in DAPI and nuclei quality was assessed by examining DAPI-stained nuclei at 40-60× magnification prior to starting the 10× Genomics Chromium snRNA-seq protocol. Nuclei suspensions with an unacceptable amount of debris (i.e., large clumps in bright field that were not DAPI+) and/or blebbing (i.e., with most nuclear membranes appearing substantially disrupted) were discarded. Because marmoset tissue is precious, a higher level of debris and/or blebbing was tolerated for marmoset nuclei suspensions. Nuclei suspensions were diluted to a target concentration of 1,000 nuclei/μL for 10× Genomics Chromium chip loading.

snRNA-seq libraries were prepared using 10× Genomics Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Protocol CG000315 Rev C or Rev D) following manufacturer instructions. Time from tissue lysis to 10× Chromium chip loading averaged ~2 h. Whenever possible, channels were loaded with enough nuclei suspension to recover a target of 10,000 nuclei. Initial cDNA amplification was performed using 13 PCR cycles and Sample Index PCR was performed using 11-12 PCR cycles. Amplified cDNA (product of protocol step 2) and libraries (product of protocol step 3) were quantified and quality-checked using both Qubit (HS dsDNA Assay) and a Fragment Analyzer. Libraries were pooled and sequenced on an Illumina NovaSeq at the Broad Genomics Platform. Data are available for download at: https://data.nemoarchive.org/biccn/grant/u01_feng/feng/transcriptome/sncell/10x_v3.1/.

Sex determination in mouse and marmoset fetal and neonate samples

Because 2-3 mouse E18.5 brain regions were pooled into a single tube for generating the nuclei suspension without sex determination, we do not have metadata about sex for these samples. Instead, we performed sex assignment on a per-nucleus basis after snRNA-seq based on the expression of Y-chromosome genes. Specifically, if a nucleus had *Zfy1*, *Zfy2*, *Usp9y*, *Uty*, *Eif2s3y*, *Kdm5d*, or *Ddx3y* expression above 2 log counts per million, it was assigned male sex. This resulted in 41.37% male nuclei for the E18.5 samples, likely an underestimate due to dropout.

Marmoset GD135 donors and one neonate donor (21-197) did not have their sex determined anatomically prior to euthanasia. To determine their sex, we performed PCR-based sex genotyping on either skin or brain tissue. Briefly, DNA was extracted from tissue using the NucleoSpin Tissue kit (Macherey-Nagel) and eluted in nuclease-free water. We genotyped for *ZfX/Y*[104] and *SRY*[105] using the following PCR primers (from 5' to 3'):

- ZfX/Y Forward (modified from the original publication): CTGTGCATAACTTTGTTCCTG
- ZfX/Y Reverse (modified from the original publication): CAGTTGCCTTTGTCATCATC
- SRY Forward: TACAGGCCATGCACAGAGAG
- SRY Reverse: CTAGCGGGTGTTCCATTGTT

And ran the following protocol on a thermocycler:

1. 94°C - 2 min
2. 98°C - 10 s
3. 58°C - 30 s
4. 68°C - 40 s
5. 35 or 40 rxns (34× or 39×)
6. 68°C - 3 min
7. 4°C - Infinite

We then digested 10uL of the ZFX/Y product using the DdeI and MseI enzymes in separate reactions (New England Biolabs). We were able to see clear separation of bands on a 2% agarose gel run at 135V for 40-45 min. Digesting the ZfX/Y PCR product with DdeI will yield double bands if the animal is XX and triple bands if the animal is XY. Digesting the ZfX/Y PCR product with MseI, which can only cut the ZfY band, will show smaller bands below the ~500 bp PCR product for XY animals and a single band for XX animals. An *SRY* band of around 218 bp indicates male sex. Our predicted sexes also matched *SRY* gene expression after snRNA-seq.

RNA fluorescence *in situ* hybridization (FISH)

Neonate (P4) or adult mice were deeply anesthetized via hypothermia (neonates) or isoflurane overdose (adults) and rapidly decapitated. Brains were extracted and frozen in OCT compound (Tissue-Tek) over dry ice and stored at -70°C. Briefly, neonate and adult marmosets were deeply sedated by intramuscular injection of ketamine (20–40 mg/kg) or alfaxalone (5–10 mg/kg), followed by intravenous injection of sodium pentobarbital (10–30 mg/kg). When the pedal with-drawal reflex was eliminated and/or the respiratory rate was diminished, animals were trans-cardially perfused with ice-cold sterile PBS. Whole brains were rapidly extracted into fresh PBS on ice for transfer to the lab (<30 min), then frozen in OCT over dry ice and stored at -70°C. Brains were sectioned sagittally on a cryostat (Leica) at -16 μm with a cutting temperature between -15 and -17°C and mounted on SuperFrost Plus (VWR/EMS) slides and stored at -70°C.

Multiplexed smFISH was performed using the RNAscope™ HiPlex kit and protocol (Advanced Cell Diagnostics). Briefly, sections were removed from 70°C, placed directly in 4% PFA, and fixed for 1 h at room temp. Next, sections were dehydrated via an ethanol series of 50%, 70%, 100%, and 100% EtOH in water, each immersion for 5 min at room temp. Samples were then treated with protease (Protease Plus for neonate mouse, Protease III for neonate marmoset, adult marmoset, and adult mouse) for 30 min at room temperature. Probes (at a 1:50 dilution factor) were hybridized for 2 h at 40°C. Following the application of amplifiers and fluorophores at 40°C, adult marmoset sections were incubated with TrueBlack Plus (Biotium) at a 1:30 dilution factor (1.5×) for 10 min. Sections were counterstained with DAPI and coverslips were affixed with ProLong Diamond Antifade Mountant (Thermo Fisher). Following each round of imaging, coverslips were removed by soaking slides in 4x SSC, fluorophores were cleaved, and the next set of tails were applied, followed by additional TrueBlack (for adult marmoset sections) and DAPI application (every other round for mouse, every round for marmoset) prior to mounting and acquiring images for the next round. The tail application and imaging order was as follows: T1-3, T4-6, T7-9, and T10-12 for mouse; T7-9, T4-6, T10-12, and T1-3 for marmoset. We changed the round order for marmoset after observing low signal-to-noise for T7-9 when imaged in the third round. Imaging control probes (housekeeping genes *POLR2A*, *PPIB*, and *UBC*) also showed significant loss of signal in the third round of imaging in adult marmoset brain.

Single-round FISH for marmoset rDEGs and sDEGs in both species was performed using the RNAscope™ Multiplex Fluorescent V2 kit and protocol (Advanced Cell Diagnostics. Briefly, sections were removed from 70°C, placed directly in 4% PFA, and fixed for 1 h at room temp. Next, sections were dehydrated via an ethanol series of 50%, 70%, 100%, and 100% EtOH in water, each immersion for 5 min at room temp. Samples were then treated with Protease Plus for 30 min at room temperature. Probes (mixed at 1:1:50 ratios of C2, C3, and C1 respectively) were hybridized for 2 h at 40°C. Following the application of amplifiers (30 min for Amp1, 30 min for Amp2, and 15 min for Amp3), HRP-C1 (TSA fluorescein), HRP-C2 (TSA Cyanine 5), and HRP-C3 (Cyanine 3) signals were developed. Adult marmoset sections with incubated with TrueBlack Plus (Biotium) at 1:30 dilution factor (1.5×) for 10 min. Sections were counterstained with DAPI and coverslips were affixed with ProLong Diamond Antifade Mountant (Thermo Fisher).

Images were acquired using an Olympus Fluoview FV3000 confocal microscope using the multi-area time-lapse (MATL) module to record stage positions for re-use in subsequent imaging rounds. Fluorophores were excited with 405 nm, 488 nm, 561 nm, and 640 nm lasers. Imaging settings were adjusted on a per-experiment (and in rare cases, per-sample) basis due to batch-level variations in signal intensity so as to maximize the dynamic range of pixel intensities. All brain regions within a slide were imaged with the same laser power and voltage settings. Fields of view consisting 3 × 3 grids or smaller (for smaller regions in neonate tissue) with the 20× magnification objective lens were obtained in each brain region of interest (PFC, striatum, and thalamus). We obtained a z-stack covering the entire axial extent of the tissue section with a z-step size of 2 μm. Raw images from the microscope were converted to.tif format, flattened using a maximum intensity projection, and separated into individual channels. For HiPlex images, round 2-4+ images were registered to round 1 using the DAPI channel and cropped to mutually overlapping area using the HiPlex Image Registration Software v2.1 (ACD, https://acdbio.com/rnascope%E2%84%A2-hiplex-image-registration-software-v21), or, in cases where v2.1 failed to register most of the field of view, HiPlex Image Registration software v1.0.0 (ACD, provided by their technical support team).

RNA quality was assessed using positive control probes targeting housekeeping genes for each species. These included *POLR2A*, *PPIB*, and *UBC* for marmoset and *Polr2a*, *PPIB*, *Ubc*, *Hprt*, *Actb*, *Tubb3*, *Bin1*, *Ldha*, *Gapdh*, *Pgk1*, *Bhlhe22*, and *Cplx2* for mouse (RNAscope HiPlex12 Positive Control Probe - Mm). For HiPlex datasets, control experiments were always performed simultaneously with the experiments for rDEG probes of interest, using the same reagents and brain slices of minimal stereotaxic distance to the slice used for rDEG probes. Control probe signal was not quantified, but the brightness and density of probes that are expected to be expressed ubiquitously in most/all cells were manually observed to judge RNA quality for the inclusion or exclusion of experiments. Control probes were omitted for RNAscope v2 experiments, as sufficient RNA quality in neighboring slides with tissue from the same animals had already been confirmed.

Mouse astrocyte labeling and ExR for mouse tissue

We stereotaxically injected 10 week-old adult C57BL/6J mice with AAV2/5 CAG-flex-GFP-4x6T (Addgene 196418) and AAV2/5 gfaABC1D-Cre-4x6T (Addgene 196410) as in Gleichman et al.[112] in prefrontal cortex, striatum, and thalamus (Figure 7A). After 3 weeks of viral expression and pre-expansion staining for GFP, astrocytes were clearly labeled in our regions of interest

(Figure 7B). We found that a single expansion step, yielding an expansion factor of ~3.5×, was sufficient to visualize complex astrocyte morphology in gray matter regions of the mouse PFC, striatum, and thalamus (Figure 7C; Videos S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, and S18). For reference, the effective resolution of ~4× expansion microscopy is ~70 nm.[113] Though not the super-resolution afforded by other techniques such as electron microscopy, ~4× ExR is advantaged by rapid sample preparation, compatibility with conventional antibody staining, and rapid imaging of large volumes on a confocal microscope.

All AAVs in this study were packaged in-house as previously described.[114] Briefly, for each 150 mm culture dish of HEK293 cells, 5.7 μg of construct DNA was transfected with 22.8 μg of capsid plasmid and 11.4 5μg of pADDeltaF6 using polyethylenimine 25K MW (Polysciences, 23966-1). Collection of cells and media for AAV harvesting began 72 h after transfection, followed by iodixanol gradient ultracentrifugation purification using a Type 70 Ti Fixed-Angle Titanium Rotor (Beckman-Coulter, 337922). Titer was calculated using droplet digital PCR (ddPCR) as described by Addgene (https://www.addgene.org/protocols/aav-ddpcr-titration/)[115] using the QX200 AutoDG Droplet Digital PCR System (BioRad, 1864100).

For stereotactic intracranial injection of AAVs, C57BL/6J mice (3 females, 5 males) aged 10 weeks old were anesthetized with isoflurane (5% induction, 1.5% maintenance) and kept on a heating pad. Depth of anesthesia was confirmed via breathing rate and bilateral toe pinch. Mice were given subcutaneous sustained-release buprenorphine (1 mg/kg) on the day of surgery and subcutaneous meloxicam (5 mg/kg) on both the day of surgery and 24 h after. Viruses were delivered using a pulled glass micropipette. A total volume of 500 nL of AAV2/5 CAG-flex-GFP-4x6T and AAV2/5 GfaABC1D-Cre-4x6T, each at final a titer of 1×10^{12} vg/mL, were co-injected at 150 nL/min to the PFC (2.68 AP, 0.75 ML, 2 DV), striatum (1 AP, 1.7 ML, 3.5 DV), and thalamus (-1.46 AP, 1 ML, 3.5 DV). Mice were monitored for 3 days post-surgery to ensure proper recovery.

3 weeks following surgery for adult C57 Bl/6J mice, animals were perfused for ExR as described.[37] Briefly, animals were deeply anesthetized using isoflurane and transcardially perfused with ice-cold 2% acrylamide in PBS followed by ice-cold 30% acrylamide and 4% paraformaldehyde in PBS (by initial volume: e.g., for two mice, we dissolved 15 g of acrylamide in 38.75 mL deionized water, added 5 mL 10x PBS, and 6.25 mL 32% PFA). Brains were post-fixed in the same fixative solution overnight, transferred to 100 mM glycine for 6 h, and stored in PBS at 4°C until sectioning. Brains were sectioned coronally at 150 μm on a vibrating microtome (Leica) and stained for GFP (primary antibody, Abcam chicken-anti-GFP at 1:1000, secondary antibody, AcX-conjugated goat-anti-chicken Alexa Fluor 488 at 1:200, see Table S30 for antibody product information). Briefly, sections were permeabilized in 1x PBS + 0.05 Triton X-100 solution for 10 min at RT, blocked for 2 h at RT in blocking buffer (5% normal goat serum (NGS) + 0.5% Triton X-100 in 1x PBS), incubated with primary antibodies in carrier solution (5% NGS + 0.25% Triton X-100 in 1xPBS) for 12-24 h at 4°C, washed in PBST (1x PBS + 0.1% Triton X-100) 3 times for 10 min each at RT, incubated with secondary antibodies in carrier solution for 12-24 h at 4°C, and washed in PBST 3 times for 10 min each at RT. Slices containing regions of interest (PFC, striatum, and thalamus) were identified using the online Allen Brain Institute adult mouse reference atlas with coronal sections.

One hemisphere containing each region of interest (the hemisphere that was injected, from the slice with the brightest GFP signal) was expanded at ~3.5× expansion followed by staining for GFP, Lectin, and GFAP for morphology characterization. For 2 of the 8 mice, another hemisphere from a neighboring slice was expanded ~18× for super-resolution imaging of astrocytic processes with synaptic proteins and rDEGs. Gels were generated using the ExR protocol.[37] Briefly, tissues were incubated in the first gelling solution for 30 min at 4°C followed by 37°C for 30 min to 2 h. To preserve blood vessel morphology, gels were treated with 0.5 kU/mL Collagenase VII overnight at 37°C, a variation on our lab's previous protocol[116] (Figure S18C). After collagenase treatment, tissue-embedded gels were incubated in ExR denaturation buffer for 1 h at 95°C. Denatured gels were fully expanded in deionized water by washing 2-4 times for 15-45 min each. ~18×-expanded gels were generated without collagenase treatment. We omitted collagenase treatment for the ExR samples for two reasons. First, the goal of this experiment was not to capture astrocyte morphology, which might be locally disrupted by blood vessel breakage (indeed, most of the fields of view we imaged contained no blood vessels). Nevertheless, despite some broken blood vessels, astrocyte morphology appeared largely continuous in non-collagenase treated samples, except specifically at astrocyte contact sites with blood vessels (Figure S18D). Second, collagenases could in principle be contaminated with other proteases, which might degrade sensitive epitopes and reduce signal captured by ExR.

For the ~3.5×-expanded gels for morphology analysis: After full expansion, we shrunk the gels for easier handling by incubating in 10x PBS for 10-30 min. We then transferred gels to blocking buffer (5% NGS and 0.5% Triton X-100 in 1x PBS) and incubated for 90 min - 2 h at room temperature. Primary antibodies (chicken-anti-GFP and mouse-anti-GFAP, see Table S30 for antibody product information) were incubated at 4°C for 16-24 h in antibody carrier solution (5% normal donkey serum (NDS) and 0.25% Triton X-100 in 1x PBS) at a dilution factor of 1:200. Gels were washed in 0.01% Triton X-100 in 1x PBS 6 times for 15 min each at room temperature, and then incubated with secondary antibodies (donkey-anti-goat AF488, donkey-anti-chicken AF488, donkey-anti-mouse AF555, and Lycopersicon Esculentum (Tomato) Lectin (LEL, TL), DyLight 649, see Table S30 for antibody product information) were incubated at 4°C for 16-24 h in antibody carrier solution (5% normal NDS and 0.25% Triton X-100 in 1x PBS) at a dilution factor of 1:200. Gels were washed in 0.05x PBST (e.g., 500 μL Triton X-100 in 1xPBS, 25 mL of 1x PBS, up to 500 mL of DIW) 6 times for 15 min each at room temperature to expand the gels for imaging.

For the ~18×-expanded gels for rDEG protein expression analysis: After full expansion, we shrunk the gels for easier handling by incubating in 10x PBS for 10-30 min. We then transferred gels to blocking buffer (5% NGS and 0.5% Triton X-100 in 1x PBS) and incubated for 90 min - 2 h at room temperature. Primary antibodies (against GFP, Cav2.1, and Gat3 or EEAT1, see Table S30 for antibody product information and dilution factors) were incubated at 4°C for 12-24 h in antibody carrier solution (5% NDS and

0.25% Triton X-100 in 1x PBS) at a dilution factor of 1:100 to 1:200. Gels were washed in 0.01% Triton X-100 in 1x PBS 6 times for 15 min each at room temperature, and then incubated with secondary antibodies (donkey-anti-goat AF488, donkey-anti-chicken AF488, donkey-anti-mouse AF555, and Lycopersicon Esculentum (Tomato) Lectin (LEL, TL), DyLight 649, see Table S30 for antibody product information) were incubated at 4°C for 12-24 h in antibody carrier solution (5% NDS and 0.25% Triton X-100 in 1x PBS) at a dilution factor of 1:200. Gels were washed in 0.05x PBST (e.g., 500 μL Triton X-100 in 1xPBS, 25 mL of 1x PBS, up to 500 mL of DIW) 6 times for 15 min each at room temperature to expand the gels for imaging.

The expansion factor for each gel in the morphology set was measured before and after expansion (in 0.05x PBST, the buffer used for washing before imaging) using a ruler (pre-expansion) and landmarks in the gel from tiled overview images obtained at 4× magnification (post-expansion). Measurement lengths smaller than 0.1 cm, the smallest demarcation on the ruler, were estimated. We used the average expansion factor across samples for a given brain region or experiment to convert from physical (pre-expansion) to biological (post-expansion) units for scale bars, area, and volume calculations (Table S31). For the low expansion factor morphology dataset, final expansion factors for PFC, striatum, and thalamus were 3.65, 3.43, and 3.34, respectively. Images were obtained on an inverted Nikon w1 confocal microscope with a 40× water magnification lens. Images for the lower expansion factor morphology dataset were obtained at a 0.5 μm z-step with 200 ms exposure and 100% laser power for each optical channel. For some image stacks, the Lectin channel (640nm) was inadvertently obtained with 50 ms exposure (as noted in Table S32), but this channel was not used for quantification, and only one of such astrocytes is shown in Figure S7 (for this astrocyte, the maximum Lectin contrast was decreased by 50% so that its Lectin contrast is similar to the others in the figure). We imaged the full z-stack for each channel separately using the Ti Z-drive. When possible, we imaged the astrocytes that were: (1) sufficiently brightly labeled with GFP, (2) fully contained within the gel (volume not cut off, though many astrocytes were partially cut off for the experiment in Figure 7, as noted in Table S32), (3) not overlapping with other brightly labeled astrocytes (though many fields of view did contain part of a second astrocyte, often more dimly labeled, for the experiment in Figure 7, as noted in Table S32).

The expansion factor for the twice-expanded ExR gels for rDEG visualization was estimated at ~18× in 0.05x PBST based on measurements using the same protocol in our prior studies.[37,117] Images for ~18× expanded ExR gels were obtained at 0.5 μm z-step with 1 s exposure and 100% laser power for each optical channel on a Nikon SoRa confocal microscope. We imaged 2048 × 2048 × 51 voxel volumes, with each z-step incrementing only after all 3 channels were imaged. In several fields of view, there are portions of empty gels lacking tissue, likely resulting from tissue being improperly anchored to the gel (see Table S32). These regions of empty did not affect our quantification of rDEG intensity in astrocytes across brain regions, as they were masked out during the GFP segmentation process (see "analysis of high-expansion factor ExR rDEG target images" section below). Furthermore, we only included fields of view containing mostly tissues for quantification.

QUANTIFICATION AND STATISTICAL ANALYSIS

Read alignment

Reads were aligned to an optimized mouse reference genome based on the mouse GRCm38 primary sequence assembly (version 2)[103] or or a modified marmoset mCalja1.2.pat.X assembly (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_011100555.1/) provided courtesy of Michael DeBerardine and Fenna Krienen (Princeton Neuroscience Institute) in which mitochondrial genes from CM021961.1 (https://www.ncbi.nlm.nih.gov/nuccore/1820101357/) were annotated using MITOS2.[118] The marmoset reference genome was generated from the .fasta and .gtf files using Cell Ranger "mkref" v7.1.0 using default parameters without filtering for any gene/transcript biotype and is available at: https://doi.org/10.5281/zenodo.16915694. 10× Genomics Cell Ranger software version 7.1 was used for alignment and counting via the 10× Genomics Cloud Analysis platform. For samples that were sequenced across several library pools, fastq files were grouped prior to alignment to create one cell-by-gene (cell × gene) counts matrix per sample. CellBender (v0.2.0) remove-background[106] was used to remove ambient RNA and call nuclei with default parameters and expected_cells = 10,000, total-droplets-included = 40,000 and the –cuda flag. CellBender-cleaned cell × gene matrix.h5 files were read into Python in the anndata[119] format using a custom function written by Stephen Fleming (https://github.com/broadinstitute/CellBender/issues/57).

Calculation of sequencing coverage statistics

To determine whether we achieved our target of 40,000 sense reads per nucleus and calculate the sequencing coverage statistics shown in Figure S1 and Table S28, we ran a light quality control on CellBender-cleaned cell × gene matrices. Briefly, nuclei with fewer than 1,000 unique molecular identifiers (UMIs) and fewer than 800 genes expressed were removed, as were genes with nonzero expression in fewer than 10 cells. We note these cutoffs are more stringent than what we used for preprocessing (see "snRNA-seq data preprocessing and quality control" section), to account for the lack of doublet removal and additional manual curation that is much more time consuming. Nuclei with greater than 4% of reads aligning to the mitochondrial genome (prefix "mt-" for mouse or "MT-" for marmoset) were removed.

snRNA-seq data preprocessing and quality control

All scripts used for preprocessing and downstream analyses are available on GitHub at https://github.com/Feng-Lab-MIT/AstrocyteHeterogeneity(version of record, release v0.1, https://doi.org/10.5281/zenodo.16911362). Preprocessing was conducted on a species-wide, cross-age, cross-region basis. Filtered counts matrices were pre-processed using scanpy.[107] Briefly, nuclei with fewer than 800 unique molecular identifiers (UMIs) and fewer than 500 genes expressed were removed, as were genes with nonzero expression in fewer than 10 cells. Nuclei with more than 4% of counts annotated as mitochondrial genes (prefix ‘‘MT-’’ in marmoset or ‘‘mt-’’ in mouse) were removed. Counts were normalized to 1 million per cell and log-transformed using scanpy’s ‘‘log1p’’ function. Highly variable genes (HVGs) were identified from raw counts on a per-batch basis using scanpy’s ‘‘seurat_v3’’ method with 4,000 top genes. HVGs present in less than 10% of batches or of mitochondrial origin (prefix ‘‘MT-’’ in marmoset or ‘‘mt-’’ in mouse) were removed. We used the scanpy-based single cell variational inference (scVI[108]) package to create and train a variational autoencoder on a subset of the cell × gene matrix with highly variable genes only with the following parameters: batch_key corresponding to 10× Genomics reaction, raw counts layer, ‘‘gene-batch’’ dispersion, and training with GPU. The resulting nonlinear embedding was used to create a neighborhood graph for clustering and calculate UMAP[120] coordinates with the scanpy ‘‘sc.pp.neighbors’’ and ‘‘sc.tl.umap’’ functions. scVI and the scvi-tools package were the basis for many downstream analyses. We used Solo, an automated doublet removal package[110] based on the scVI model, to calculate doublet scores for each nucleus on a per-batch basis. For the mouse data, one nucleus was removed from the ‘‘Exp074_mmP35_2D’’ 10× reaction to circumvent a known bug in the Solo package (https://discourse.scverse.org/t/solo-scvi-train-error-related-to-batch-size/1591/2). Doublet/singlet thresholds were determined manually by examining the doublet- vs. singlet-score scatter plot and a predicted doublet rate of 13-15%, a conservative estimate based on 10× Genomics’ predicted ~8% for a target recovery of 10,000 nuclei. We decided to use this manual threshold to prevent removal of developing cells that are more likely to be flagged doublets automatically.

Global (species-wide, cross-age, cross-region) integration and annotation

After automated doublet removal, highly variable genes were re-calculated and the scVI model was re-trained. Leiden clustering was performed with a resolution of 1. Top differentially expressed genes in each cluster (i.e., putative marker genes) were identified using scanpy’s rank_genes_groups function with log-normalized counts and the Wilcoxon rank-sum method. Top marker genes, expression of known marker genes, and dendrograms were used to annotate clusters per the following convention: [Cell class]_[Excit or Inh*]_[region*]_[cortical layer*]_[age*]_[known marker*]_[first rank_genes_group marker]-[second rank_genes_group marker*], where the asterisked attributes were used variably, as applicable. Low-quality clusters were manually identified and removed, highly variable genes re-calculated, the scVI model re-trained, and clustering and annotation processes were repeated. Remaining doublet clusters were manually removed, and the neighborhood space and UMAP coordinates were recalculated. Neuronal annotations were subsequently refined based on predicted mapping to the Allen Brain Cell Atlas using MapMyCells (see alignment to Allen Brain Cell Atlas (ABCA) with MapMyCells section below).

Adult (29-32 month, referred to as 30 month in most of the paper) marmoset data used in this study was generated previously,[7] and includes data from 4 donors. CellBender background-removed cell × gene matrices for the four regions of interest (annotated as ‘‘pfc’’, ‘‘m1’’, ‘‘striatum’’, and ‘‘thal’’) were preprocessed as described above, except that mitochondrial genes were not annotated and therefore not used for quality control. Adult marmoset cell × gene matrices were randomly downsampled to 40,000 nuclei per region to match the approximate number of nuclei in each age-region combination in the developmental dataset. Of note, the adult marmoset snRNA-seq reads were aligned to the cj1700 transcriptome, which lacks mitochondrial genome annotation (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009663435.1/). 22,582 genes overlapped between the mCalJa1.2.pat.X (developmental and aged data, 31,308 genes) and cj1700 (adult data, 27,304 genes) reference-aligned datasets. Data from GD135, neonate, 7-month, 14-month, and aged timepoints were integrated with the downsampled adult (29-32 month) data using scVI using only highly variable genes, clustered, and annotated and described above. Clusters with the vast majority or all nuclei derived from adult (29-32 month) marmoset data were removed, as they likely derive from differences in dissection strategies between the two studies. We also removed a small subcluster of 233 adult astrocytes that clustered with immature astrocytes, because they derived primarily from a single adult replicate and were not found in other donors 14 months and older. Neuronal annotations were subsequently refined based on predicted mapping to the Allen Brain Cell Atlas using MapMyCells[41] (see alignment to Allen Brain Cell Atlas (ABCA) with MapMyCells section below). In designing and implementing downstream analyses (e.g. compositional, pseudotime, and cell-cell interaction analyses), we relied heavily on the Single Cell Best Practices e-book (https://www.sc-best-practices.org/).[121]

Integration quality analysis

We utilized code from Cao and Gao[38] to quantify integration quality separately for each species using neighbor consistency, average silhouette width, and donor mixing. Neighbor consistency relied on coordinates in a pre-integration embedding and post-integration embedding, for which we used scanpy’s PCA function (50 components) to recalculate the pre-integration embedding and scVI for the post-integration embedding. We calculated the average silhouette width with a wrapper function of the sklearn package with the scVI embedding, and used ‘‘supercluster’’ as the group labels. Finally, donor mixing was determined using the ‘‘seurat-alignment_score’’ function with the scVI embedding and donor ID.

Alignment to Allen Brain Cell Atlas (ABCA) with MapMyCells

MapMyCells was accessed at https://portal.brain-map.org/atlases-and-data/bkp/mapmycells (RRID: SCR_024672). Both mouse and marmoset mature neurons (GABAergic, glutamatergic, and mixed from mouse P14 and above or marmoset neonate and above) or astrocytes were subsetted from the larger cell × gene matrix of each species by age and region, with some metadata removed to shrink the file size. To use the maximum number of genes available for each marmoset time point, pre-integrated adult or developmental gene counts (that is, with ~27,000 genes for 29-32 month marmoset data and ~31,000 genes for developmental and aged data) were used as input. Per the MapMyCell input requirements (https://portal.brain-map.org/explore/file-requirements-and-limits), cell × gene matrix entries were set to raw counts and NCBI gene symbols were converted to mouse Ensembl gene IDs using g:Profiler's g:Convert.[122] Marmoset NCBI gene symbols were converted to mouse NCBI gene symbols using a table downloaded from Ensembl BioMart (https://useast.ensembl.org/info/data/index.html),[123] available in Table S29. If no match was found in the table, the marmoset gene name was converted to sentence case. Due to non-uniqueness after converting marmoset IDs to mouse, *Bex2* was removed from the adult marmoset neuron cell × gene matrix, while *Cstl* was removed from the developmental and aged marmoset neuron cell × gene matrices, before mapping to Ensembl IDs. ~25,300 mouse and ~14,800 marmoset genes were successfully mapped to Ensembl IDs, and the others were excluded from MapMyCells analysis. These smaller (<2Gb) cell × gene matrices were uploaded to the Allen Brain Maps's MapMyCells web portal (https://knowledge.brain-map.org/mapmycells/process/) and aligned to the 10× Whole Mouse Brain (CCN20230722) reference taxonomy[1] with the hierarchical mapping algorithm. Using the output of MapMyCells, which is a "class", "subclass", "supertype", and "cluster" assignment for each barcode, we examined the most abundant (as the percentage of our cells in each of our clusters mapping to that subclass) ABCA subclass (about the same taxonomic rank) for each of our cross-region, cross-age, within-species embedding Leiden clusters. Broadly, our annotations were identical or highly similar to the most abundant ABCA mappings. When our annotation disagreed with or was not as specific as the most abundant ABCA subclass mapping, and we were not confident in our annotation, we updated the annotation based on the most abundant ABCA subclass (e.g., comprising over 60% of the cells in the cluster for mice, or over 20% for marmoset). For clusters that had a more uniformly distributed mapping onto ABCA subclasses (e.g., less than 10% mapping onto each subclass), we examined several of the top mapping clusters and their anatomical locations (using the web resource from Yao et al.[1] available at https://knowledge.brain-map.org/data/5C0201JSVE04WY6DMVC/summary), and if they were in the same taxonomic or anatomical neighborhood, annotated our cluster accordingly. In a few cases, such as for marmoset thalamic neurons mapping to a midbrain ABCA population and immature astrocytes mapping to the Allen olfactory bulb/immature neuron subtype, we did not adopt the ABCA subclass label. The class, subclass, supertype, and cluster mappings for each of our leiden clusters (as proportion of cells in that cluster mapping to each ABCA taxonomic rank) for both species are provided in Table S4.

Compositional analysis with scCODA

The proportional breakdown of each leiden cluster by developmental time point (age), brain region, and sex are provided in Table S3 (assigned brain region) and displayed in Figure S4 (dissected brain region). To more quantitatively assess cell type composition changes across these variables, we used the scanpy-based single-cell compositional data analysis (scCODA)[41] (https://github.com/theislab/scCODA), which implements a Bayesian model of cell type counts to address the issue of low sample sizes in snRNA-seq data. For this analysis, we merged the PFC and MO into one "cortex" assignment, due to their high degree of similarity in cell type composition. Per the tutorial in the Single Cell Best Practices e-book (https://www.sc-best-practices.org/)[121] Section 17.4, we generated an scCODA model of type "cell_level" with the cell type identifier as either "leiden" (cluster) or "cell_type", sample identifier as "10x_batch" and "sex" and age, region, sex, and replicate as covariate observation. We ran the model with the formula "region + age + sex", automatic selection of reference cell type for leiden-level analysis and Mural (marmoset) or Astrocyte (mouse) as the reference cell type for cell type-level analysis, and default false discovery rate of 0.05. The "final parameter", which is a boolean value that indicates whether or not there is a significant effect of age, region, or sex on the composition of each cell type and cluster for both mouse and marmoset are provided in Tables S5, S6, S7, and S8.

Region reassignment of cross-contaminant nuclei in fetal and neonate marmoset

We observed a modest amount of cross-region contamination, particularly between striatum and thalamus in our late embryonic and neonate samples, which is not unexpected given our coarser dissection strategy for these regions (Figures S5 and S6). For example, in marmoset, 4% of medium spiny neurons came from the thalamic dissection and 23% of thalamic excitatory neurons, which were mostly from GD135 and neonate time points, came from the striatal dissection. To reduce the effect of this cross-contamination on downstream analyses, we reassigned the region annotation for neurons, astrocytes, and OPCs (the most strongly region-segregated cell types on which we focused our analysis) at these ages. We subsetted the cell type of interest from the cross-age, cross-region integrated cell × gene matrix, re-computed neighbors using the scVI latent space, and recomputed the UMAP space for each cell type. We then assigned each nucleus as either *FOXG1*+ or *FOXG1*− based on an expression threshold of 4 logCPM, and performed Leiden clustering at low resolution (0.3). Any GD135/neonate astrocyte nucleus from the thalamic dissection that co-clustered with primarily telencephalic astrocytes was assigned striatum if it mapped to the telencephalic astrocyte ABCA subclass. Any GD135/neonate astrocyte from the striatum dissection in the same Leiden cluster was the thalamic nuclei was assigned thalamus if it was *FOXG1*− and it mapped to the non-telencephalic ABCA subclass. Any GD135/neonate astrocyte from the thalamic dissection that did not meet the first condition but was *FOXG1*+ or mapped to the telencephalic ABCA subclass was assigned striatum. Any GD135/neonate GABAergic neurons from the striatum clustering with thalamic GABAergic neurons (either TRN or midbrain-derived) was assigned thalamus, any thalamic neuron clustering with medium spiny neurons was assigned striatum, and any *FOXG1*+ nucleus from the thalamic dissection was assigned striatum.

Any GD135/neonate neurons clustering with thalamic glutamatergic neurons were assigned thalamus, and any GD135 glutamatergic neurons not clustering with thalamic glutamatergic neurons were assigned PFC. Since OPCs did not cluster by region, we did not reassign region based on clustering, but instead, used only *FOXG1*: any OPCs from the thalamic dissection that were *FOXG1*+ were assigned striatum. The resulting assigned region annotation is saved in the ''region''.obs variable of the annotated data files, while the original region is saved in ''region_dissected''.

Region reassignment of cross-contaminant nuclei in mouse

Cross-region contamination was higher in the E18.5 and P4 mouse timepoints, but present in all ages, primarily between striatum and thalamus (Figure S6). Similarly to marmoset, we subsetted the cell type of interest from the cross-age, cross-region integrated cell × gene matrix, re-computed neighbors using the scVI latent space, and recomputed the UMAP space for each cell type. We then assigned each nucleus as either *Foxg1*+ or *Foxg1*− based on an expression threshold of 4 logCPM, and performed Leiden clustering at low resolution (0.3). Nuclei in immature astrocyte clusters composed of mixed regions were reassigned based on *Foxg1* expression or assignment in ABCA subclass by MapMyCells. Astrocytes from the thalamic dissection in these mixed region clusters were reassigned as striatum if they were *Foxg1*+ or had the subclass of ''Astro-TE-NN''. Immature clusters exhibiting more clear telencephalic and diencephalic divisions (mostly from P4) were manually reassigned, either from thalamus to striatum, or striatum to thalamus, to match the predominant region of origin for that cluster. All thalamic dissected astrocytes in the predominantly telencephalic and P4 cluster were manually reassigned to striatum. For the primarily thalamic and P4 cluster, we reassigned all cells to match their ABCA subclass (all ''Astro-TE NN'' being labeled striatum and all ''Astro-NT NN'' being labeled thalamus) and any cells that were *Foxg1*+ were also labeled striatum. Because this cluster contained *Foxg1*+ cells, we did not reassign all nuclei in it to thalamus. For the mature astrocytes, thalamic nuclei were reassigned to striatum if they clustered with telencephalic astrocytes and were assigned the ''Astro-TE NN'' ABCA subclass, and vice-versa for striatal astrocyte nuclei clustering with thalamic astrocyte nuclei. For OPCs, thalamic dissected cells were reassigned striatum if they were *Foxg1*+. For excitatory neurons, any striatum dissected cells were reassigned to cortex or thalamus based on the predominant region of the nuclei in their cluster. Any thalamic dissected excitatory neurons were reassigned as cortex if they clustered with cortical excitatory neurons and were *Foxg1*+. Finally, for inhibitory neurons, any thalamic dissected cells outside of the TRN cluster in E18.5 or P4 were reassigned either cortex or striatum based on the prominent region of the cluster they grouped with, and any cell at E18.5 or P4 in the excitatory TRN cluster were reassigned as thalamus. At all other time points, any thalamic excitatory neuron that was *Foxg1*+ was reassigned as cortex. The resulting assigned region annotation is saved in the ''region''.obs variable of the annotated data files, while the original region is saved in ''region_dissected''. 4.4% of astrocytes, 0.2% of OPCs, 2.6% of excitatory neurons, and 3.8% of inhibitory neurons were reassigned from their dissected region. There was also a batch of 90 weeks nuclei that was mislabeled prior to sequencing as striatum, but reassigned to cortex due to almost all cells aligning to cortical clusters within our data and to the ABCA clusters (see ''Notes'' column in Table S1).

Calculation of rDEGs, aDEGs, and sDEGs using a pseudobulk method

Regional differentially expressed genes (rDEGs, Figure 2) were calculated as previously described[7] with some modifications. rDEGs were calculated for each individual developmental time point, cell type, and species separately. To use the maximum number of genes available for each marmoset time point, pre-integrated adult or developmental gene counts (that is, with ~27,000 genes for adult marmoset data and ~31,000 genes for developmental and aged data) were used for rDEG calculation. We first created per-region metacells for each cell type (astrocyte, OPC, GABAergic neuron, and glutamatergic neuron) by averaging the raw counts of all cells of each cell type per region and per replicate. If a metacell of one region, cell-type, and replicate combination had fewer than 50 cells, it was omitted. Because they had few rDEGs between them, motor and prefrontal cortices were grouped for rDEG analysis. Metacell counts were normalized to 100,000 counts total and log10-transformed. We required rDEG candidates have at least 10 transcripts per 100,000 in at least one metacell per region and be expressed in at least 33% of nuclei in the metacell. 33% was chosen to require a significant portion of cells in a metacell to express the gene, but also be low enough to account for dropout.[124] Genes with $>10^{0.5}$ (3.16) fold-change (logFC) expression in the same cell type between two regions (pairwise) were considered rDEGs.

For marmoset, at the 30-month timepoint we restricted this analysis to the 2 30-month donors (bi005 and bi007, from our previous study[7]) that were represented in each regions. Because we averaged and normalized gene expression across cells within a region/age/species during metacell creation to calculate differential expression, we do not explicitly account for the relative contribution of each biological donor and/or replicate to differential expression. For this reason, for marmoset rDEGs, where the biological donors are balanced within each developmental time point (i.e., each animal donated each brain region), we filter rDEGs to those shared by at least 2 biological donors. For the mouse data, in which some regions were donated by different animals within a time point due to the aforementioned failures, and for age- and species- comparisons, which are inherently not replicate-balanced (no donor is represented in multiple ages or species), we do not enforce this criterion. Thus, for mouse rDEGs, aDEGs, and sDEGs, we merged all biological replicates within a group into one metacell. As an alternative source of replicate specificity information, we calculated the fraction of replicates in each group that had at least 33% of cells expressing each gene.

To plot rDEG expression heatmaps as in Figures 2B and 3B, we plotted the rDEG lists from fetal, early adolescent, and adult time points, ordered first by the age the gene was differentially expressed and second by the region(s) that the gene that was more highly expressed in, including repeats if an rDEG was detected at multiple time points. Marmoset rDEGs were only included if they were present in both replicates of a given time point, while mouse rDEGs did not have a replicate restriction. Because of the high heterogeneity in the mouse striatal astrocytes due to the multiple immature populations, these cells were reordered to place cells of similar populations together in the heatmap. To do so, striatal astrocytes were clustered at a low leiden resolution (0.1) and these clusters were ordered using a combination of the pseudotemporal ordering and the ages present in the cluster. A similar logic was applied to order marmoset striatal nuclei in the rDEG expression heatmaps.

Raster plots underneath the rDEG expression heatmaps were generated as separate figures using custom Python scripts with the help of ChatGPT 4.0. In brief, we created a matrix encoding the age and region(s) for which each gene was an rDEG, and plotted a line of the corresponding color in the corresponding raster row. These separate figures were manually aligned below the scanpy-generated rDEG expression heatmaps to the best of our abilities, but the x-axes may not be exactly aligned. To create the UpSet[125] plots in Figures 2, 3C, 3D, 5C, 5D, 6F, S16B, and S16C, we used the package UpSetPlot (https://github.com/jnothman/UpSetPlot). The ''active'' dots were manually colored according to the group's color scheme for clarity.

The pairwise region rDEG scatter plots in Figure S7 were generated as follows. First, we recalculated rDEGs for each species within each region pair using the same metacell method described above, with one modification to allow more lowly expressed genes to pass the filter: requiring a gene be expressed by a minimum of 33% of cells in *any* region of either cortex, striatum, or thalamus. We then plotted the log fold-changes (calculated as described above) for each gene in (x,y) with x being the logFC value for one region pair (e.g. cortex-striatum), and y being the logFC value for another region pair (e.g. cortex-thalamus). The sign on the logFC value was determine with respect to the region shared across the pairwise comparisons (e.g. cortex). Pearson's correlation coefficient, r, was calculated using the scipy stats package (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html). A gene was marked as rDEG in one region if it had greater than 0.5 magnitude logFC in that region, or both if it had greater than 0.5 magnitude logFC in both.

aDEGs (Figure 5) were calculated in a similar manner, except for that the metacell axis was age instead of region, and that no cross-replicate consistency was imposed for either species, because no donor or biological replicate provided tissue for multiple developmental time points. For aDEG and rDEG calculation in marmoset, lateral septal GABAergic neurons and putative hippocampal Cajal-Retzius neurons were removed, as we could not assign their region to either cortex, striatum, or thalamus, before metacell generation. Similarly for mouse, we omitted lateral septal GABAergic neurons and glutamatergic neurons from the anterior olfactory nucleus before aDEG and rDEG calculation. To calculate species differentially expressed genes (sDEGs, Figure 6), we first combined downsampled mouse and marmoset datasets of all developmental time points and brain regions using the intersection of genes with 1:1 orthologs converted to mouse gene IDs using Table S29. If no match was found in the table, the marmoset gene name was converted to sentence case. We then created metacells of each supercluster for each species, where species was the metacell axis, and compared gene expression between species within each supercluster. As for aDEGs, we did not impose a cross-replicate requirement on sDEGs for either species because no replicate can be a member of both species.

To aid the interpretation of region-shared vs. region-specific aDEG profiles, we divided astrocyte aDEGs into 3 groups: group CA-RS (cell type agnostic, region specific) are astrocyte aDEGs that are also aDEG in neurons and OPCs for a given brain region; group AS-RA (astrocyte specific, region agnostic) are shared between astrocytes of all brain regions; and group AS-RS (astrocyte specific, region specific) are specific to astrocytes in a given brain region. CA-RS aDEGs are shared between all analyzed cell types in a given region and not shared with other regions. We found very few (3 or less) CA-RS aDEGs within marmoset GD135-14-month comparisons. Full lists of all mouse and marmoset aDEGs and associated UniProt GO annotations (as for rDEGs) for P4 vs. P90 and GD135 vs. 14-month comparisons respectively, including CA-RS, AS-RA, and AS-RS lists, are provided in Tables S19 and S20.

Group AS-RA aDEGs reflect universal aspects of astrocyte transcriptional identity at each developmental stage, regardless of brain region. We found 74 of these within marmoset GD135 vs. 14-month aDEGs and 56 within mouse P4 vs. P90 aDEGs. Group AS-RS aDEGs reflect the brain region's influence on the maturation of astrocytes *only* in a given brain region. In contrast to group CA-RS (above) they are unique to astrocytes (vs. OPCs and glutamatergic and/or GABAergic neurons) in a given brain region, and (unlike group AS-RA) are not shared with astrocytes in other brain regions. We found 20 of these in striatum, 51 in cortex, and 125 in thalamus within GD135 vs. 14-month aDEGs. Plotted genes (Figures 5F and S16D) were manually selected to include genes that are both developmentally up- and downregulated, and which have varied annotated functions that may be of interest to those studying astrocyte biology and/or neuron-astrocyte communication, but are not meant to suggest increased interest or importance compared to other cortical group AS-RS aDEGs.

We calculated the overlap between rDEGs and aDEGs between mouse and marmoset (Figures 6B and 6C) by converting marmoset gene names to mouse gene names based on 1:1 orthologs using Table S29 (as before, if no match was found in the table, the marmoset gene name was converted to sentence case) and taking the intersection between the two lists. To calculate whether a marmoset astrocyte rDEG was more likely to be a mouse astrocyte rDEG and vice-versa, we performed a Fisher's exact test using scipy stats's ''fisher_exact'' (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html). 2 × 2 contingency tables were generated at each time point as follows: first row, first column was the number of overlapping rDEGs. First row, second column was the number of mouse rDEGs that were not marmoset rDEGs. Second row, first column was the number of marmoset rDEGs that were not mouse rDEGs. Second row, second column was the number of non-rDEG (in either species) overlapping 1:1 orthologs that men minimum expression criteria (at least 10 transcripts per 100,000 in at least one metacell per region and be expressed in at least 33% of nuclei in the metacell) to be considered as rDEGs. The test was run analogously for aDEGs. These tests are implemented in the Jupyter notebook "rDEGs_overlaps_xSpecies_current.ipynb" and "aDEGs_overlaps_xSpecies_current.ipyb".

Automated UniProt annotation and querying of SFARI Gene 3.0

Code to query UniProt programmatically was modified from the companion document to the publicly available "Programmatic access to UniProt using Python" webinar at https://colab.research.google.com/drive/1i9UtVqa4m9WQ4ZVJkbGdwWbto_W7zmP_. Our scripts are available on our GitHub repository with the "mine_uniprot" prefix within rDEG, aDEG, and sDEG subfolders. Briefly, gene symbols from rDEG/aDEG/astrocyte sDEG lists and the appropriate species taxonomy ID were used to search for the UniProt accession number for each gene. The first result's primary accession number was then used to return the full protein name, GO Cellular Component, GO Biological Process, and GO Molecular Function for each protein. Multiple results were concatenated with commas and populated in a data frame for export with r/a/sDEG lists. Genes with no results were populated with "NaN" (empty) values. Genes beginning with the prefix "LOC" (unnamed) or "MT-" (mitochondrial) were ignored.

SFARI 3.0 genes from the 4/3/25 release were downloaded from the SFARI Human Gene Module (https://gene.sfari.org/database/human-gene/). Telencephalic and diencephalic sDEGs that were higher in marmoset (vs. mouse, hereafter marmoset sDEGs) were converted to human gene symbols based on 1:1 orthologs using Table S29 (if no match was found in the table, the marmoset gene name was converted to all capital letters). For each set of marmoset sDEGs (diencephalic and telencephalic astrocytes), the following contingency table was constructed: number of SFARI genes that marmoset sDEGs in first row, first column; number of SFARI genes that were not marmoset sDEGs in first row, second column; number of non-SFARI marmoset sDEGs in second row, first column; and number of non-SFARI, non-sDEG genes (assuming 20,000 protein-coding genes in the human genome) in the second row, second column. A Fisher's exact test was run on this contingency table using Scipy stat's "fisher_exact" function in Python.

Astrocyte subclustering within each brain region

We conducted sub-clustering for subsetted astrocytes from each brain region for each species separately (Figures S12 and S13). For marmoset, using the subsetted cell × gene matrix for each region (all cell types), we removed small 10× Chromium batches with smaller than 500 cells (from reassigned regions), highly variable genes were recalculated (minimum number of batches equal to the floor of the total number of 10× Chromium batches divided by 10, to avoid batch-specific differentially expressed genes while including region-specific variation), the scVI model re-trained, and the neighborhood graph and UMAP coordinates re-calculated on subsetted astrocytes as described above. For mouse, highly variable genes were recalculated per region based on the original dissected region rather than the reassigned region due to a known issue with small batch sizes with the sc.pp.highly_variable_genes function. Our sub-clustering procedure was inspired by earlier methods.[61] Scanpy's "tl.leiden" function was used to identify clusters over a range of decreasing resolution parameters, purposefully starting with an intentionally high resolution that led to over-clustering. At each resolution, the minimum number of pairwise and one-versus-rest (where one group is compared to a metacell of all other groups combined) differentially expressed genes (DEGs) for each cluster and pair of clusters was calculated using the metacell method described above for rDEGs/aDEGs/sDEGs, with metacells calculated for each cluster, not incorporating replicate information. We first conducted a course resolution scan with increments of 0.05, followed by a fine-grained resolution scan in the target resolution range with increments of 0.01. We chose a clustering resolution that resulted in at least 3, but often 10 or more, pairwise and one-vs-rest DEGs for each subcluster and subcluster pair. Subclusters were named according to the following convention: [abbreviated_region]_Ast[subcluster number]_[known subtype name (immature, fibrous, and/or protoplasmic)]_[top marker genes]. More details are available in the scripts used to generate subclusters on our GitHub repository, with the prefix "subcluster_mouse_astrocytes" for mouse, or "integrate_xAge_subcluster" for marmoset.

Pathway analysis with WebGestalt

We employed WebGestalt 2024[53,54] (https://www.webgestalt.org/) over-representation analysis (ORA) for pathway analysis. For the reference gene set, we used either all genes present in the final mouse counts matrix ("adata.var_names"), all genes present in the fetal, neonate, 7-month, 14-month, and aged marmoset data (generated in the current study with the mCalja1.2.Pat.X reference), or all genes present in the adult (29-32 month) marmoset data (generated previously[7] and aligned to cj1700). We used human as the host species for marmoset, mouse as the mouse species for mouse, the Gene Ontology (GO) Biological Process[126,127] noRedundant and KEGG[128] pathway functional databases, weighted set cover redundancy reduction for pathway display, and default advanced WebGestalt settings. Because differentially expressed genes were calculated in a pairwise manner, pathway analyses were performed on genes differentially expressed in both directions (i.e., either up- or downregulated, as opposed to unidirectionally). Lollipop plots were generated from "Description", "Ratio", and "FDR" columns of the weighted set cover redundancy-reduced pathway table using custom Python scripts with the assistance of ChatGPT 4.0.

Cross-species integration with scANVI

scANVI[72] is a semi-supervised variational autoencoder variant of the previously described scVI model that utilizes cell type label information in its latent space. We used a random downsample of 20,000 cells from any age or region for each species. Next, we added a less granular "supercluster" annotation to each cell based on its leiden cluster annotation. For example, two marmoset cowrtical excitatory neuron L6IT clusters were combined into one supercluster. In some cases, the leiden cluster to supercluster mapping was one:to:one, as for OPCs, lateral septal inhibitory neurons, diencephalic and telencephalic astrocytes, and others. To create the embedding space, we used a small subset of highly variable genes for cross-species integration that was inspired by an earlier approach.[61] Briefly, we converted the marmoset gene names to their mouse orthologs and subsetted both datasets to the intersection of shared genes. Next, we took the top 50 most highly expressed genes in each supercluster (as determined by a Wilcoxon rank-sum test in scanpy) in each species separately, then used the intersection of those lists as the highly variable genes entered into the model. We created a scANVI model using the superclusters as the labels key, the dispersion set to ‘‘gene-batch’’, and calculated on the raw counts layer. Finally, we calculated the neighborhood graph and UMAP using scanpy’s previously mentioned functions.

Cross-species integration with SATURN

SATURN[73] is a deep learning method used to integrate cells from multiple species in the same low-dimensional space. It utilizes the ESM2[129] language model and reference genomes from Ensembl (for marmoset: https://useast.ensembl.org/Callithrix_jacchus/Info/Index, https://useast.ensembl.org/Mus_musculus/Info/Index for mouse, as in the original publication) to generate protein embeddings that predict similarity of genes across species. The marmoset protein embedding space was generated by the lead author of the SATURN study and is linked at https://github.com/snap-stanford/SATURN/issues/19, and the mouse protein embedding is available at http://snap.stanford.edu/saturn/data/protein_embeddings.tar.gz. SATURN uses a combination of cell type annotations from the user and the protein embeddings to create ‘‘macrogenes’’, groups of genes that are predicted to be ‘‘functionally related’’ and ‘‘coexpressed across species’’[73]. It then uses a weakly supervised autoencoder to refine the macrogene space by using a triplet loss function that incorporates within-species cell type annotations. Before running SATURN, we first randomly downsampled the annotated cell × gene matrices for both species to 100,000 nuclei total to reduce computational burden. We provided SATURN with a mapping between mouse and marmoset supercluster names, which were identical except for 2-3 unique superclusters per species. We trained the SATURN model as instructed in the tutorial (https://github.com/snap-stanford/SATURN/blob/main/Vignettes/frog_zebrafish_embryogenesis/Train%20SATURN.ipynb), using raw (not normalized) counts, 2,000 macrogenes, 8,000 highly-variable genes, and our supercluster mapping as the cell type mapping file. The resulting SATURN embedding in UMAP space was used to generate the plots in Figure S17.

We performed SATURN cross-species integration on the same downsampled, supercluster-annotated cell × gene expression matrix as for our scANVI approach (with the exception of L5IT cortical excitatory neurons labeled separately for marmoset). By visual inspection, the species-integrated UMAP calculated from the SATURN embedding yielded similar results as our scANVI approach: broad conservation with a few species-specific clusters as described above (Figure S17C). However, because SATURN does not explicitly rely on cell type annotation to group cells from different species, it was able to merge cortical glutamatergic L5IT supercluster mouse neurons with the corresponding population in marmoset (Figure S17C, dark green cluster with the asterisked label), despite the fact that L5IT mouse neurons were not separately annotated as such, but grouped with L4/5IT cortical glutamatergic neurons. The scANVI approach did not result in this merging unless we grouped L5IT and L4/5IT neuron superclusters in both species, which is the approach we adopted for the results shown in Figure 6. However, the relative position of superclusters in the SATURN UMAP seems to be less meaningful than for the scVI/scANVI embeddings: related neuronal clusters are no longer adjacent in UMAP space, OPCs and MOLs are far away from one another, and microglia were well integrated, despite species differences apparent in the scANVI integration and past studies suggesting they should be species-divergent.[76,130] Closer examination of the SATURN-integrated astrocytes illustrates concordant results with the scANVI integration: telencephalic marmoset and mouse astrocytes were well-integrated, diencephalic astrocytes were more separated, and immature astrocytes were almost completely separated (Figure S17D).

Cell-cell communication analysis

We performed cell-cell communication analysis on a per-region (prefrontal and motor cortex pooled), per-age, per-species basis in Python using CellPhoneDB[64] via Liana,[109] a scanpy-friendly package that integrates several cell-cell interaction inference methods. To avoid spurious findings due to differences in cluster proportion (i.e., bias towards or away from rare cell types), we first randomly downsampled all clusters to a maximum of 1,000 nuclei and dropped all clusters with fewer than 100 (thalamus and cortex) or 70 (striatum, to avoid dropping the sparse *CHAT*+ neuron cluster) nuclei. As recommended in the Liana documentation, we ran CellPhone DB on the log1p-transformed normalized counts matrix, grouping by leiden cluster, using the ‘‘consensus’’ resource for marmoset and the ‘‘mouse consensus’’ resource for mouse, and requiring that 33% of nuclei in a cluster express a gene at nonzero levels for it to be considered for ligand-receptor enrichment. CellPhoneDB generated two outputs of interest: 1) ‘‘lr_means’’, a measure of interaction magnitude, and is simply the mean expression of the ligand in the source cluster averaged with the mean expression of the receptor in the target cluster, and 2) the permutation-based p value, a measure of interaction specificity. Briefly, the specificity p value for a ligand-receptor pair is calculated as the proportion of null distribution (generated by randomly permuting the cluster labels of all cells) means that are greater than or equal to the actual mean expression calculated. For all ligand-receptor pairs shown, we required a CellPhoneDB p-value of less than 10^{-6}. Because many ligands and receptors are expressed by both neurons and astrocytes, we ran CellPhoneDB on only the neuronal and astrocytic clusters to increase specificity by removing cell types that could deflate the specificity p-value. To generate the dotplots shown in Figures 4A, 4B, S14A, and S14B, we restricted the plot to only those ligand-receptor (L-R) pairs that were near-unique (p-value below 10^{-6}) to 2 or fewer (except for mouse P90, which was relaxed to 3 or fewer) neuronal clusters and the dominant astrocyte subtype (or vice-versa). Otherwise, the top 25 L-R pairs shown are mostly shared across all neuronal subtypes.

For the upset plots shown in Figures 4C, 4D, S14C, and S14D, non-filtered (non near-unique) L-R pairs were used. Because immature versions of the most abundant neuronal and astrocytic cluster dominated at early developmental time points in both species, L-R pairs between these immature clusters were used at these time points only, and only if their abundance was greater than the corresponding mature cluster. For example, in neonate mouse, we used L-R pairs between the 'Neuron_Excit_Ctx_Immature_L23IT_Ptprk' Leiden cluster and the 'Astrocyte_Immature_Ptprz1' Leiden cluster, while in adult mouse, we used L-R pairs between 'Neuron_Excit_Ctx_L23IT_Cam2ka' and 'Astrocyte_Telen_Mature_Slc1a2'. Details are available in the Jupyter notebooks in our GitHub repository (https://github.com/Feng-Lab-MIT/AstrocyteHeterogeneity).

We ran three analyses at a single developmental time point (14-month marmoset) to check whether differences in neuron-astrocyte and astrocyte-neuron L-R pair expression across regions were driven solely by neuronal expression differences across regions. We first ran a region-scrambled CellPhoneDB analysis that included local neuronal and astrocytic populations from all regions in the same run of CellPhoneDB (as opposed to running CellPhoneDB separately by region as in the manuscript Figures 4 and S14). See ''CCC_check_specificity_toneuron_astrocyte_shuffle_nofilter_allnas_current.ipynb'' in our GitHub repository for more details. We compared the magnitude and specificity of N-A L-R pairs across regions by examining L-R pairs between local interneuron clusters (cortical MGE-derived *PVALB*+ interneurons, striatal MGE-derived interneurons, and thalamic midbrain-derived interneurons) and local astrocyte populations (cortical telencephalic astrocytes, striatal telencephalic astrocytes, and thalamic diencephalic astrocytes). Our results showed that neuron-astrocyte L-R enrichment patterns are not the same across all astrocyte subtypes for a given local neuronal subtype. In a second analysis, we compared the current astrocyte-neuron/neuron-astrocyte L-R pairs, and their sharing/divergence across regions, to OPC-neuron/neuron-OPC L-R pairs. See ''CCC_check_specificity_toneuron_opc_shuffle_nofilter_allnos_current.ipynb'' in our GitHub repository for more details. We found that only 40-60% of L-R pairs overlapped between astrocytes and OPCs within a given region. Finally, we quantified the proportion of genes in L-R pairs of a given neuron-astrocyte pairing that are also DEGs in that neuron cluster, or astrocyte regional subtype, relative to others. Our analysis showed that N-A L-R pairs contained DEGs from both neurons and astrocytes, with the proportion of ligands or receptors that were also cluster DEGs ranging from ∼20-70%. In fact, astrocytes had a much larger proportion of overlap between ligand (when source) or receptor (when target) and leiden-level cluster DEGs. However, the increased astrocyte DEG presence in L-R pairs should be interpreted with caution, because there are many more neuronal than astrocytic Leiden clusters, potentially leading to more general ''astrocytic'' genes being DEGs, while neuronal DEGs are more leiden cluster-specific.

Pseudotime analysis and calculation of gene change scores

We performed pseudotime analysis using Palantir,[67] a scanpy-friendly package that orders cells along pseudo-temporal trajectories based on diffusion space and assigns each cell a probability of differentiating into each user-defined terminal state based on a Markov chain. Importantly, unlike single-cell velocity based methods that rely on estimates of spliced and unspliced counts,[131],[132] Palantir runs diffusion maps in a latent space calculated from the regular counts matrix, which is in our case the scVI latent space (pca_key= ''X_scVI'', n_components=5). The ''determine_multiscale_space'' parameter n_eigs was set to 5 to avoid a documented error (https://github.com/dpeerlab/Palantir/issues/84). Palantir imputes missing gene expression data in log1p-transformed counts per million space using MAGIC,[124] which is useful for visualizing gene expression. The root and terminal cells were manually specified for each cell type (oligodendrocyte lineage or astrocyte) based on their location in UMAP space combined with cluster and age information, and 500 waypoints were used. We calculated pseudotime trajectories for the astrocyte and oligodendrocyte lineages on a per-species region-combined basis. To reduce computational load, marmoset oligodendrocytes were randomly downsampled to 100,000 nuclei. We first ran Palantir pseudotime analysis on the oligodendrocyte lineage as a ''sense check'' for the algorithm, as the biological ground truth of oligodendrocyte lineage differentiation is well known (OPC→COP→NFOL→MFOL→MOL[68]), and found that Palantir's calculated pseudotime was able to recapitulate the known differentiation trajectory.

We then used Mellon[69] to identify genes whose expression changes significantly in pseudotime transition periods, identified as regions of low cell-state density, for each pseudotime trajectory branch (AST-DI and AST-TE). We followed the basic Mellon tutorial ''Density Estimator for scRNA-seq data'' (https://mellon.readthedocs.io/en/latest/notebooks/basic_tutorial.html) and the ''Gene change analysis'' tutorial (https://mellon.readthedocs.io/en/latest/notebooks/gene_change_analysis_tutorial.html). Mellon, a companion algorithm to Palantir, estimates cell-state ''densities'', that is, density of cells in each state, from high-dimensional single cell data using a Bayesian model. It also computes ''local variability'' on a per-cell, per-gene basis as the maximum (across nearest neighbors) normalized (by the distance between cells in state space) difference in MAGIC-imputed expression of a gene in a cell compared to that of its nearest neighbors. Gene change scores are then calculated using a specified set of cells (e.g., a pseudotime trajectory branch) as the inverse density-weighted (such that lower densities have higher weights) average local variability for a gene across all cells. As stated in the source paper, ''gene-change analysis quantifies the influence of a gene in driving state transitions in low-density regions''. Briefly, we subsetted the integrated all-region, all-age data to astrocytes, calculated the diffusion maps, calculated densities, calculated pseudotime using Palantir (as described above), computed local variability, and calculated gene change scores separately for each pseudotime branch (AST-TE and AST-DI). We then computed gene expression trends from the MAGIC-imputed expression data and visualized the trends over pseudotime for the top 25 highest change-scoring genes (Figure S15C). We also plotted the average trend overlaid on the per-cell trend scatterplot vs. pseudotime for each of these 25 genes, and selected 1 example gene from each trajectory in Figure S15D. These plots are available in our GitHub repository for marmoset in the ‘‘marmoset_mellon_gene_change_allages_regcombined_current.ipynb’’ notebook and for mouse in the ‘‘mouse_mellon_gene_change_allages_regcombined_current.ipynb’’ notebook. A list of the highest change-scoring genes is provided for both species in Table S18. We gratefully acknowledge the assistance of Dominik Otto on this analysis (https://github.com/settylab/Mellon/issues/13).

FISH quantification with CellProfiler

Quantification of RNAscope HiPlex FISH images was preregistered during or after data collection on the Open Science Framework (https://osf.io/crs7v/; https://doi.org/10.17605/OSF.IO/6KUB2) and conducted with the analyzer blinded to rDEG identity, for mouse (for all genes except *Sparc*, for which we modified the analysis pipeline post-hoc, see below). The following text has been adapted from the preregistration, with major deviations from the preregistration being noted here. CellProfiler 4.2.5 was used to quantify RNA signal in each nuclei of different regions. Variations between experiments such as tissue quality or probe freshness caused variations in fluorescence, so it was necessary to use separate CellProfiler pipelines for each experiment for mouse (with the exception of one adult mouse experiment that needed to be split by replicate) and neonate marmoset to account for those differences. The adult marmoset tissue had more dramatic difference in quality between replicates, requiring a separate pipeline for each replicate. Crucially, the CellProfiler pipelines were always identical within each slice of tissue, allowing for each replicate to have unbiased comparison between regions.

We first created binary masks to eliminate large artifacts (e.g., very large/bright debris), large tears in the tissue, areas of very high autofluorescence, out-of-focus areas, poorly registered areas, and/or nearby brain regions that were not the region of interest. In most cases, the cropped region of interest was greater than ~80% of the original. For adult marmoset tissue, which exhibited persistent lipofuscin autofluorescence in later rounds despite quenching with TrueBlack Plus, we automatically generated masks to eliminate this punctate fluorescence as follows: minimum intensity projection of all three channels in the third round, gaussian filter with sigma = 2, intensity thresholding at 1.5 standard deviations above the mean, and median filtering with radius of 2 pixels. We then combined this autofluorescence mask with the large artifact mask mentioned above. This quality control step was not mentioned in our preregistration, largely because we did not anticipate the extent of remaining autofluorescence. The registered, cropped, masked max-projected images were processed through the CellProfiler quantification pipeline.

In CellProfiler, all images were scaled to stretch to the full intensity range and masked to regions of interest. The lipofuscin mask was used to filter nuclei covered more than 20% by lipofuscin and crop any smaller areas within remaining nuclei. Some regions had a higher background intensity than others, so the images were binarized to eliminate background noise. Binarization was achieved using the CellProfiler ‘‘Threshold’’ module, with the threshold set manually after using the Otsu algorithm to suggest options for values that would separate foreground and background. The manual threshold was used instead of Otsu, so that the definition of foreground and background would be consistent between regions. The exception to this was *Slc1a3*, our astrocyte marker, because its signal was known to be different between regions. Using this binarized signal, we were able to quantify the fluorescence from RNA in each nucleus. There were two rDEG probes in the adult mouse that required additional processing steps prior to binarization: *Sparc* and *Clmn* were particularly noisy in the thalamus specifically, with punctate signal outside of nuclei. Those genes required the ‘‘EnhanceOrSuppressFeatures’’ to suppress speckles. We also applied the ‘‘Smooth’’ module to *Sparc* in all adult mouse regions to reduce noise. From the binarized signal, we calculated mean intensity, or the fraction of the nucleus that is covered by the probe signal, and integrated intensity, the number of thresholded pixels within the area of the nucleus.

All nucleus objects identified by DAPI signal were expanded by 3 pixels to reflect RNA signal likely also existing outside of the nucleus, with the exception of P4 mouse nuclei, as they were more tightly packed together. To identify astrocytes, we first filtered by *SLC1A3*/*Slc1a3* mean intensity, using 5% for neonate mouse astrocytes, 6.5% for adult mouse astrocytes, 12.5% for neonate marmoset astrocytes, and 8% for adult marmoset astrocytes. This value was determined to reflect expected expression levels and similar fraction of astrocytes out of total cells based on the snRNA-seq data. Next, we filtered out the astrocytes that had at least 50% mean intensity for neuronal markers *GAD2*/*Gad2*, *SLC17A6*/*Slc17a6*, and *SLC17A7*/*Slc17a7*. We did not filter based on *OLIG2*/*Olig2* because a subset of astrocytes are known to express *OLIG2*/*Olig2*.[133] We then measured the intensity of the binarized signal in our rDEG probes inside astrocytes and in all nuclei. To count a cell positive for a probe, the cell needed to either have at least 0.03 mean intensity or 6 integrated intensity. Both measurements were used to account for variations in cell size, as smaller cells would likely have lower integrated intensity and larger cells would likely have lower mean intensity. To compare rDEG expression between regions, we used the fraction of cells counted positive for a probe and the average mean intensity of all cells within the region. For mouse, these values were averaged within slices of the same replicate and then treated as a single data point.

We used DAPI signal, positive control probes, and cell type marker probe signals to qualitatively assess RNA quality (e.g., degraded or not degraded based on the brightness of fluorescence intensity) before running images through the analysis pipeline. If we observed unacceptably low tissue (e.g. over-digested, damaged, or folded slices or extremely high background) or RNA quality (little to no signal) for a given sample, we collected more images from a different slice or animal to maintain sample size prior to starting analysis. Given the low number of samples per region/age/species, we did not remove outliers, as we were not adequately powered to detect outliers post-hoc. Of note, changes were made to marmoset analysis pipelines after unblinding due to unrealistic detection of *FOXG1* expression in the thalamus. This gene is known to be a telencephalic patterning factor,[134] a finding confirmed *in situ* by other groups (see for example the RIKEN Marmoset Gene Atlas[135,136] at https://gene-atlas.brainminds.jp/gene-image/?gene=370-6), triggering reevaluation of the lipofuscin masks and *FOXG1*'s binarization threshold. Ultimately, we excluded *FOXG1* from our HiPlex quantification results because no version of our pipeline could overcome the large variations in signal-to-noise between brain regions.

Another notable change to the pipeline after unblinding was an increase of the intensity threshold for binarizing *Sparc* images in neonate mouse, as the minimal expression of the gene at the neonate timepoint led to an erroneously low intensity threshold under blinding. The *Sparc* mouse neonate analysis was also rerun with altered masks to remove the pial and ventricle surfaces of the PFC and striatum. At P4 in mouse, we noticed that *Sparc* expression was particularly high in endothelial cells, and what look to be pial astrocytes or other pial-associated non-neuronal cells lining the edges of the PFC, as well as in ependymal cells bordering the striatum (Figure S10E). We suspected that co-localization of *Slc1a3* (which is not exclusive to astrocytes) and *Sparc* in these brain border- and vascular-associated cells was driving a discrepancy between the snRNA-seq data and our FISH image quantification. We verified this by repeating *Sparc* quantification using a mask that removed the thin border sections (pial surface, ventricle) of the tissue in the PFC and striatum and increased intensity threshold in all regions, and found expression trends that more closely align with our snRNA-seq predictions.

For the RNAscope v2 data, the images did not require advanced masking to account for lipofuscin and were only masked for the regions of interest. Because there was no labeling for other cell type markers to filter the cells, the threshold of mean *SLC1A3* intensity for astrocytes was increased to 15% and the fraction of astrocytes per region was comparable to the HiPlex quantification. Because of differing signal-to-noise ratios of the RNAscope protocols, the threshold for a positive cell was increased to at least 5% mean intensity or 10 integrated intensity.

Because we did not have sufficient sample size to test for statistical significance in marmoset, we report only observations and trends. For mouse data, we used 2 univariate ANOVAs with Benjamini–Hochberg p value correction for each measure presented: fraction of astrocytes positive for a probe and mean fraction of the astrocyte covered by the probe (i.e., mean intensity). If the effect of brain region was significant overall, Tukey's multiple comparisons (i.e., the Tukey HSD method) test was performed separately on each variable to test for the significance of pairwise differences between brain regions. Statistical analyses were performed in GraphPad Prism. Contrary to our preregistration, we did not first use a one-way repeated measures *multivariate* ANOVA, as we could not find a package to do so, but we believe this divergence to be minor given that we only report two outcome variables.

Not all marmoset astrocyte rDEGs followed the predicted expression pattern by RNAscope HiPlex FISH, and that there is a high degree of variability between the two replicates (Figure S9). For example, *FOXG1* expression was spuriously detected in adult marmoset thalamus (at lower levels than PFC and striatum) by our CellProfiler pipeline, perhaps due to low and variable signal-to-noise. Nevertheless, *FOXG1* expression appears negligible in thalamic astrocytes when compared to PFC and striatum in our higher-quality marmoset sample, highlighting the limitations of a thresholding-based quantification approach (Figure S8). In another example, the fraction of astrocytes positive for *KCND2* differs by several fold in the PFC between the two adult marmoset replicates. Our inability to quantitatively validate rDEG patterns in some replicates is likely due to variable tissue quality (even between regions, with striatum usually exhibiting less autofluorescence than PFC or thalamus), sampling a different position on the medial/lateral axis (especially for neonate marmoset samples), decreased signal-to-noise in later rounds of the RNAscope HiPlex protocol, overlapping RNA signal from adjacent cells of other cell types, or a combination of these factors. We found the RNAscope HiPlex protocol to produce variable results in marmoset tissue, especially in later rounds, even after several attempts at optimization. For this reason, we probed for *SPARC* and *KCNH7*, two probes with lower signal-to-noise in the HiPlex protocol, using RNAscope Multiplex Fluorescent v2 protocol, and produced much higher quality data. We suspect that the signal-to-noise for other rDEGs would be similarly improved, and some of the aforementioned issues resolved, if we used the RNAscope v2 protocol for all rDEGs though we did not perform these experiments ourselves.

In mouse, our quantification suggested *Gfap* expression was higher at both time points than predicted by snRNA-seq, as was *Sparc* expression in neonates (Figure S11D), findings which were not readily apparent by visual inspection of the images (Figure S10). For *Gfap*, this discrepancy may reflect under-detection in the snRNA-seq dataset, as evidenced by higher *Gfap* expression in the mouse ABCA MERSCOPE v1 dataset (Figure S11B). For *Sparc*, this may reflect low overall expression, which reduces the signal-to-noise ratio of FISH and makes our quantification methods more vulnerable to false positives. In this and other cases, discrepancies between snRNA-seq expression and FISH signal may be due to distinct gene capture strategies, or the use of probe target sequences on exons not highly expressed in astrocytes.

Quantification of mouse astrocyte rDEG expression in the ABCA Whole Mouse Brain atlas

We utilized the Allen Brain Cell Atlas MERFISH spatial transcriptomics dataset[1] available at https://knowledge.brain-map.org/abcatlas and using the tutorial posted at https://alleninstitute.github.io/abc_atlas_access/descriptions/MERFISH-C57BL6J-638850.html. Next, we subsetted the data to astrocytes only by subclass, only '318 Astro-NT NN' or '319 Astro-TE NN'. To subset the data for the regions of interest, we selected striatum, thalamus, hypothalamus, midbrain, and cerebellum cells by the parcellation divisions 'STR', 'TH', 'HY', 'MB', or 'CB' respectively.

We selected PFC cells by parcellation structure being 'AId', 'PL', 'ORBl', 'AIp', 'ORBvl', 'ORBm', 'AIv', 'ILA', or 'FRP'. We selected motor cortex cells by parcellation structure 'MOp', or 'MOs.' Finally, we selected somatosensory cortex cells by parcellation structures of ‘SS-bfd’, ‘SSp-ll’, ‘SSp-m’, ‘SSp-n’, ‘SSp-tr’, ‘SSp-ul’, ‘SSp-un’,’ or ‘SSs.’ We then averaged the gene expression for each gene of interest per region and plotted a heatmap of the resulting values.

Analysis of low-expansion factor ExR morphology images

Parameters for the following analyses were determined in advance of results compilation and statistical testing. All ExR images were background-subtracted in Fiji using the rolling ball algorithm, radius of 50 pixels, as in our previous studies.[37,117] To quantify astrocyte morphology from ~3.5× expanded astrocytes, images were processed as follows after background subtraction: conversion to grayscale, gaussian filtering with sigma = 2 using MATLAB’s “imgaussfilt3”, binarization with an intensity threshold at 1 standard deviation above the mean across the whole stack, and size filtering with a minimum of 10^6 voxels using MATLAB’s “bwareaopen”. Importantly, the resulting segmentations allowed our morphological profiling to be robust to variations in GFP intensity observed in the astrocytes we imaged (Figure 7C; Videos S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, and S36). If no astrocytes were detected after thresholding and filtering, the intensity threshold was lowered to 33.33% of the original (i.e., 0.3333 standard deviations above the mean intensity of the whole stack) and size filtered at the same minimum size. If no astrocytes were detected after lowering the intensity threshold, no segmentation was created for the image and it was excluded from further analysis. If there were two connected components (MATLAB’s “bwconncomp”, connectivity of 26) over the minimum size filter, a new size filter was set at 10 voxels less than the volume of the largest connected component, so that only the largest astrocyte remained. Volume, surface area, and equivalent diameter were then calculated from the largest connected component using MATLAB’s “regionprops3”, and converted to cubic and square microns using a weighted average of the physical pixel size divided by the expansion factor (e.g., $((^2/_3)*(0.1625/\text{expansion\_factor}) + (^1/_3)*(0.5/\text{expansion\_factor}))^3$ for volume). Aspect ratio was calculated as the length of the first principal axis divided by the length of the second principal axis of the ellipsoid with the equivalent normalized second central moments as the connected component, calculated using MATLAB’s “regionprops3”.

After segmentation, we excluded 7 astrocytes that had a large-volume second astrocyte in the 3D binary segmentation, and/or had low-quality 3D segmentations due to sparse signal (for 2 of such astrocytes, no 3D binary segmentation was created). Notes on the quality of each imaged astrocyte region of interest are provided in Table S32. We assessed whether the results of our statistical testing (see below) held up when more stringent quality control quality control (QC) criteria were applied to the 3D astrocyte segmentations being analyzed. Specifically, we removed astrocytes with sparse segmentations (6), segmentations including portions of second astrocytes in the field of view (22), cracked blood vessels (27), or which were partially cut off in the z-dimension (63). This higher-quality subset of the data included 74 astrocytes, most of which came from the thalamus (38%), followed by striatum (36%), and lastly PFC (26%). We repeated our linear mixed effects modeling on this dataset, and report the results in Figure S18A and Table S25).

Sholl analysis was run using Fiji’s SNT[137] Sholl Analysis[80] plugin, installed and accessed via the Neuroanatomy plugin, as described: https://imagej.net/plugins/snt/#installation. The center of each soma was manually annotated using the point selection tool and saved as an overlay. Sholl analysis was run in batch mode for each image using a custom macro that calls the legacy version of SNT’s Sholl Analysis, with the following parameters: start radius of 2 μm, end radius of 60 μm, step size 2 μm, and no polynomial fitting (see “batch_sholl_frombin.ijm” in our GitHub repository). The number of intersections at each radius for each cell was averaged over all non-excluded astrocytes in each region to create the plots in Figure 7Evi.

Analysis of fractal dimension was performed in MATLAB on 3D binary segmentations using the box counting method via “boxcount”, available on the MATLAB file exchange at https://www.mathworks.com/matlabcentral/fileexchange/13063-boxcount (written by Frederick Moisy, accessed August 2024). Briefly, a fractal set (in this case, a binary image of an astrocyte) is one that exhibits self-similarity at progressively smaller scales. The box counting method can be used to characterize the extent to which a set is fractal by counting the number of boxes of size R in a grid needed to cover the edges (in our case, transition from black to white) of an image at progressively smaller grid sizes.[79] A fractal image will have exponentially more detail (edges) at smaller grid sizes, require exponentially more boxes of size R to cover the set, and therefore will have a negatively sloped line on the log-log plot of N, the number of boxes vs. R, the size of each box in the grid. For a straightforward explanation of box counting, see https://fractalfoundation.org/OFC/OFC-10-5.html. The local fractal dimension can therefore be calculated as the slope of the log(N) vs. log(R) plot (df = -diff (log(n))./diff(log(r)), as in the boxcount function “Examples” tab). We took the mean fractal dimension over all 12 local slopes on the line to arrive at a single measure of fractal dimension for each astrocyte.

Statistical significance of differences between astrocytes from different brain regions was determined using a linear mixed effects model via the scipy statsmodels package’s “mixedlm” function (https://www.statsmodels.org/stable/mixed_linear.html), where the outcome variable was the measure of interest, the coefficient was region assignment, and “animal” was the random effect group variable (see the associated Jupyter notebook, “analyze_ExR_4x_astromorph_results.ipynb” on our GitHub repository). *p* values on the coefficients from striatum and thalamus were corrected for multiple comparisons with the Benjamini-Hochberg method using the scipy stats package’s “false_discovery_control” function (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.false_discovery_control.html). For the analysis where more stringent QC criteria were applied, the model failed to converge for fractal dimension. Results are provided in Table S25.

Analysis of high-expansion factor ExR rDEG target images

All ExR images were background-subtracted in Fiji using the rolling ball algorithm, radius of 50 pixels, as in our previous studies.[37,117] To create a segmentation of astrocyte processes (and if applicable, soma), the GFP channel was processed as follows: gaussian filtering with sigma = 10 using MATLAB's "imgaussfilt3", binarization with an intensity threshold at 0.75 standard deviations above the mean across the whole stack, median filtering with size 5 × 5 × 3 voxels, and size filtering with a minimum of 200 voxels using MATLAB's "bwareaopen". We then calculated the mean intensity (Figure 7F) of the target channel (either Glast or Gat3) in the GFP+ region as follows: divided the raw pixel value of the target image by 2e16 (images were 16-bit), multiplied the resulting target image by the binary segmentation of the GFP channel, and divided the sum of this image by the number of nonzero pixels. We did the same for the reference synaptic protein channel, Cav2.1, which can be used to normalize the intensity values for the target channel (Figure S18Bi). The enrichment ratio was calculated as the mean intensity (of either the target channel or Cav2.1) within the GFP+ segmentation to the mean intensity outside the GFP+ region (Figure S18Bii). Fields of view with poor Cav2.1 staining or lots of empty gel were excluded from analysis (see Table S32). Results are provided in Table S26.

Generative AI

ChatGPT (OpenAI, GPT-3.5 or 4.0) and GitHub Copilot (in Visual Studio Code) were used to aid computational scripting and debugging. On these occasions, we prompted it to interpret error messages and some lines of code generated by others, generate bash commands and scripts, accelerate existing Python code, generate some Python functions and code chunks for niche tasks (e.g., generating Venn diagrams and raster plots), and/or used the Copilot autocompletion function. ChatGPT and CoPilot-generated code blocks are annotated as such in our Python notebooks. However, generative AI did not meaningfully contribute to the intellectual development of experimental design, data analysis, or data interpretation of this manuscript. No generative AI was used to write or edit the manuscript.

ADDITIONAL RESOURCES

Raw data (sequencing reads in fastq format) for all 10× Chromium snRNA-seq samples and CellBender-cleaned aligned counts matrices (in .h5 format) genrerated in this study are publicly available for download on the Neuroscience Multi-omic Data Archive (NeMO) at: https://data.nemoarchive.org/biccn/grant/u01_feng/feng/transcriptome/sncell/10x_v3.1/ and on the Gene Expression Omnibus (GEO) at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE309702. Pre-processed, clustered, and annotated data (in.h5ad format) is available for download, exploration, and gene search on the Broad Single Cell Portal at: https://singlecell.broadinstitute.org/single_cell/study/SCP2719/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-mouse-brain (mouse) and https://singlecell.broadinstitute.org/single_cell/study/SCP2706/a-multi-region-transcriptomic-atlas-of-developmental-cell-type-diversity-in-marmoset-brain (marmoset). Registered RNAscope HiPlex/v2 FISH image stacks, raw and background-subtracted expansion microscopy image volumes and binarized 3D tracings (both in.tif format) are available for download on BossDB at https://bossdb.org/project/schroeder2025. Custom scripts used to analyze data are available at https://github.com/Feng-Lab-MIT/AstrocyteHeterogeneity.

Neuron, Volume *113*

Supplemental information

A transcriptomic atlas of astrocyte heterogeneity across space and time in mouse and marmoset

Margaret E. Schroeder, Dana M. McCormack, Lukas R. Metzner, Jinyoung Kang, Katelyn X. Li, Eunah Yu, Lisa Melamed, Kirsten M. Levandowski, Heather Zaniewski, Qiangge Zhang, Edward S. Boyden, Fenna M. Krienen, and Guoping Feng

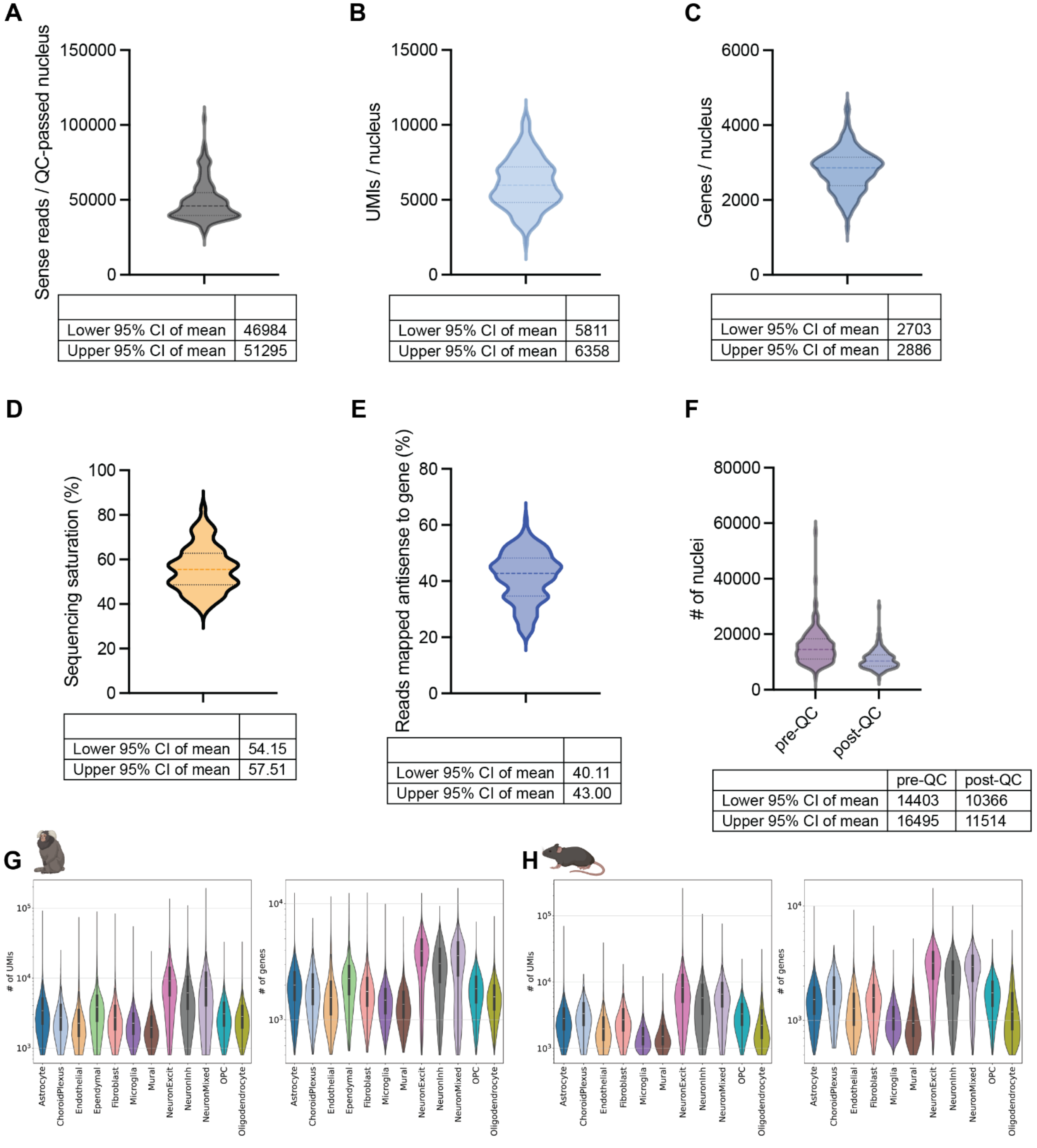
Supplementary Information

A transcriptomic atlas of astrocyte heterogeneity across space and time in mouse and marmoset

Margaret E. Schroeder, Dana M. McCormack, Lukas R. Metzner, Jinyoung Kang, Katelyn X. Li, Eunah Yu, Lisa Melamed, Kirsten M. Levandowski, Heather Zaniewski, Qiangge Zhang, Edward S. Boyden, Fenna M. Krienen, and Guoping Feng

Supplementary Information Contents

Supplementary Figures



A

Lower 95% CI of mean	46984
Upper 95% CI of mean	51295

B

Lower 95% CI of mean	5811
Upper 95% CI of mean	6358

C

Lower 95% CI of mean	2703
Upper 95% CI of mean	2886

D

Lower 95% CI of mean	54.15
Upper 95% CI of mean	57.51

E

Lower 95% CI of mean	40.11
Upper 95% CI of mean	43.00

F

	pre-QC	post-QC
Lower 95% CI of mean	14403	10366
Upper 95% CI of mean	16495	11514

Figure S1. Summary of sequencing coverage for the cross-region developmental snRNAseq atlas, related to Fig. 1. A, Violin plot showing the estimated population distribution of sense reads per rough quality control (QC)-passed (low-UMI and low-gene but not doublet removal, see **Methods**) for all nuclei across all 144 10x Chromium reactions sequenced in the current study (not including adult marmoset data from our previous study[1]). Dashed lines indicate the median and quartiles. **B**, Violin plot (as described for **(A)**) of the median number of unique molecular identifiers (UMIs) per nucleus as described in **(A)** (n = 144 10x Chromium reactions). **C,** Same as **(B)**, for the median number of genes per nucleus. **D**, Same as **(C)**, for sequencing saturation as calculated by 10x Cell Ranger software (v7.1). **E**, Same as **(D)**, for the fraction of reads mapped antisense to genes, as calculated by 10x Cell Ranger software (v7.1). **F**, Number of nuclei called by 10x Cell Ranger software before (pre-QC) and after (post-QC) rough QC (low-UMI and low-gene but not doublet removal, see **Methods**). **G,** Violin plot (scanpy's default) showing the estimated distribution of the number of UMIs (left) and genes (right) for full QC-passed marmoset nuclei, as presented in the main figures, grouped by cell type. **H**, Same as **(G)**, for mouse nuclei.

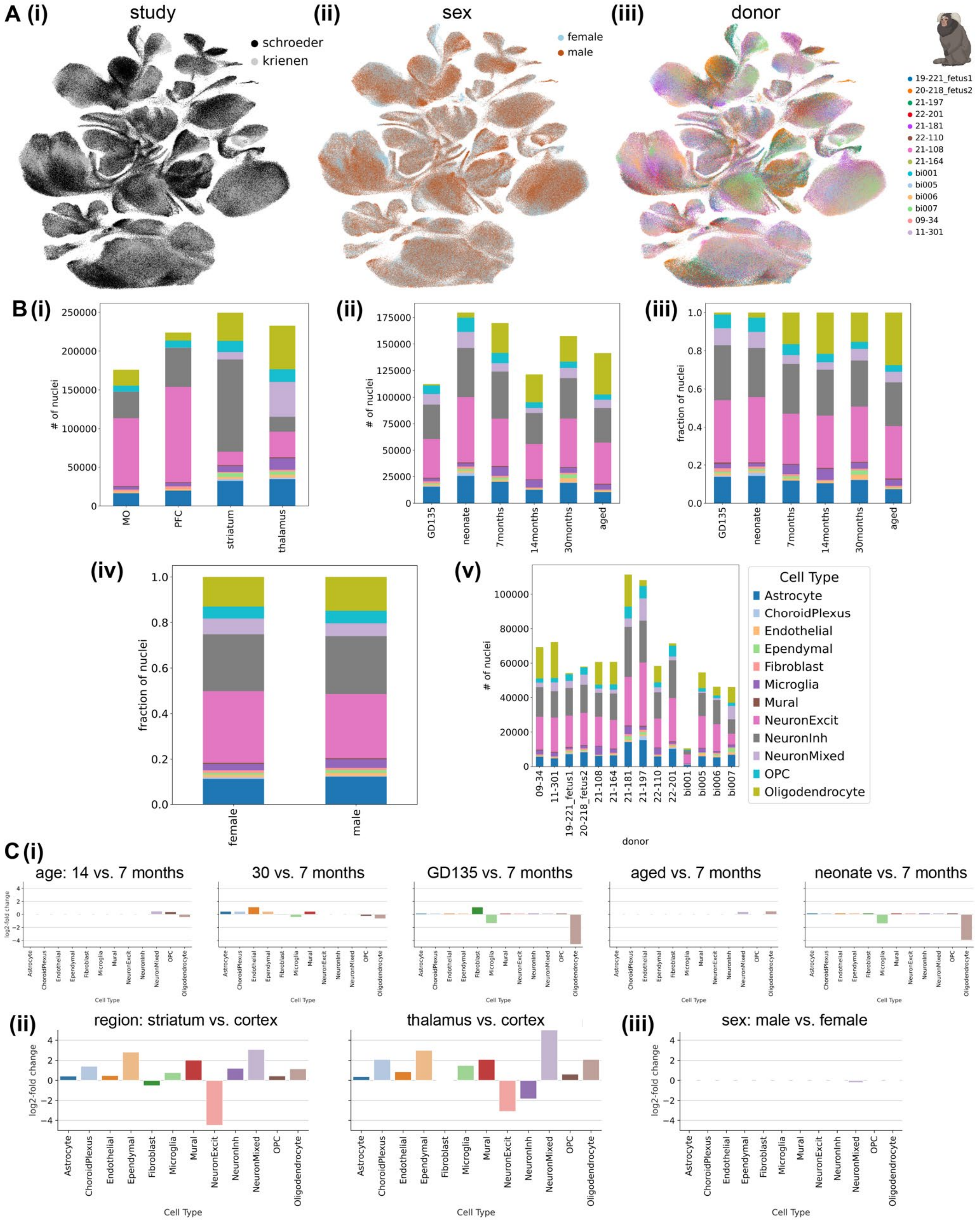


Figure S2. Distribution of nuclei across cell types, dissected regions, ages, sexes, and study for marmoset single-nucleus RNAseq data, related to Fig. 1. A, Integrated UMAP embedding of 881,832 marmoset nuclei from PFC, motor cortex, striatum, and thalamus (as dissected, before reassignment) across all developmental time points assayed and a randomly downsampled portion of adult nuclei from our previous study[1] colored by **(i)** study (present, “schroeder” or previous, “krienen”), (**ii)** sex, and **(iii)** biological replicate (for marmoset equivalent to single donor). **B**, Stacked bar plots showing the number or proportion of full QC-passed nuclei for each **(i)** dissected brain region, **(ii)** age (number of nuclei), **(iii)** age (proportion of nuclei) **(iv)** sex (proportion of nuclei) and **(v)** biological replicate, broken down by assigned cell type. **C**, Log2-fold change in cell type composition as estimated by single-cell compositional data analysis (scCODA)[2] for the coefficients on **(i)** age (reference = 14 months), **(ii)** region (reference = cortex), and **(iii)** male sex (reference = female).Only fold-changes for credible effects (i.e., significantly nonzero coefficient) are shown. The 881,832 marmoset nuclei were composed of 12 broad cell classes: glutamatergic excitatory neurons (29.7%, GABAergic inhibitory neurons (25.2%), neurons belonging to mixed glutamatergic and GABAergic clusters (6.24%), astrocytes (11.7%), oligodendrocytes (14.0%), oligodendrocyte progenitor cells (OPCs, 5.35%), microglia and macrophages (3.45%), mural cells (0.51%), endothelial cells (1.22%), vascular-associated fibroblasts (1.05%), ependymal cells (1.11% in), and choroid plexus cells (0.45%).

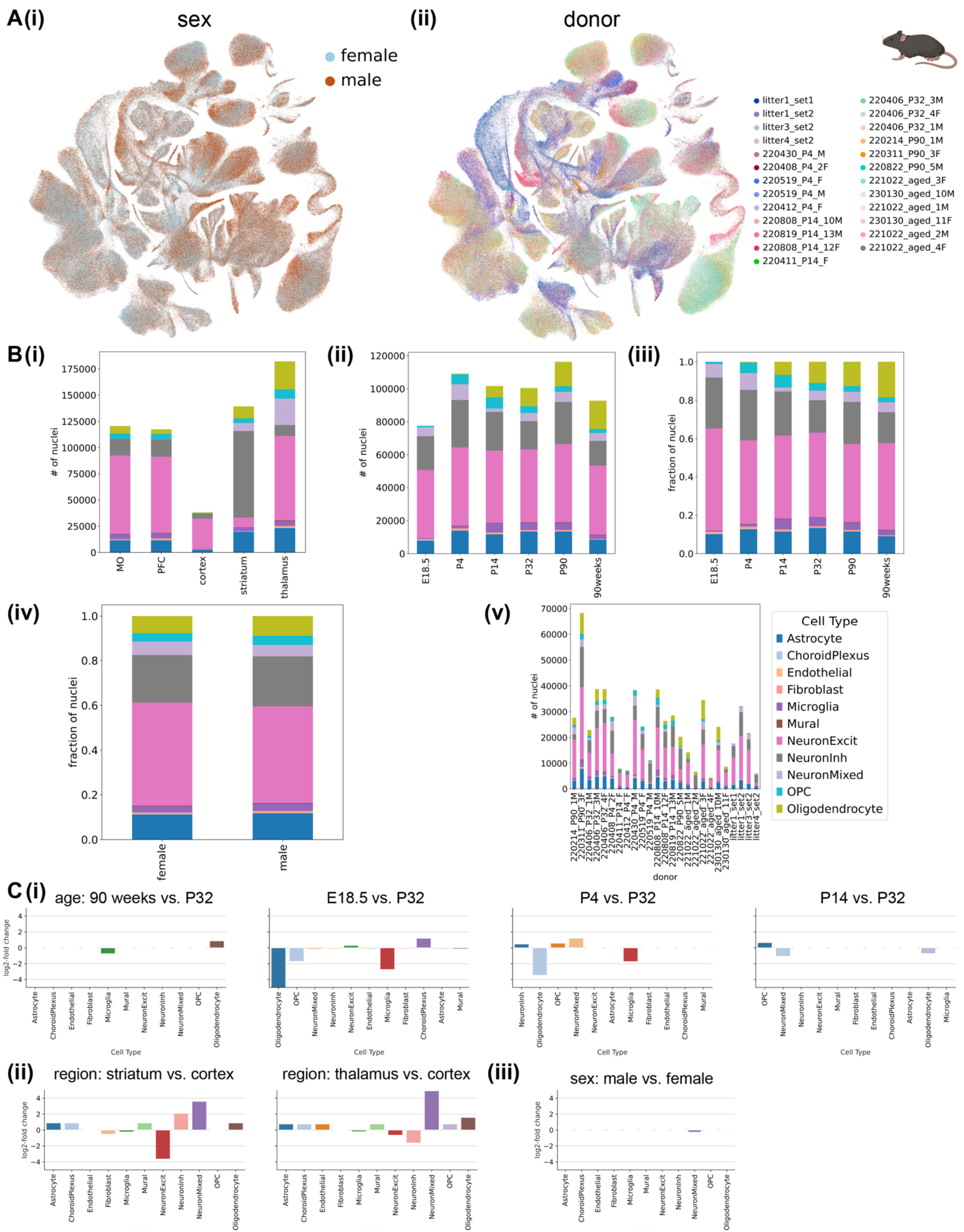


Figure S3. Distribution of nuclei across cell types, regions, ages, sexes, and study for mouse single-nucleus RNAseq data, related to Fig. 1. A, Integrated UMAP embedding of mouse (597,668 nuclei) from PFC, motor cortex, striatum, and thalamus (as dissected, before reassignment) across all developmental time points assayed and a randomly downsampled portion of adult nuclei from our previous study[1] colored by **(i)** sex and **(ii)** biological replicate (for most mouse time points, is the same as individual donor). **B**, Stacked bar plots showing the number or proportion of full QC-passed nuclei for each **(i)** dissected brain region, **(ii)** age (number of nuclei), **(iii)** age (proportion of nuclei) **(iv)** sex (proportion of nuclei) and **(v)** biological replicate, broken down by assigned cell type. **C**, Log2-fold change in cell type composition as estimated by scCODA for the coefficients on **(i)** age (reference = P32), **(ii)** region (reference = cortex), and **(iii)** male sex (reference = female) for. Only fold-changes for credible effects (i.e., significantly nonzero coefficient) are shown. The 597,668 mouse nuclei were composed of 12 broad cell classes: glutamatergic excitatory neurons (44.7%), GABAergic inhibitory neurons (21.8%), neurons belonging to mixed glutamatergic and GABAergic clusters (5.54%), astrocytes (11.5%), oligodendrocytes (8.41%), oligodendrocyte progenitor cells (OPCs, 3.91%), microglia and macrophages (2.98%), mural cells (0.28%), endothelial cells (0.26%), vascular-associated fibroblasts (0.62%), ependymal cells (0%; these instead co-clustered with astrocytes), and choroid plexus cells (0.12%).

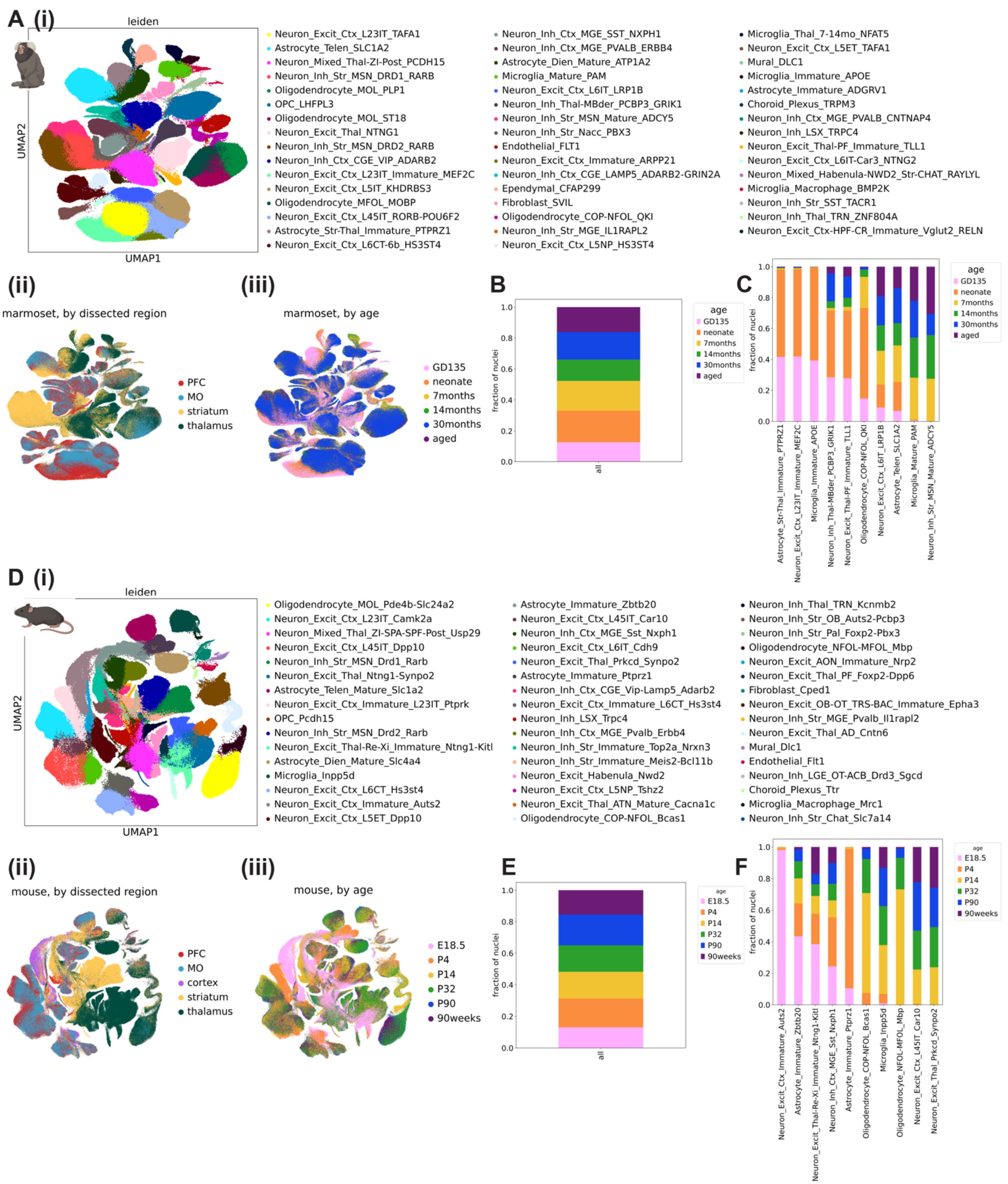


Figure S4. Leiden cluster annotations and proportions for the cross-region developmental snRNAseq atlas, related to Fig. 1. A, Integrated UMAP embedding of marmoset (**A**, 881,832) or mouse (**B**, 597,668) nuclei from PFC, motor cortex, striatum, and thalamus (as dissected, before reassignment) across all developmental time points assayed and a randomly downsampled portion of adult nuclei from our previous study[1] colored by **(i)** Leiden-algorithm determined cluster annotation (hereafter, Leiden cluster), **(ii)** dissected region and **(iii)** developmental time point. **B**, Stacked bar plot showing the proportion of all marmoset nuclei from each developmental time point (see legend in **A(iii)**). **C,** Stacked bar plot showing the proportion of each Leiden cluster from each developmental time point for a hand-selected set of 10 clusters that are developmentally enriched or depleted. **D,** Stacked barplots showing the number of nuclei assigned to each of the top 5 most abundant Allen Brain Cell Atlas mouse whole-brain atlas class or subclass using The Allen Institute's MapMyCells (see **Methods**) for **(i)** excitatory neurons **(ii)** inhibitory neurons and **(iii)** astrocytes, broken down by dissected brain region. **E-H,** Same as **A-D,** for mouse.

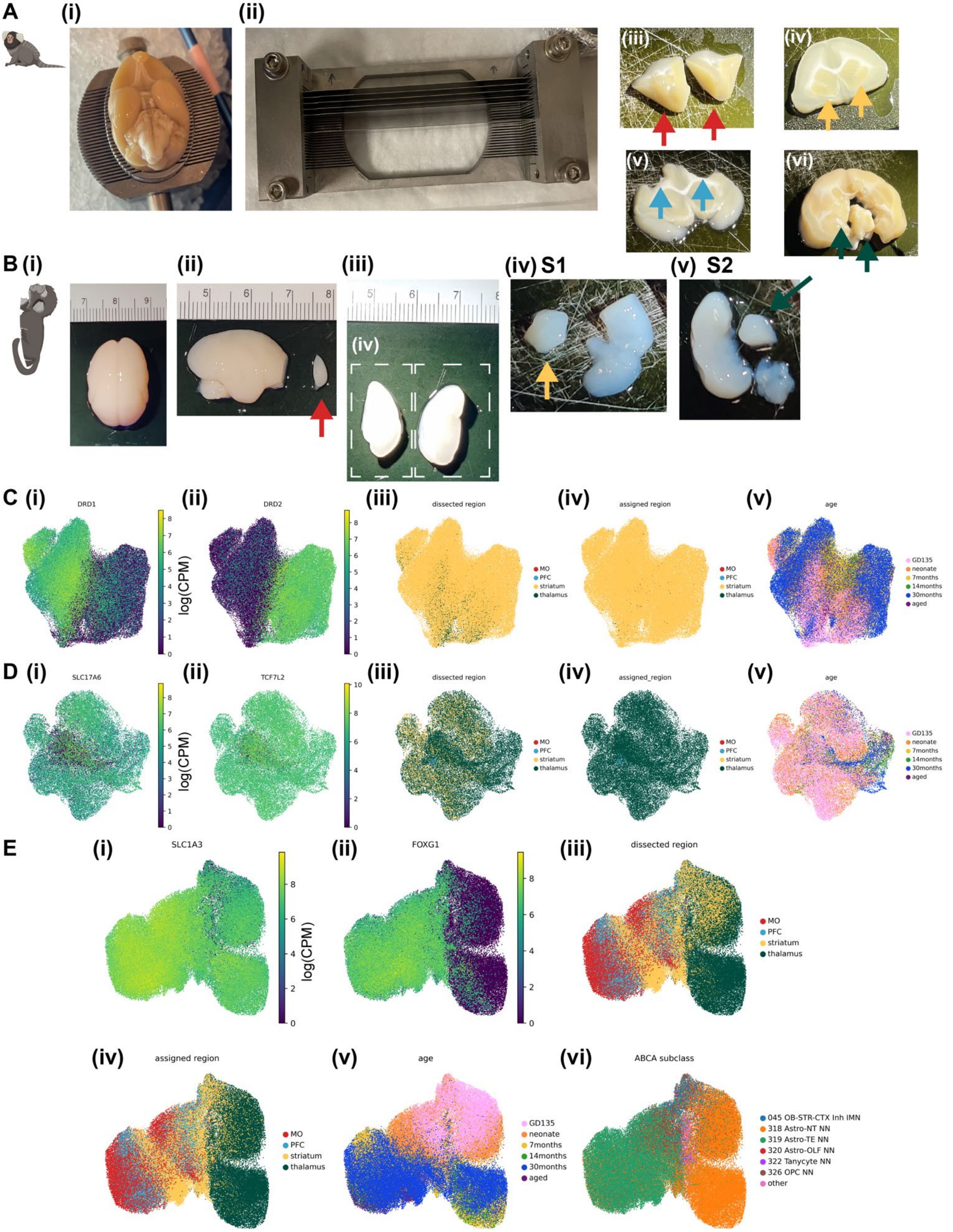


Figure S5. Dissection strategies for marmoset brains and region reassignment to mitigate cross-region contamination for developing samples, related to Fig. 1. A, Dissection strategy for 7-month, 14-month, and aged marmosets used in this study. **(i)** Ventral view of brain from female 14-month marmoset donor 21-164 in brain matrix. **(ii)** Blades positioned in blade holder to create desired tissue slabs for dissection. Arrows on the blade holder indicate anterior direction. **(iii)** Tissue slab for prefrontal cortex (PFC) dissection, with red arrows indicating PFC. Some white matter was removed, and then the entire slab was frozen for nuclei isolation. **(iv)** One out of two tissue slabs used for striatum dissection, with yellow arrows indicating striatum (not yet dissected). **(v)** Tissue slab used for motor cortex dissection, with blue arrows indicating where motor cortex was removed (already removed in this photo). **(vi)** Tissue slab used for thalamus dissection, with dark green arrows indicating the thalamus (already dissected from the left hemisphere, but not yet removed). **B,** Dissection strategy for GD135 marmoset (similar for neonate). **(i)** Dorsal view of brain from female GD135 donor 19-221_fetus1. **(ii)** Prefrontal cortex as grossly dissected from lateral surface of the brain. **(iii)** Two slabs cut from the bissected region spanning approximately from the anterior beginning of the temporal lobe to its posterior end, used to dissect the striatum (slab 1, S1) and thalamus (slab 2, S2). **(iv)** Striatum as dissected from slab 1 (S1). **(v)** Thalamus as dissected from slab 2 (S2). **C**, Analysis of dissection cross-contamination in marmoset (all ages, but only nuclei from GD135 and neonate were reassigned) using medium spiny neurons (MSNs). **(i-v)** UMAP projection of marmoset MSNs colored by **(i)** *DRD1* expression in logCPM units, **(ii)** *DRD2* expression in logCPM units, **(iii)** dissected brain region, **(iv)** reassigned brain region, and **(v)** age. **(iii)** Shows contamination of thalamus in striatal dissections for GD135 and neonate samples. **D,** Same as **(C)**, for marmoset thalamic excitatory neurons, with **(i)** showing *SLC17A6* expression and **(ii)** showing *TCF7L2* expression. **(iii)** Shows contamination of striatum in thalamic dissections for GD135 and neonate samples. **E,** Analysis of dissection cross-contamination in marmoset (all ages, but only nuclei from GD135 and neonate were reassigned) astrocytes of all clusters. UMAP projection of marmoset astrocytes colored by **(i)** *SLC1A3* expression, **(ii)**, *FOXG1* expression, **(iii)** dissected brain region, **(iv)** reassigned brain region, **(v)** age, and **(vi)** ABCA subclass, as determined using MapMyCells.

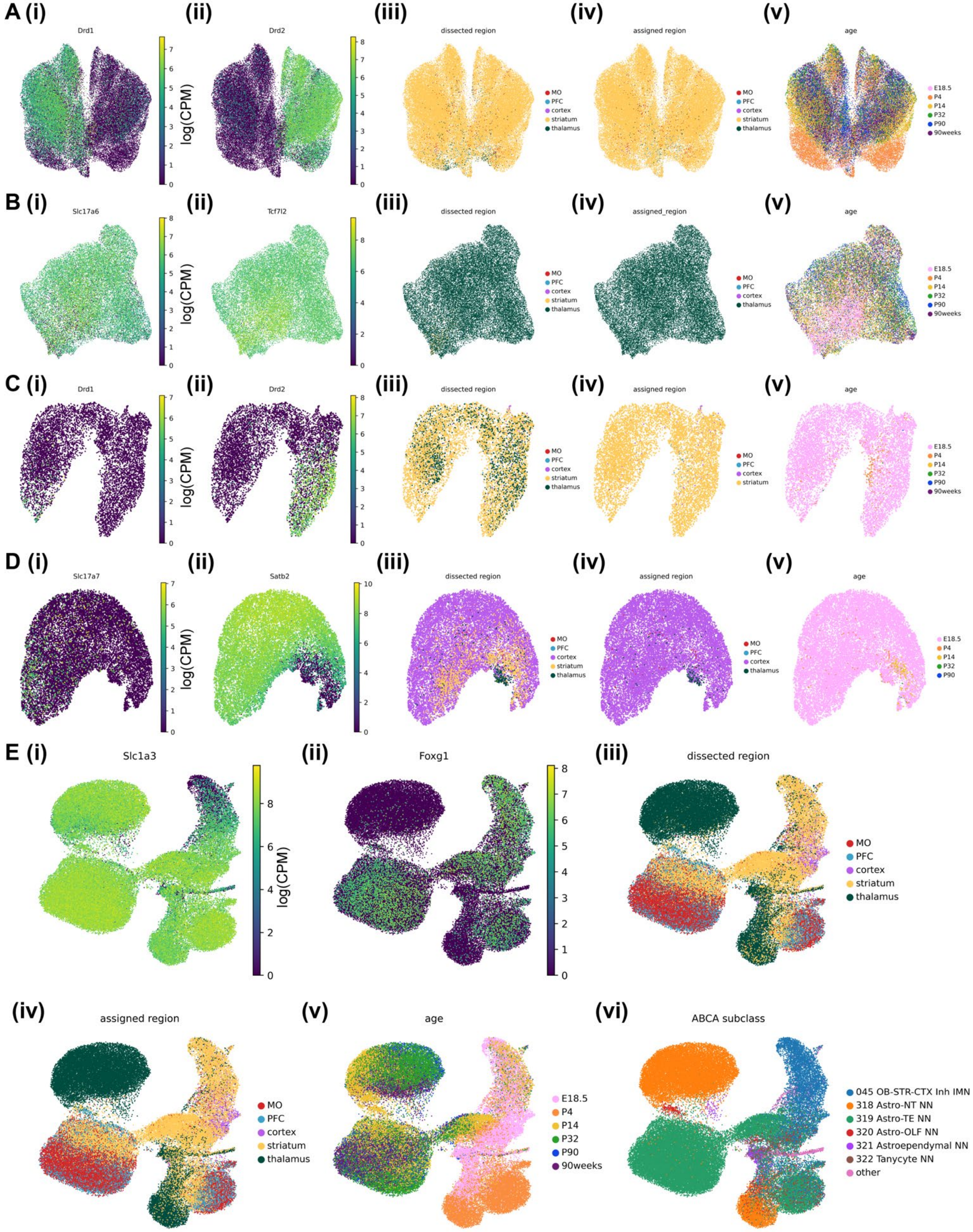


Figure S6. Region reassignment for cross-contaminant nuclei in mouse samples, related to Fig. 1. A, Analysis of dissection cross-contamination in mouse (all ages) using medium spiny neurons (MSNs). **(i-v)** UMAP projection of mouse MSNs colored by **(i)** *Drd1* expression in logCPM units, **(ii)** *Drd2* expression in logCPM units, **(iii)** dissected brain region, **(iv)** reassigned brain region, and **(v)** age. **B,** Same as **(A)**, for mouse thalamic excitatory neurons, with **(i)** showing *Slc17a6* expression and **(ii)** showing *Tcf7l2* expression. **C,** Same as **(A),** for immature mouse MSNs, showing a higher degree of striatum-thalamus cross-contamination in E18.5 mouse samples. **D,** Same as **(A-C)**, for immature mouse cortical neurons, with **(i)** showing *Slc17a7* expression and **(ii)** showing *Satb2* expression. **(iii)** Illustrates cortical contamination of striatal and thalamic dissections in the mouse E18.5 data. **E,** Analysis of dissection cross-contamination in mouse (all ages) astrocytes of all clusters. UMAP projection of mouse astrocytes colored by **(i)** *Slc1a3* expression, **(ii)**, *Foxg1* expression, **(iii)** dissected brain region, **(iv)** reassigned brain region, **(v)** age, and **(vi)** ABCA subclass, as determined using MapMyCells.

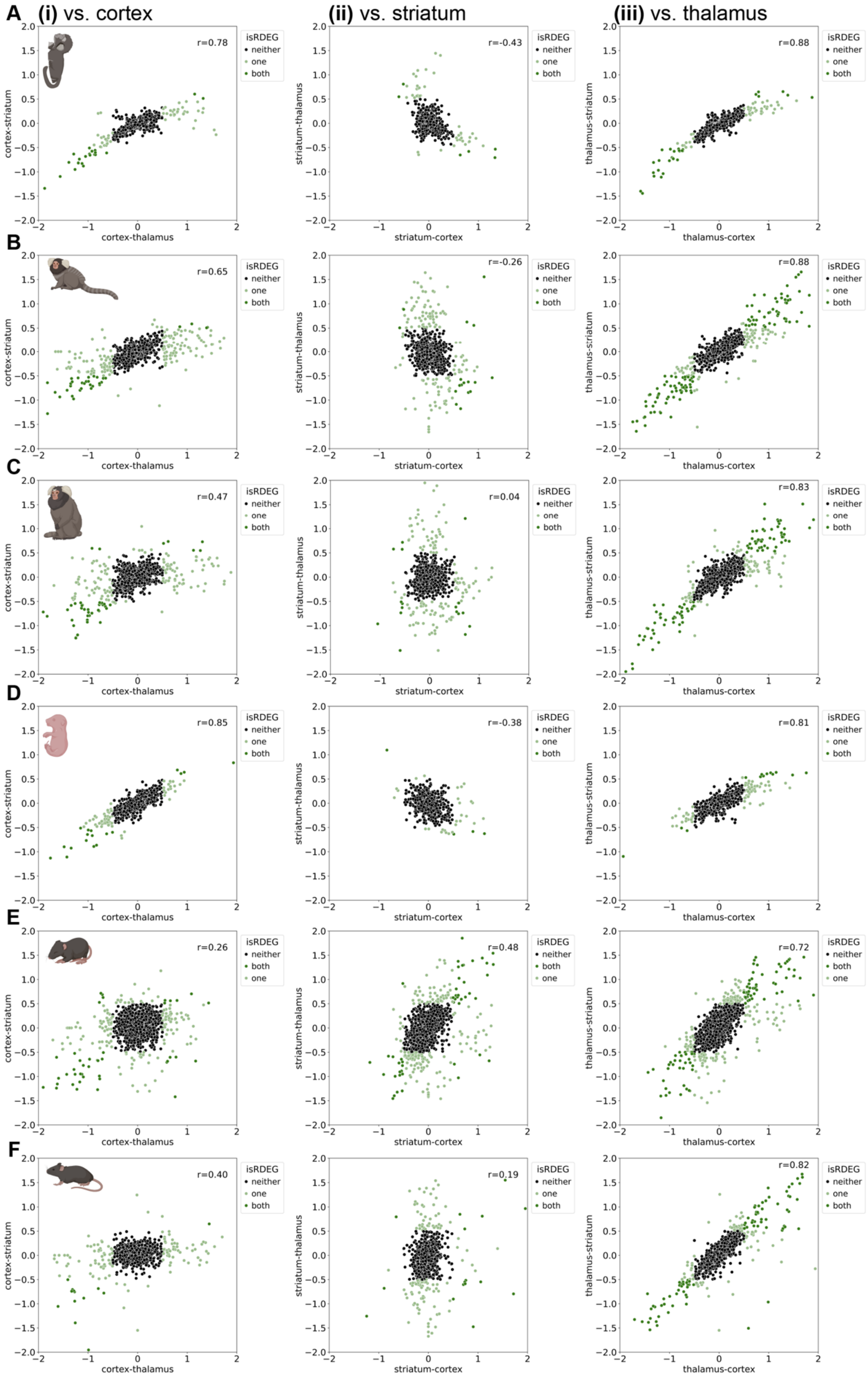


Figure S7. Correlation of pairwise astrocyte rDEG log-fold change between region pairs across development in mouse and marmoset, related to Figs. 2 and 3. A, Scatterplot of log fold-change (logFC) difference in expression in marmoset GD135 astrocytes for genes meeting minimum metacell expression criteria (relaxed from the criteria for rDEGs shown in **Figs. 2-3**, see **Methods**) for 2 different region pairs: **(i)** cortex-striatum vs. cortex-thalamus (a negative value indicates downregulation in cortex), **(ii)** striatum-thalamus vs. striatum-cortex (a negative value indicates downregulation in striatum), and **(iii)** thalamus-striatum vs. thalamus-cortex (a negative value indicates downregulation in thalamus). Light green dots indicate genes who meet rDEG (magnitude of logFC > 0.5) criteria for one region pair, and dark green dots indicate genes who meet rDEG criteria for both region pairs. In marmoset, logFC is averaged across replicates. r, Pearson's correlation coefficient calculated using all the points shown. **B,** Same as **(A)** for 7-month marmoset. **C,** Same as **(A)** for 30-month marmoset. **D**, Same as **(A)** for E18.5 mouse. **E**, Same as **(A)** for P14 mouse. **F**, Same as **(A)** for P90 mouse.


A
(i) PFC
(ii) striatum
(iii) thalamus
SLC1A3 FOXG1 DNM3 merge
RNAscope HiPlex
B
GFAP KCND2 CSMD1 merge
C
SLC1A3 SPARC KCNH7 merge
RNAscope v2
D
(i) PFC
(ii) striatum
(iii) thalamus
SLC1A3 FOXG1 DNM3 merge
RNAscope HiPlex
E
GFAP KCND2 CSMD1 merge
F
SLC1A3 SPARC KCNH7 merge
RNAscope v2

Figure S8. Validation of marmoset astrocyte rDEG expression in situ using multiplexed and single-round FISH, related to Fig. 2. A, Single-channel and composite registered (across rounds) maximum-projected images of telencephalic astrocyte rDEGs for 3 exemplary astrocytes (low autofluorescence and high signal-to-noise) from one adult female marmoset in **(i)** PFC, **(ii)** striatum, and **(iii)** thalamus, generated using the RNAscope HiPlex protocol. *SLC1A3* was used to identify astrocytes. Throughout the images, DAPI nuclear counterstain is shown in gray. **B,** Same as **(A)**, for diencephalic astrocyte rDEGs. Scale bar, 5μm. **C,** Single-channel and composite images of rDEGs for 3 representative astrocytes from one adult female marmoset generated using the RNAscope v2 protocol. Scale bar, 5μm. **D-F,** Same as **A-C**, for one male **(D-E)** and one female **(F)** neonate marmoset. Contrast was manually adjusted by setting minimum and maximum intensity values, using the same values across regions within each age, except for *SLC1A3* in adult **(A)**, for which we reduced the maximum intensity value in the thalamus (vs. cortex and striatum), to account for lower expression in order to visualize its expression in astrocytes.

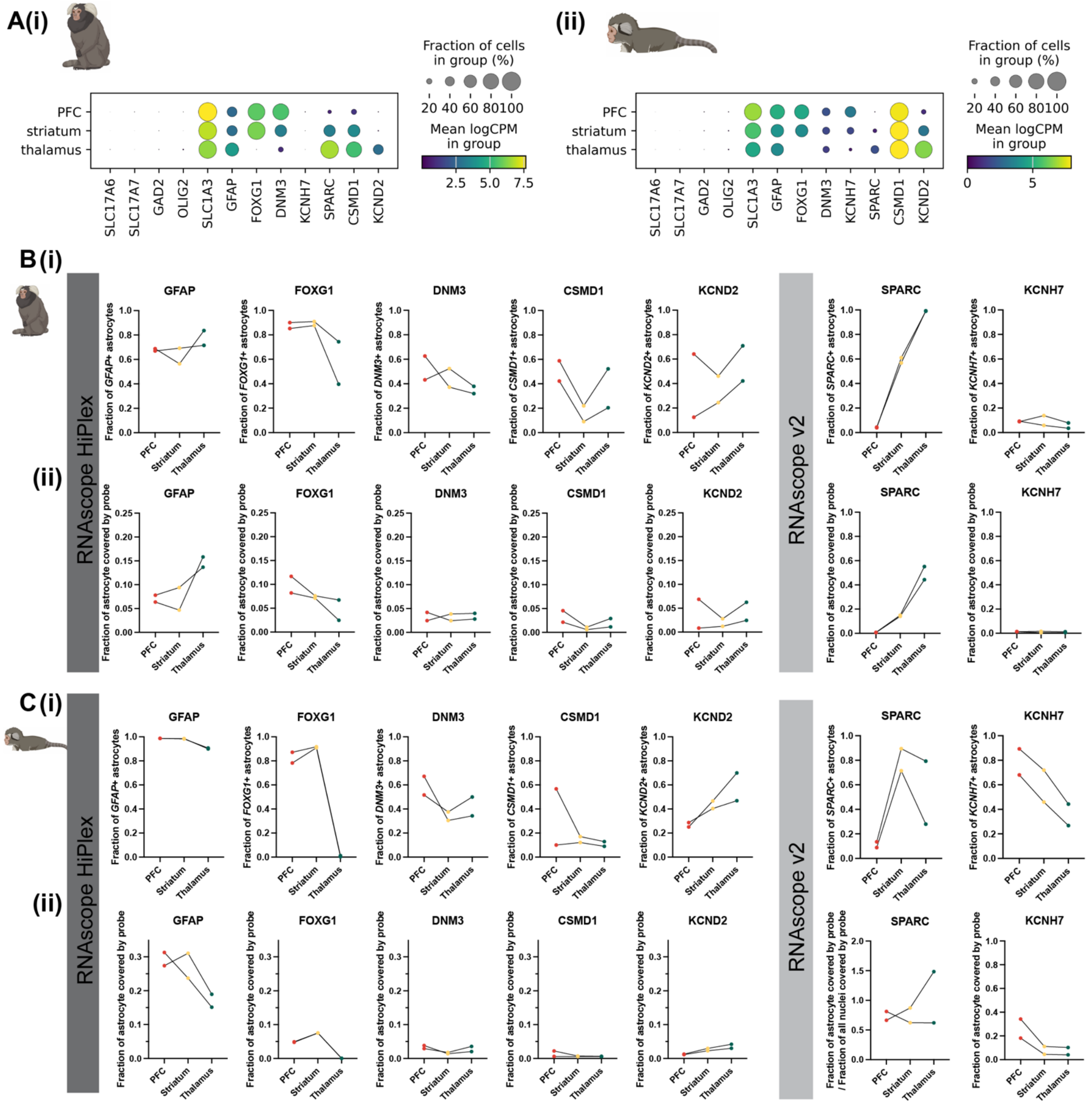


Figure S9. Quantification of selected rDEG and astrocyte subtype marker expression *in situ* in adult and neonate marmoset, related to Fig. 2. A, Mean expression (in logCPM from the snRNAseq data) of selected rDEG (*DNM3, KNC7, SPARC, CSMD1, KCND2*) and astrocyte subtype (*GFAP*) genes by region in the **(i)** adult (30 month) marmoset and **(ii)** neonate marmoset. **B, (i)** Line plots showing the fraction of astrocytes positive for these genes in PFC (red), striatum (yellow), and thalamus (green) from the RNAscope HiPlex or v2 FISH data (**Fig. S8**) for 2 adult female marmosets. **(ii)** Line plots showing the mean fraction of astrocytes covered by probes as in **(i)**. See **Methods** for details on quantification, which was performed semi-automatically using CellProfiler. Colored dots connected by lines are data originating from the same marmoset donor. **C,** Same as **(B)**, for two neonate marmoset donors, one male and one female.

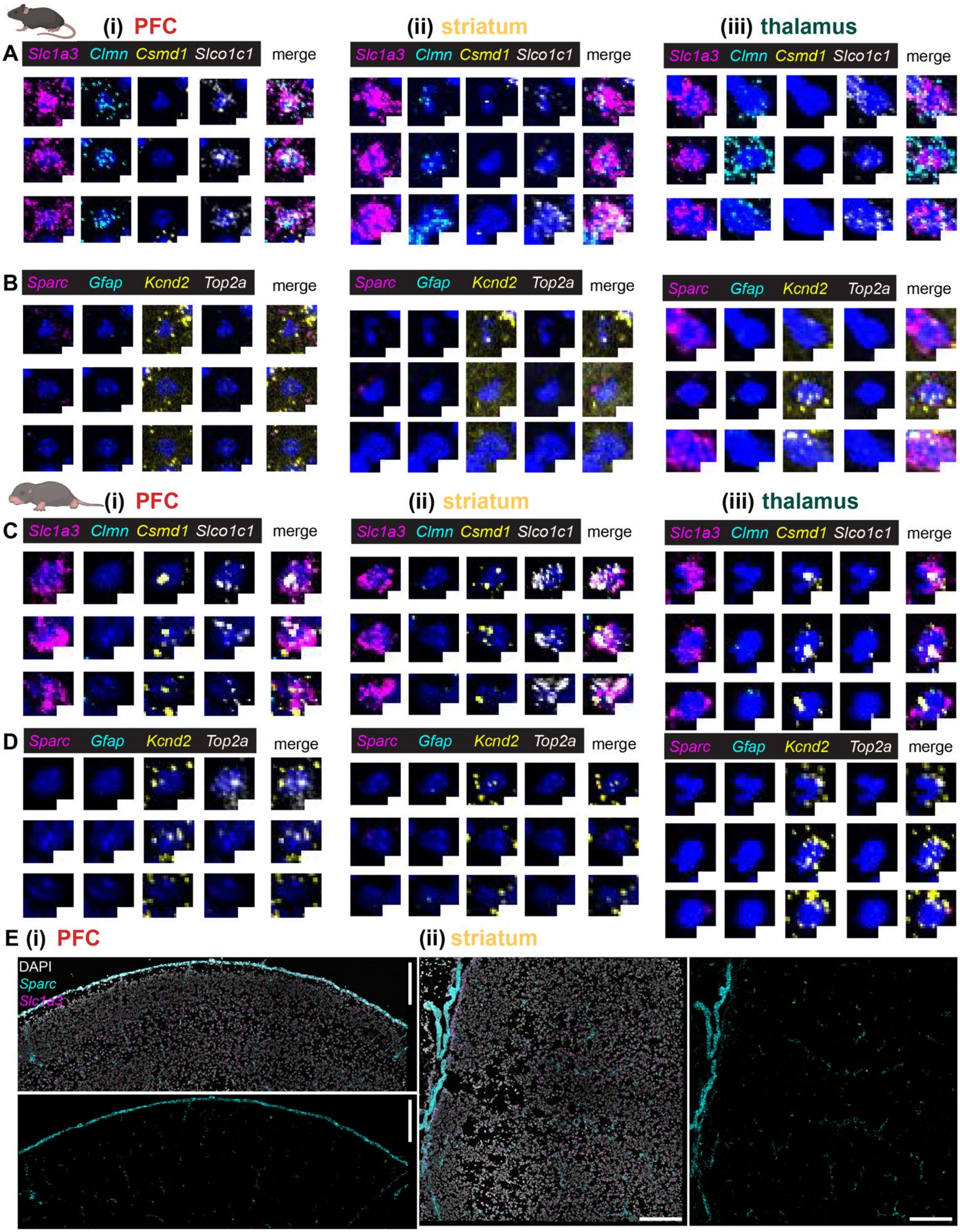


Figure S10. Validation of mouse astrocyte rDEG expression in situ using multiplexed FISH, related to Fig. 3. A, Single-channel and composite registered (across rounds), maximum-projected images of telencephalic astrocyte rDEGs for 3 exemplary astrocytes in the **(i)** PFC **(ii)** striatum and **(iii)** thalamus in one female P90 mouse, generated using RNAscope HiPlex FISH. **B,** Same as **(A)**, for diencephalic rDEGs (*Sparc, Kcnd2*) and other astrocyte subtype markers (*Gfap* and *Top2a*). Scale bar, 5µm. **C-D,** Same as **A-B**, for one male P4 mouse. Contrast was manually adjusted by setting minimum and maximum intensity values, using the same values across regions within each age. **E,** *Sparc* expression pattern in early postnatal mouse. Composite and single-channel registered, max-projected, stitched images of the entire field of view imaged for **(i)** PFC and **(ii)** striatum in one P4 female mouse, generated using RNAscope HiPlex FISH. *Sparc* staining is shown alone to aid visualization of its high signal intensity along brain borders (pial surface and ventricles) and along the vasculature (based on morphology) at this time point. *Sparc* contrast was increased to aid visualization in the larger, lower-zoom images. The PFC image has been rotated 90° from its original orientation. Scale bar, 200µm.

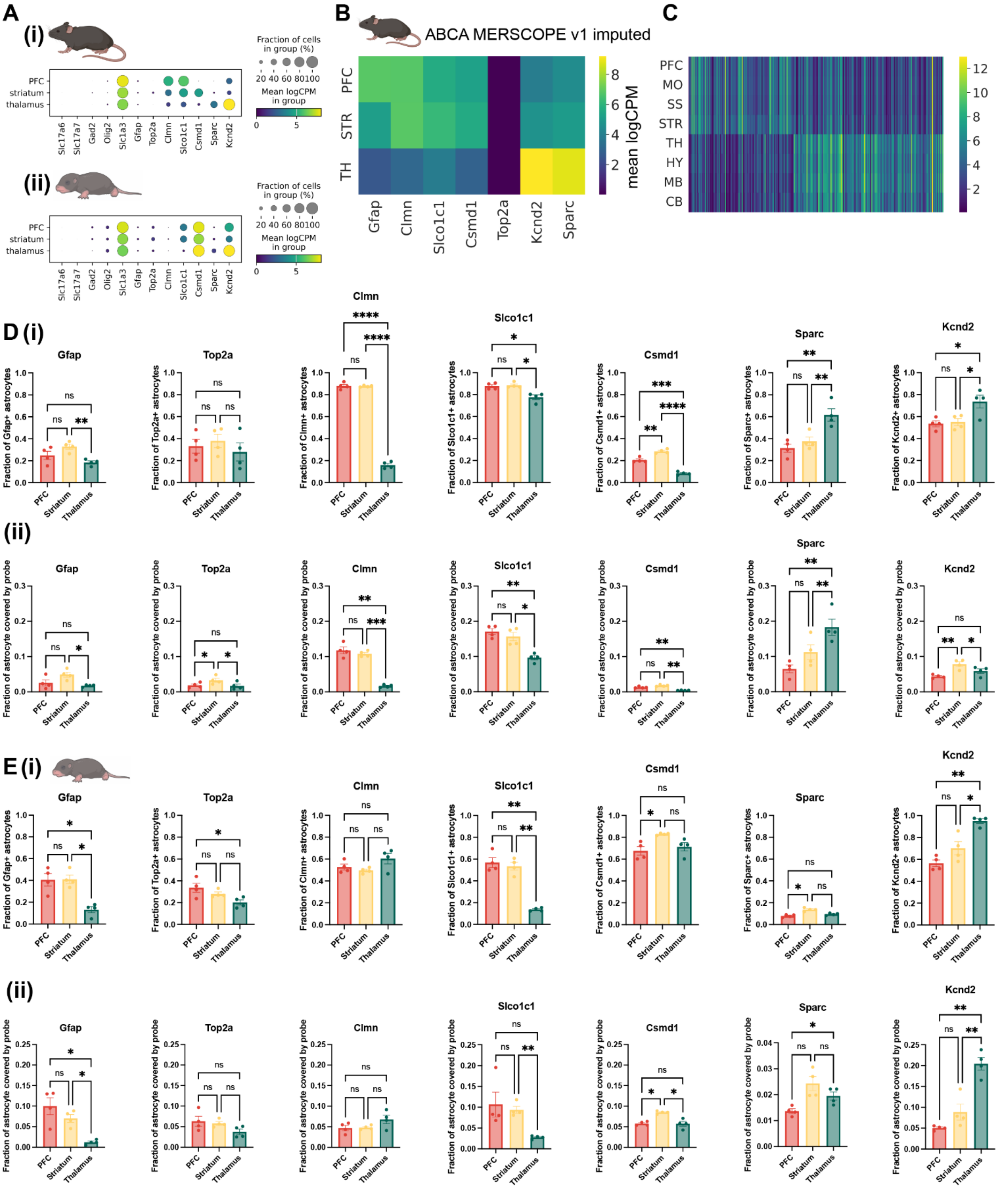


Figure S11. Quantification of selected rDEG and astrocyte subtype marker expression *in situ* in adult and neonate mouse, related to Fig. 3. A, Mean expression (in logCPM from the snRNAseq data) of selected rDEG (*Clmn, Slco1c1, Csmd1, Sparc,* and *Kcnd2*) and astrocyte subtype (*Gfap, Top2a*) genes by region in the **(i)** adult (P90) and **(ii)** neonate (P4) mouse. **B,** Mean expression (across astrocyte nuclei in each brain region, in logCPM) of the genes in **(A)** in astrocytes from the Allen Brain Cell Atlas (ABCA) MERSCOPE v1 whole mouse brain dataset with imputed genes from 10x Genomics scRNAseq (https://knowledge.brain-map.org/abcatlas)[3]. PFC, prefrontal cortex. STR, striatum. TH, thalamus. **C,** Mean expression (across astrocyte nuclei in each brain region, in logCPM) of all mouse P90 astrocyte rDEGs from the ABCA MERSCOPE v1 whole mouse brain dataset as in **(B)**, with additional brain regions added. SS, somatosensory cortex. HY, hypothalamus. MB, midbrain. CB, cerebellum. **(i)** Bar plots showing the mean fraction of astrocytes positive for these genes in PFC (red), striatum (yellow), and thalamus (green) from FISH data in P90 mice (**Fig. S10**, n = 4 mice, 2 of each sex, data point shows average of 2 tissue sections each). **(ii)** Bar plots showing the mean fraction of astrocytes covered by probes for the genes in **(i)** in PFC (red), striatum (yellow), and thalamus (green) from FISH data (**Fig. S10**, n = 4 mice, 2 of each sex, data point shows average of all astrocytes from 2 tissue sections each). Error bars indicate standard error of the mean. See **Methods** for details on quantification, which was performed semi-automatically using CellProfiler. Statistical significance was determined in GraphPad Prism using a repeated-measures one-way ANOVA, followed by Tukey's multiple comparisons test (see **Table S13** for full statistics for each gene). ns, non significant, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$. **D,** Same as **(C)**, for P4 mice (n = 4, 2 of each sex, data point shows average across 2 tissue sections each).

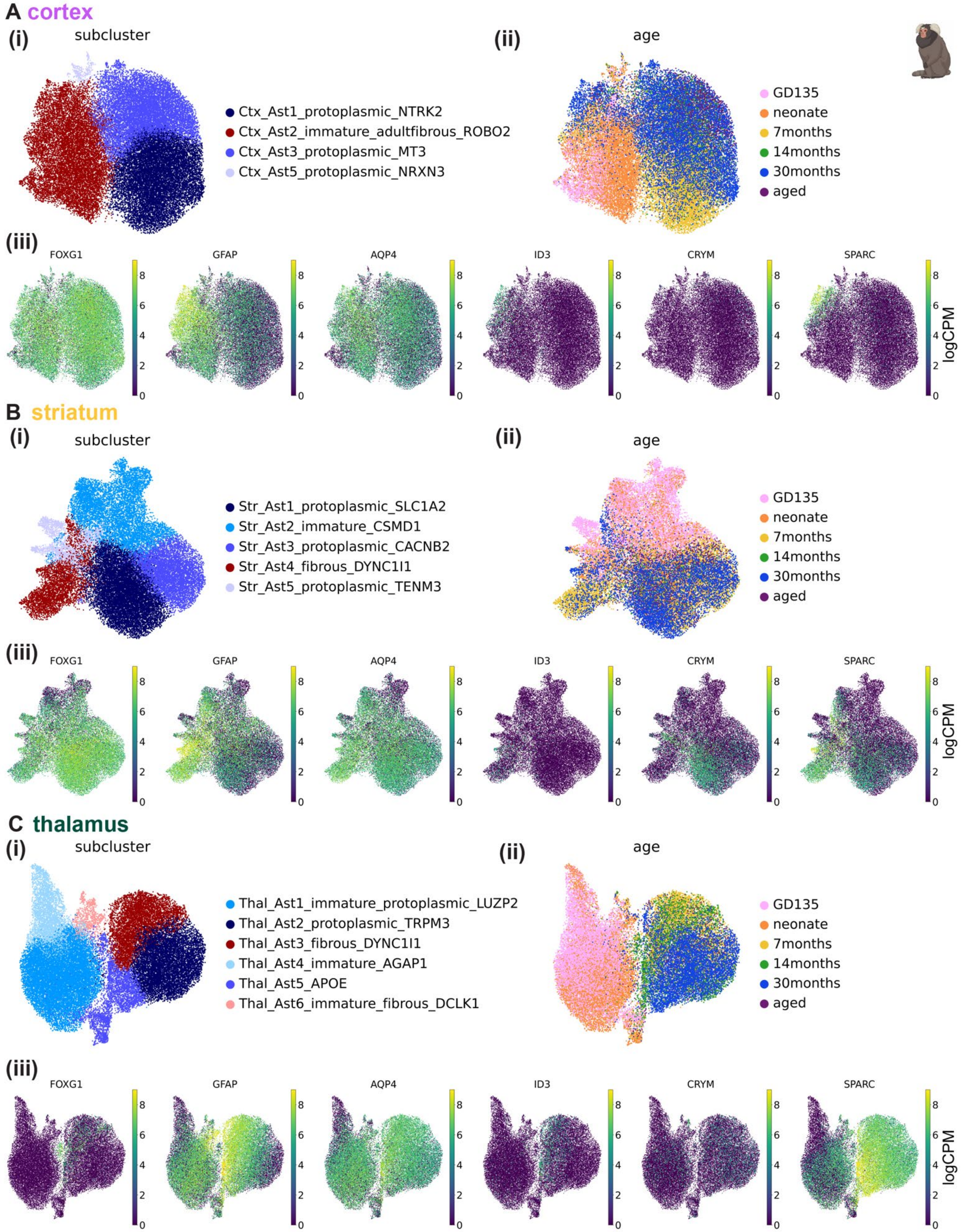


Figure S12. Astrocyte sub-clustering captures intra-regional heterogeneity in marmoset, related to Fig. 2. **A,** UMAP embedding of 36,136 marmoset cortical astrocytes from all developmental time points colored by **(i)**

subcluster (see **Methods**), **(ii)** age, or **(iii)** *FOXG1, GFAP, AQP4, ID3, CRYM,* or *SPARC* expression in logCPM units. *FOXG1* marks telencephalic astrocytes, *GFAP, AQP4*, and *ID3* mark fibrous and interlaminar astrocytes[4], *CRYM* marks striatal astrocytes[5], and *SPARC* marks thalamic astrocytes[6]. **B,** Same as **(A)** for 29,931 marmoset striatal astrocytes. **C,** Same as **(A)** for 35,493 marmoset thalamic astrocytes. There are 1,380 *FOXG1+* astrocytes in the thalamus clusters, ~81% of which are assigned to the non-telencephalic astrocyte ABCA subclass, most but not all of which come from the donors from the current study, and may reflect contamination from neighboring regions such as globus pallidus or septum. **D-F**, same as **(A-C)** for mouse astrocytes: **(D)** 25,558 from cortex **(E)** 19,687 from striatum, and **(F)** 23,240 in thalamus. *Top2a* marks immature and SVZ rostral migratory stream astrocytes.

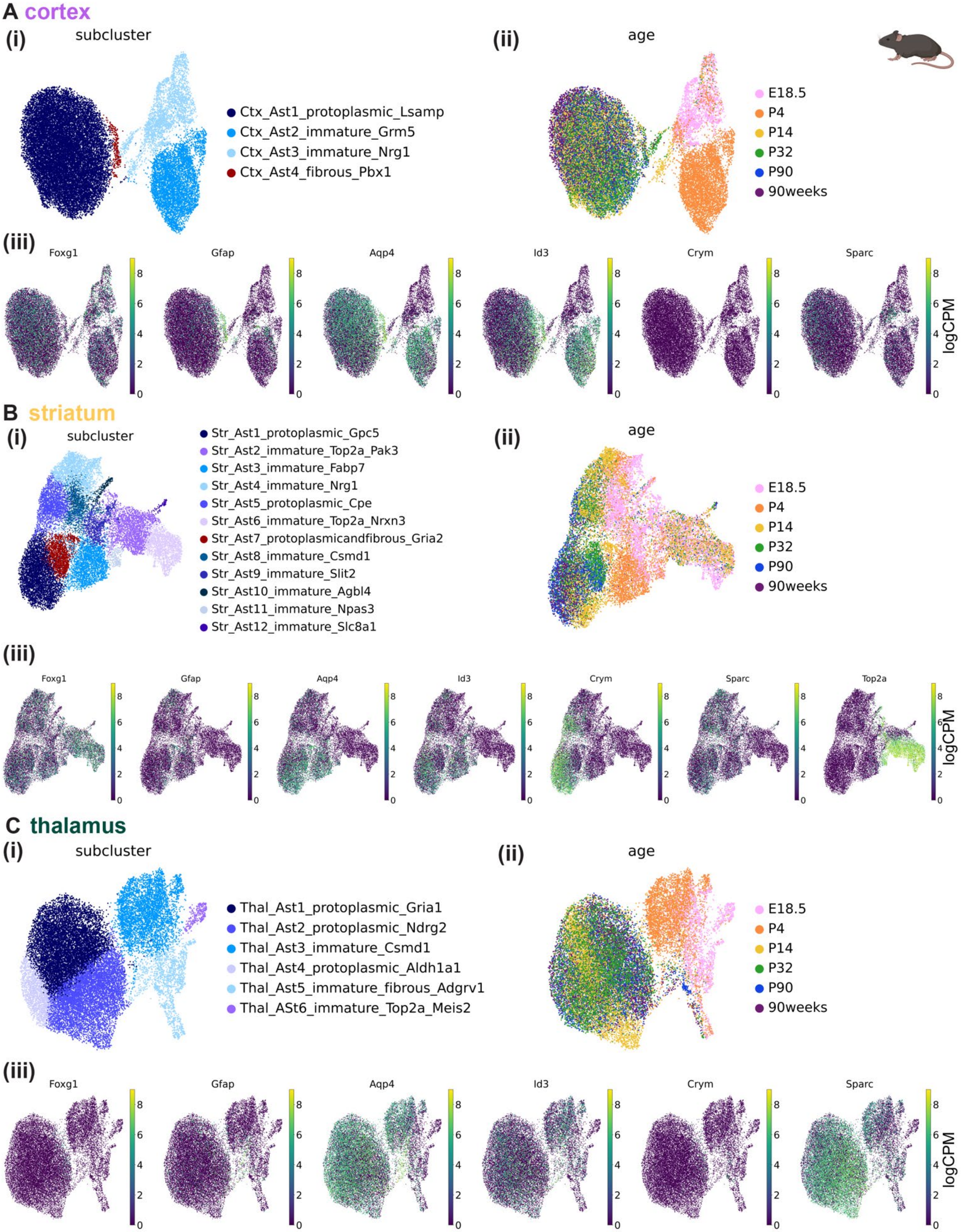


Figure S13. Astrocyte sub-clustering captures intra-regional heterogeneity in mouse, related to Fig. 3. A, UMAP embedding of 36,136 mouse cortical astrocytes from all developmental time points colored by **(i)** subcluster (see **Methods**), **(ii)** age, or **(iii)** *Foxg1, Gfap, Aqp4, Id3, Crym, Sparc* expression in logCPM units. *Foxg1* marks telencephalic astrocytes, *Gfap, Aqp4*, and *Id3* mark fibrous and interlaminar astrocytes[4], *CRYM* marks striatal astrocytes[5], and *SPARC* marks thalamic astrocytes[6]. **B,** Same as **(A)** for 19,687 striatal astrocytes, with the addition of *Top2a* plotted in **B(iii)**. **C,** Same as **(A)** for 23,240 thalamic astrocytes.

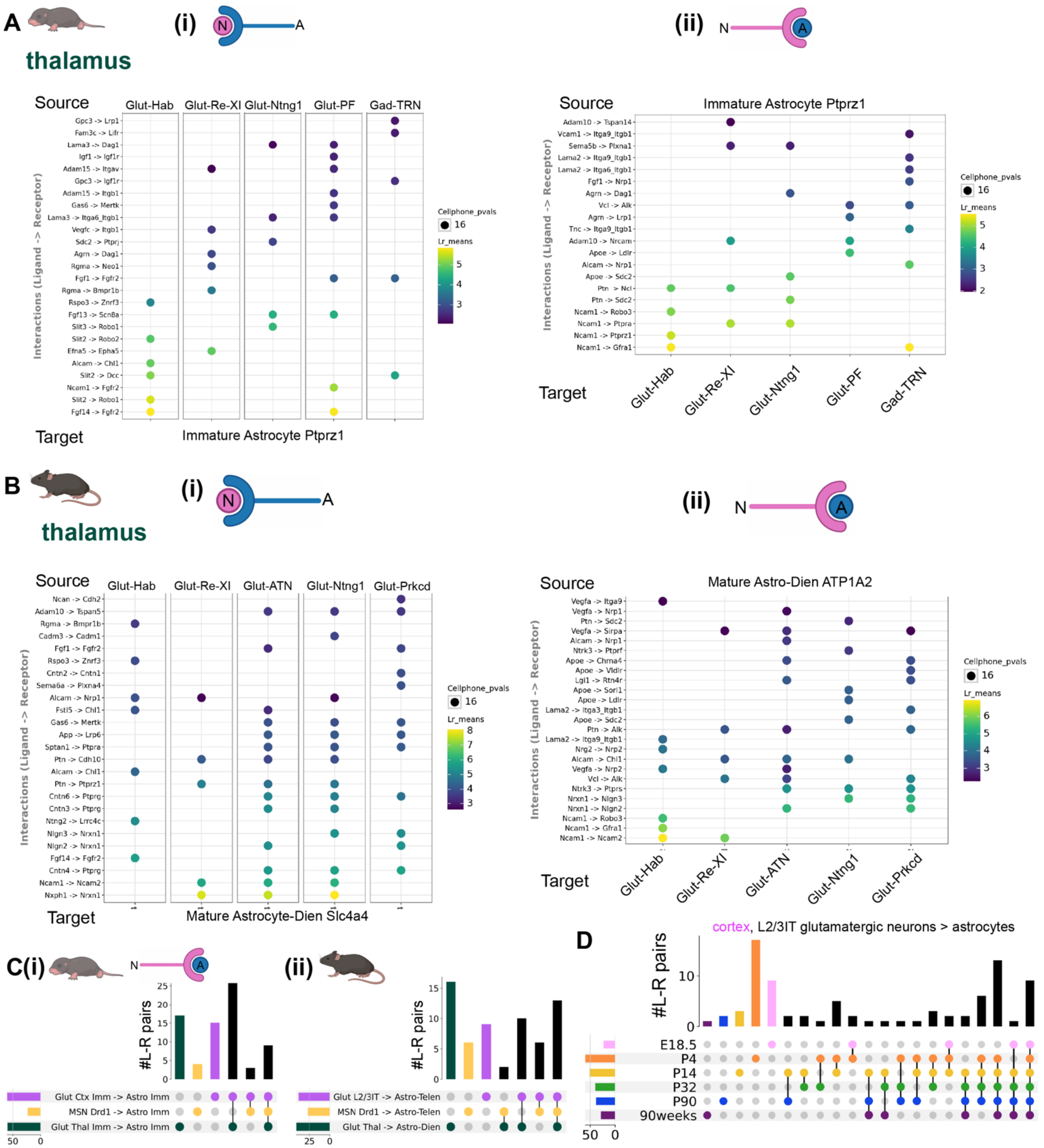


Figure S14. Cell-cell communication analysis for neuron-astrocyte and astrocyte-neuron predicted ligand-receptor pairs across regions and developmental time points in mouse, related to Fig. 4. A, Dot plot showing magnitude and specificity of the top 25 near-unique (shared with at most one other neuronal cluster) CellPhoneDB-predicted **(i)** neuron-astrocyte and **(ii)** astrocyte-neuron ligand receptor pairs for the most abundant astrocyte and neuronal Leiden clusters in the neonate mouse thalamus. The source cell (top of the plot) expresses the ligand (left side of arrow on the row labels), while the target cell (bottom of the plot) expresses the receptor (right side of arrow on the row labels). The color of the dot indicates ligand-receptor expression magnitude (“Lr_means”, calculated as the average of the mean expression of the ligand in the source group and the mean expression of the receptor in the target group), while the size of the dot is inversely related to the p-value on ligand-receptor expression sensitivity (see **Methods**). Glut-Hab, glutamatergic habenula; Glut-Re-XI, glutamatergic reunions/xiphoid nucleus; Glut-PF, glutamatergic parafascicular nucleus; Gad-TRN, GABAergic thalamic reticular nucleus. **B,** Same as **(A)**, for the P90 mouse thalamus, with the uniqueness criteria relaxed to up to 3 neuronal clusters. Glut-ATN, glutamatergic anterior thalamic nuclei. **C,** UpSet plot showing the number of overlapping neuron-astrocyte predicted ligand-receptor pairs between regions, from the most abundant neuronal and astrocyte subtypes in each region for **(i)** fetal and **(ii)** late adolescent mouse. For cortex (purple), glutamatergic L2/3IT neurons to cortical astrocytes; striatum (yellow), *Drd1*+ medium spiny neurons to striatal telencephalic astrocytes; and thalamus (green) thalamic glutamatergic neurons to thalamic astrocytes. Unlike in panel **A**, all L-R pairs meeting minimum expression criteria, including pairs shared with other neuronal and astrocytic clusters, were included in this analysis and that for panel **(D)**. **D**, UpSet plot (as in **(C)**) showing the number of overlapping cortical glutamatergic L2/3IT neuron to cortical astrocyte predicted ligand-receptor pairs between ages.

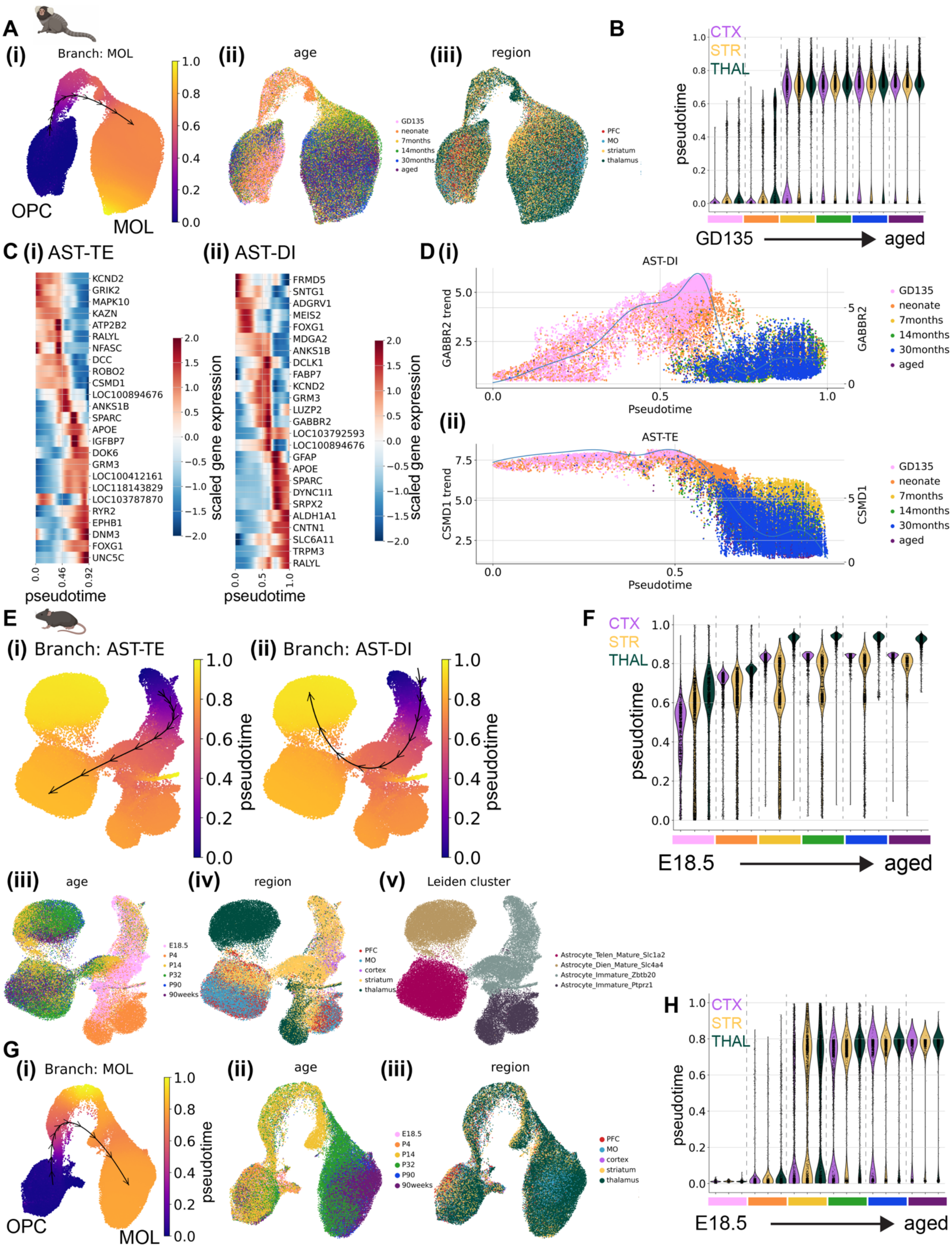

A
(i) Branch: MOL
OPC
MOL
1.0
0.8
0.6
0.4
0.2
0.0
(ii) age
GD135
neonate
7months
14months
30months
aged
(iii) region
PFC
MO
striatum
thalamus
B
CTX
STR
THAL
pseudotime
1.0
0.8
0.6
0.4
0.2
0.0
GD135
aged
C (i) AST-TE
KCND2
GRIK2
MAPK10
KAZN
ATP2B2
RALYL
NFASC
DCC
ROBO2
CSMD1
LOC100894676
ANKS1B
SPARC
APOE
IGFBP7
DOK6
GRM3
LOC100412161
LOC118143829
LOC103787870
RYR2
EPHB1
DNM3
FOXG1
UNC5C
0.0
0.46
0.92
pseudotime
scaled gene expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
(ii) AST-DI
FRMD5
SNTG1
ADGRV1
MEIS2
FOXG1
MDGA2
ANKS1B
DCLK1
FABP7
KCND2
GRM3
LUZP2
GABBR2
LOC103792593
LOC100894676
GFAP
APOE
SPARC
DYNC1I1
SRPX2
ALDH1A1
CNTN1
SLC6A11
TRPM3
RALYL
0.0
0.5
1.0
pseudotime
scaled gene expression
D (i)
AST-DI
GABBR2 trend
5.0
2.5
0.0
0.5
1.0
Pseudotime
GABBR2
5
0
GD135
neonate
7months
14months
30months
aged
(ii)
AST-TE
CSMD1 trend
7.5
5.0
2.5
0.0
0.5
Pseudotime
CSMD1
5
0
GD135
neonate
7months
14months
30months
aged
E
(i) Branch: AST-TE
(ii) Branch: AST-DI
1.0
0.8
0.6
0.4
0.2
0.0
pseudotime
(iii) age
E18.5
P4
P14
P32
P90
90weeks
(iv) region
PFC
MO
cortex
striatum
thalamus
(v) Leiden cluster
Astrocyte_Telen_Mature_Slc1a2
Astrocyte_Dien_Mature_Slc4a4
Astrocyte_Immature_Zbtb20
Astrocyte_Immature_Ptprz1
F
CTX
STR
THAL
pseudotime
1.0
0.8
0.6
0.4
0.2
0.0
E18.5
aged
G
(i) Branch: MOL
OPC
MOL
1.0
0.8
0.6
0.4
0.2
0.0
pseudotime
(ii) age
E18.5
P4
P14
P32
P90
90weeks
(iii) region
PFC
MO
cortex
striatum
thalamus
H
CTX
STR
THAL
pseudotime
1.0
0.8
0.6
0.4
0.2
0.0
E18.5
aged

Figure S15. Pseudotime inference in mouse and marmoset oligodendrocyte lineage and mouse astrocytes, related to Fig. 5. A, Integrated UMAP embeddings of 100,000 (out of 170,786, randomly downsampled) marmoset oligodendrocyte lineage cells colored by **(i)** Palantir-predicted pseudotime, with trajectory path (black lines and arrows) overlaid, **(ii)** developmental time point, and **(iii)** brain region. **B,** Violin plot (scanpy's default) showing the estimated distribution of pseudotime values for the oligodendrocyte lineage cells in **(A)** grouped brain region and developmental time point. Color scheme for ages as in **(ii)**, with vertical dashed lines indicating divisions between time points. **C,** Heatmap of scaled gene expression over pseudotime for the 25 genes with the highest Mellon change scores for the **(i)** AST-TE and **(ii)** AST-DI branches of marmoset astrocytes. **D,** Individual (single dots) and average (blue line) gene trends (MAGIC-imputed expression over pseudotime) for 2 of the top change-scoring genes for the **(i)** AST-DI (*GABBR2*) and **(ii)** AST-TE (*CSMD1*) trajectory branches. Dots represent single astrocytes colored by developmental timepoint. **E-F,** Same as **(A-B)**, for 73,638 mouse oligodendrocyte lineage cells. **G,** Integrated UMAP embeddings of 68,485 mouse oligodendrocyte lineage cells colored by **(i)** Palantir-predicted pseudotime, **(ii)** developmental time point, and **(iii)** brain region. **F,** Violin plot (scanpy's default) showing the estimated distribution of pseudotime values for the oligodendrocyte lineage cells in **(E)** grouped by brain region and developmental time point. Color scheme for ages as in **(ii)**.

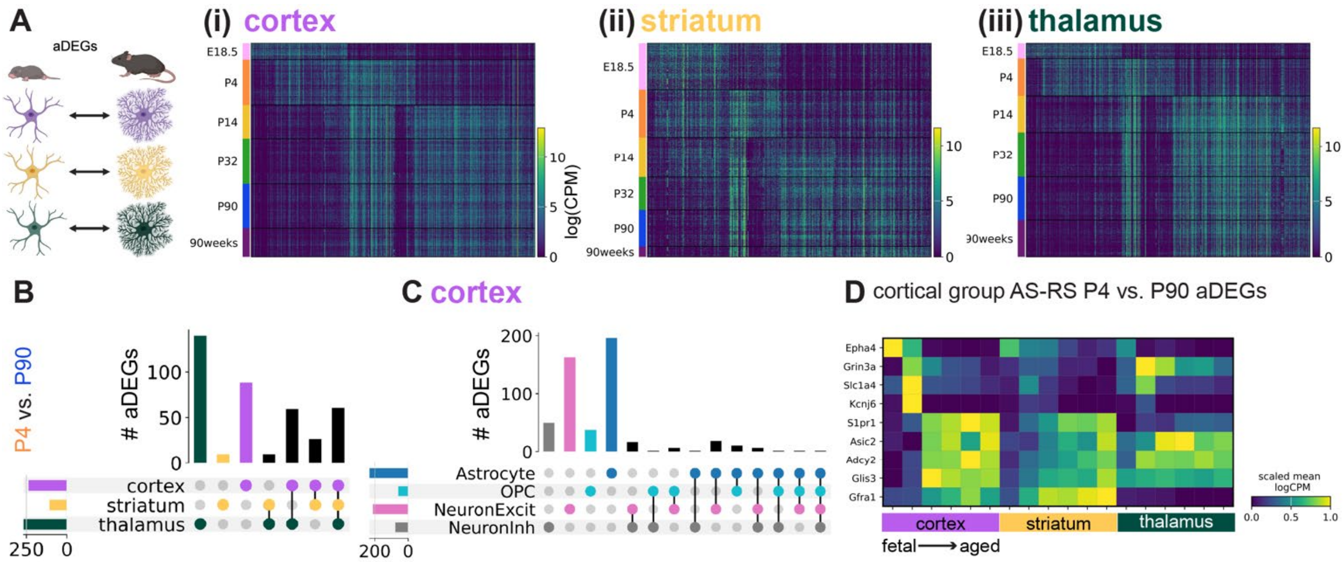


Figure S16. Gene expression signatures underlying the postnatal developmental specification of mouse astrocytes within and across brain regions, related to Fig. 5. A, Heatmaps (rows corresponding to nuclei and columns to gene) showing expression in logCPM of astrocyte age differentially expressed genes (aDEGs) in astrocytes from **(i)** cortex, **(ii)** striatum, and **(iii)** thalamus, grouped by developmental time point as indicated on the left of the heatmap. The strategy for calculating aDEGs is schematized on the left. **B**, UpSet plot showing the number of overlapping P4 vs. P90 astrocyte aDEGs between cortex, striatum, and thalamus. The colored dots below each vertical bar indicate which region(s) share that set of aDEGs, while the colored horizontal bars indicate the total number of cortex-thalamus aDEGs for each region. Overlap categories with 0 aDEGs are not shown. **C**, UpSet plot (as in **(B)**) showing the number of overlapping P4 vs. P90 cortical astrocyte aDEGs between OPCs (light blue), astrocytes (dark blue), excitatory neurons (pink), and inhibitory neurons (gray). **D,** Matrix plot showing mean expression of selected cortex group astrocyte-specific, region-specific (AS-RS) aDEGs (rows) in mouse astrocytes grouped by region and developmental time point (columns, blocked by region first and then by increasing age within each region block). Expression units of mean logCPM are standardized between 0 and 1 by subtracting the minimum and dividing by the maximum for each trait.

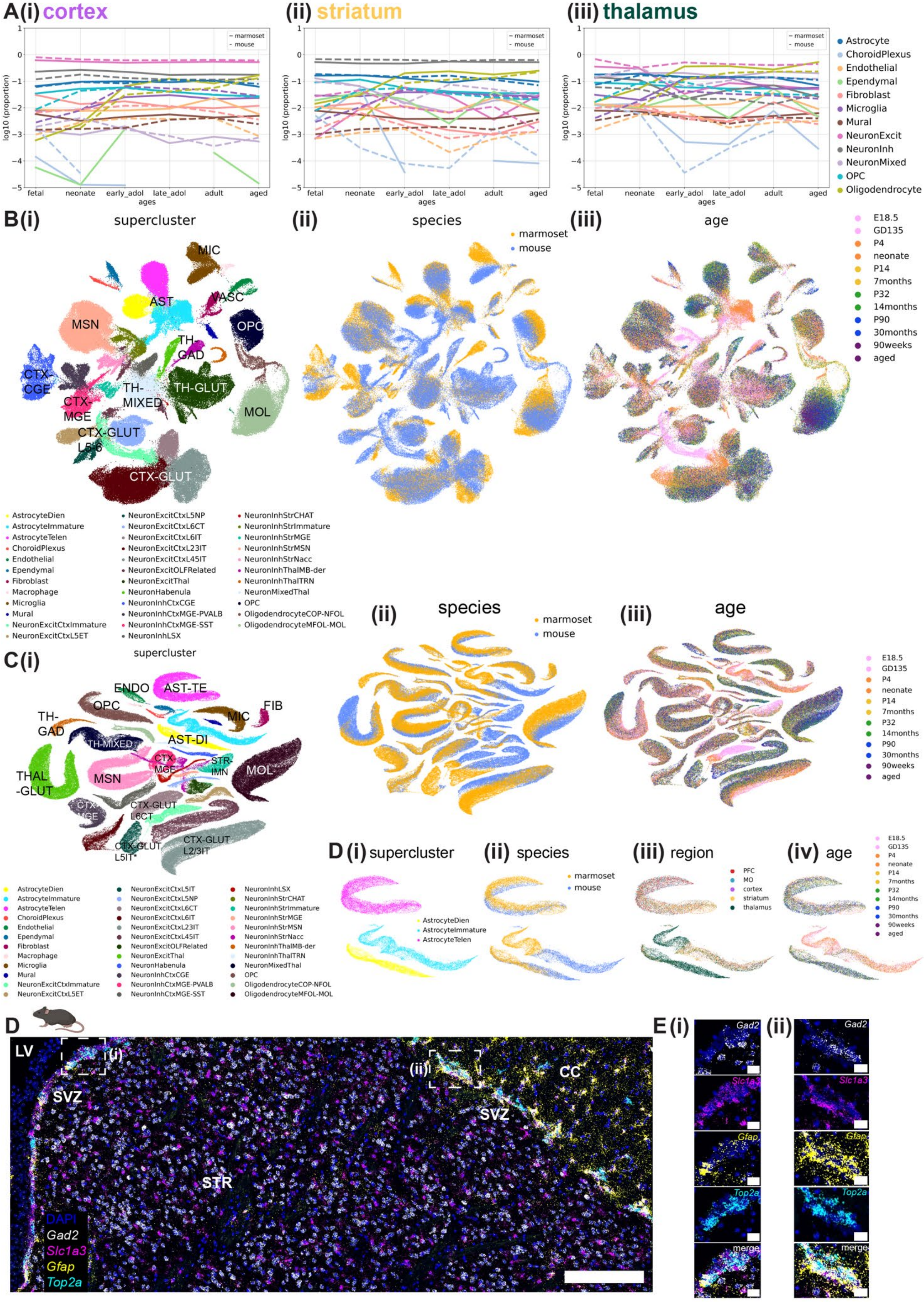

A(i) cortex
(ii) striatum
(iii) thalamus
marmoset
mouse
log10 (proportion)
1
0
-1
-2
-3
-4
-5
fetal
neonate
early_adol
late_adol
adult
aged
ages
Astrocyte
ChoroidPlexus
Endothelial
Ependymal
Fibroblast
Microglia
Mural
NeuronExcit
NeuronInh
NeuronMixed
OPC
Oligodendrocyte
B(i) supercluster
(ii) species
(iii) age
MIC
AST
VASC
OPC
MSN
TH-GAD
CTX-CGE
CTX-MGE
TH-MIXED
TH-GLUT
MOL
CTX-GLUT L5/6
CTX-GLUT
E18.5
GD135
P4
neonate
P14
7months
P32
14months
P90
30months
90weeks
aged
AstrocyteDien
AstrocyteImmature
AstrocyteTelen
ChoroidPlexus
Endothelial
Ependymal
Fibroblast
Macrophage
Microglia
Mural
NeuronExcitCtxImmature
NeuronExcitCtxL5ET
NeuronExcitCtxL5NP
NeuronExcitCtxL6CT
NeuronExcitCtxL6IT
NeuronExcitCtxL23IT
NeuronExcitCtxL45IT
NeuronExcitOLFRelated
NeuronExcitThal
NeuronHabenula
NeuronInhCtxCGE
NeuronInhCtxMGE-PVALB
NeuronInhCtxMGE-SST
NeuronInhLSX
NeuronInhStrCHAT
NeuronInhStrImmature
NeuronInhStrMGE
NeuronInhStrMSN
NeuronInhStrNacc
NeuronInhThalMB-der
NeuronInhThalTRN
NeuronMixedThal
OPC
OligodendrocyteCOP-NFOL
OligodendrocyteMFOL-MOL
C(i) supercluster
(ii) species
(iii) age
ENDO
AST-TE
OPC
FIB
TH-GAD
MIC
TH-MIXED
AST-DI
CTX-MGE
STR-IMN
THAL-GLUT
MSN
MOL
CTX-CGE
CTX-GLUT L6CT
CTX-GLUT L5IT*
CTX-GLUT L2/3IT
NeuronExcitCtxL5IT
D (i) supercluster
(ii) species
(iii) region
(iv) age
PFC
MO
cortex
striatum
thalamus
D
LV
(i)
(ii)
SVZ
CC
SVZ
STR
DAPI
Gad2
Slc1a3
Gfap
Top2a
E (i)
(ii)
Gad2
Slc1a3
Gfap
Top2a
merge

Figure S17. Cell type composition across development in both species and cross-species integration with SATURN, related to Fig. 6. A, Line graphs showing the proportions of major cell classes in the marmoset (solid lines) and mouse (dashed lines) in log10 scale across developmental time points in the **(i)** cortex, **(ii)** striatum, and **(iii)** thalamus. **B,** scANVI-integrated UMAP embeddings of downsampled marmoset and mouse nuclei, colored by **(i)** supercluster, **(ii)** species, and **(iii)** age. **C,** SATURN-integrated UMAP embeddings of downsampled marmoset and mouse nuclei, colored by **(i)** supercluster, **(ii)** species, and **(iii)** age. **D,** SATURN-integrated UMAP embeddings of marmoset and mouse astrocytes, colored by **(i)** supercluster, **(ii)** species, **(iii)** age, and **(iv)** region. **D,** Expression of Top2a in astrocytes of the mouse subventricular zone via multiplexed FISH. Composite image of a registered (across imaging rounds), cropped maximum intensity projection of imaged fields of view in the **(i)** PFC **(ii)** striatum and **(iii)** thalamus in one female P90 mouse. Scale bar, 200μm. **E,** Single-channel and composite images of the boxed regions of the subventricular zone in **(A)**. Scale bar, 20μm. LV, lateral ventricle. SVZ, subventricular zone. CC, corpus callosum. Contrast was manually adjusted by setting minimum and maximum intensity values.

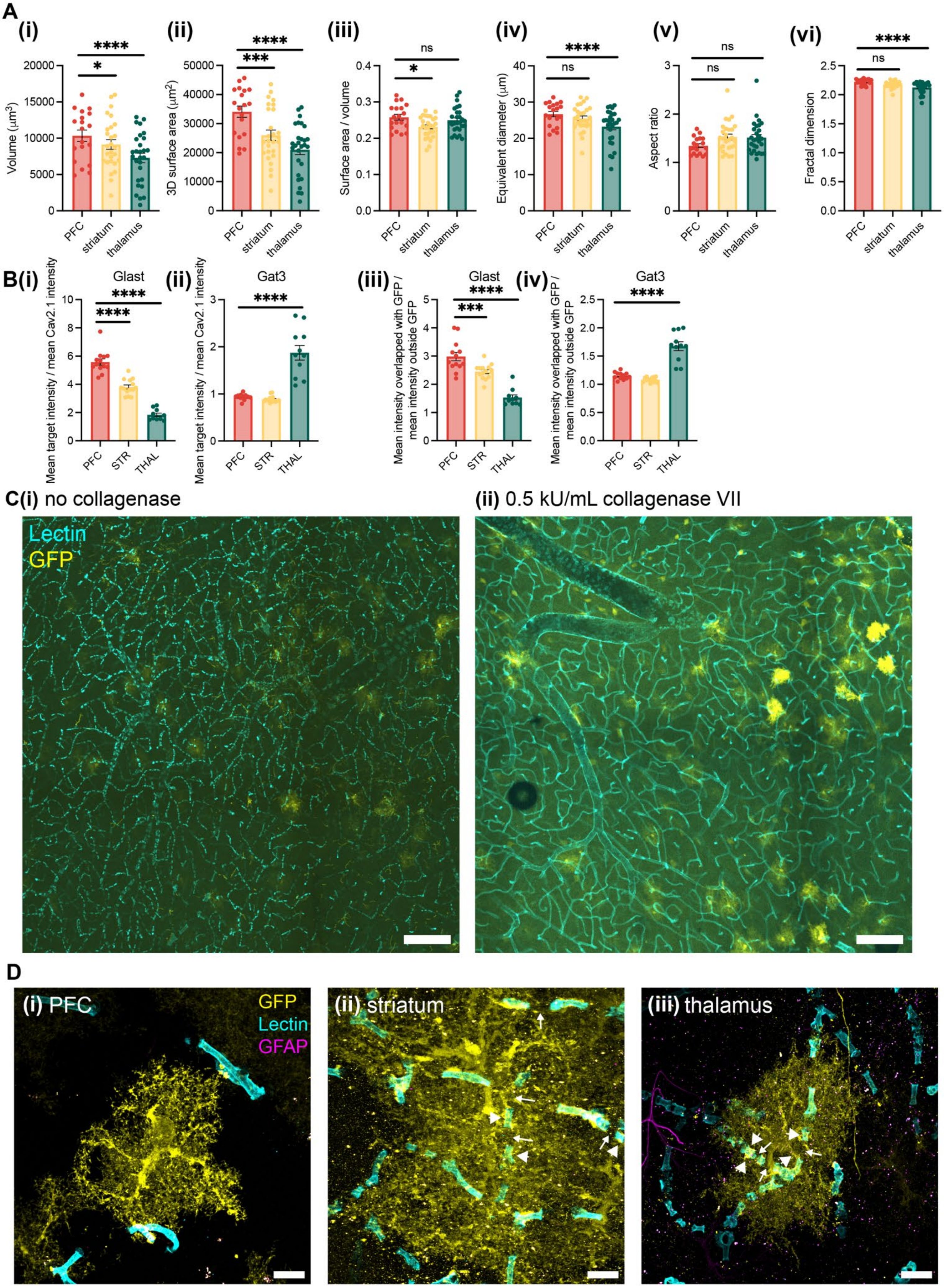

A
(i)
Volume (μm³)
20000
15000
10000
5000
0
PFC
striatum
thalamus
*
****
(ii)
3D surface area (μm²)
50000
40000
30000
20000
10000
0
***
****
(iii)
Surface area / volume
0.4
0.3
0.2
0.1
0.0
*
ns
(iv)
Equivalent diameter (μm)
40
30
20
10
0
ns
****
(v)
Aspect ratio
3
2
1
0
ns
ns
(vi)
Fractal dimension
2.5
2.0
1.5
1.0
0.5
0.0
ns
****
B(i)
Glast
Mean target intensity / mean Cav2.1 intensity
10
8
6
4
2
0
****
****
PFC
STR
THAL
(ii)
Gat3
Mean target intensity / mean Cav2.1 intensity
3
2
1
0
****
(iii)
Glast
Mean intensity overlapped with GFP / mean intensity outside GFP
5
4
3
2
1
0
***
****
(iv)
Gat3
Mean intensity overlapped with GFP / mean intensity outside GFP
2.5
2.0
1.5
1.0
0.5
0.0
****
C(i) no collagenase
Lectin
GFP
(ii) 0.5 kU/mL collagenase VII
D
(i) PFC
GFP
Lectin
GFAP
(ii) striatum
(iii) thalamus

Figure S18. Additional detail on astrocyte morphology and rDEG protein expression analysis with ExR in mouse, related to Fig. 7. A, Quantitative measures of mouse astrocyte morphology across regions (as in **Fig. 7E**) on a subset of astrocytes with no fractured blood vessels in the field of view and complete capture in the axial dimension (n = 19-28 astrocytes from 3 female and 5 male mice for each region, with statistical significance determined using a linear mixed effects model with “animal” as the random effect group variable). **B,** Bar plots showing **(i, ii)** quantified normalized mean intensity (to Cav2.1 synaptic reference channel) within GFP+ regions and **(iii, iv)** enrichment ratio (mean intensity in GFP+ regions divided by mean intensity outside GFP+ regions) of either **(i, iii)** Glast or **(ii, iv)** Gat3 (from ~18x ExR data as in **Fig. 7F**, n = 12-19 fields of view from 2 mice, with statistical significance determined using a linear mixed effect model as described in the **Methods**). In **A** and **B**, data points are individual fields of view, and error bars are standard error of the mean. **C-D,** High-concentration collagenase treatment preserves blood vessel morphology in ExR samples. **C, (i)** Maximum intensity projection cropped field of view obtained at 4x magnification of ~4x expanded tissue from an Aldh1l1-Cre (Jackson Laboratories # 023748, a gift from Dr. Mriganka Sur) mouse injected at P1-P2 with PhP.eB CAG-FLEX-GFP and perfused with ExR fixative after 5 weeks. Tissue was sectioned at 80μm, pre-stained for GFP and processed according to the original ExR protocol without collagenase treatment[7]. Blood vessels (cyan, Lectin stain) appear fragmented. This animal is from a litter of mice from a separate experiment whose data was not included in the current paper. **(ii),** Same as **(i)**, for a mouse from a separate experiment whose data was not included in the current paper, sectioned at 150μm, pre-stained for GFP, and treated overnight with collagenase VII at 0.5kU/mL (see **Methods**). Blood vessels appear continuous. Both fields of view are from the thalamus. Scale bar, 420μm in physical units (~100-120μm in biological units, assuming a 3.5-4.2x expansion factor; the expansion factor was lower in the experiment in **(i)**). **D,** Maximum intensity projections of 40x magnification astrocyte fields of view in the **(i)** PFC, **(ii)** striatum, and **(iii)** thalamus from the experiment shown in **(A)(i)**. Fields of view in **(i)** and **(ii)** are from one male mouse, while **(iii)** is from a female mouse. Contrast was manually adjusted in Fiji at or above 35% saturation to show similar GFP brightness across regions. Despite fragmented blood vessels, astrocyte morphology appears almost entirely continuous (compare to **Fig. 7**), except at end feet directly touching blood vessels. Smaller white arrows indicate breaks within blood vessels where astrocyte end feet may exhibit possible local distortion, while larger white arrowheads indicate unbroken blood vessels, where astrocyte processes appear continuous and in good registration with the vessel. Scale bar, 35μm in physical units (~10μm in biological units, assuming a 3.5x expansion factor).

References

1. Krienen, F.M., Levandowski, K.M., Zaniewski, H., del Rosario, R.C.H., Schroeder, M.E., Goldman, M., Wienisch, M., Lutservitz, A., Beja-Glasser, V.F., Chen, C., et al. (2023). A marmoset brain cell census reveals regional specialization of cellular identities. Science Advances *9*, eadk3986. https://doi.org/10.1126/sciadv.adk3986.

2. Büttner, M., Ostner, J., Müller, C.L., Theis, F.J., and Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. Nat Commun *12*, 6876. https://doi.org/10.1038/s41467-021-27150-6.

3. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature *624*, 317–332. https://doi.org/10.1038/s41586-023-06812-z.

4. Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68. https://doi.org/10.1038/s41586-019-1506-7.

5. Chai, H., Diaz-Castro, B., Shigetomi, E., Monte, E., Octeau, J.C., Yu, X., Cohn, W., Rajendran, P.S., Vondriska, T.M., Whitelegge, J.P., et al. (2017). Neural Circuit-Specialized Astrocytes: Transcriptomic, Proteomic, Morphological, and Functional Evidence. Neuron *95*, 531-549.e9. https://doi.org/10.1016/j.neuron.2017.06.029.

6. Morel, L., Chiang, M.S.R., Higashimori, H., Shoneye, T., Iyer, L.K., Yelick, J., Tai, A., and Yang, Y. (2017). Molecular and Functional Properties of Regional Astrocytes in the Adult Brain. Journal of Neuroscience *37*, 8706–8717. https://doi.org/10.1523/JNEUROSCI.3956-16.2017.

7. Sarkar, D., Kang, J., Wassie, A.T., Schroeder, M.E., Peng, Z., Tarr, T.B., Tang, A.-H., Niederst, E.D., Young, J.Z., Su, H., et al. (2022). Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy. Nat. Biomed. Eng *6*, 1057–1073. https://doi.org/10.1038/s41551-022-00912-3.