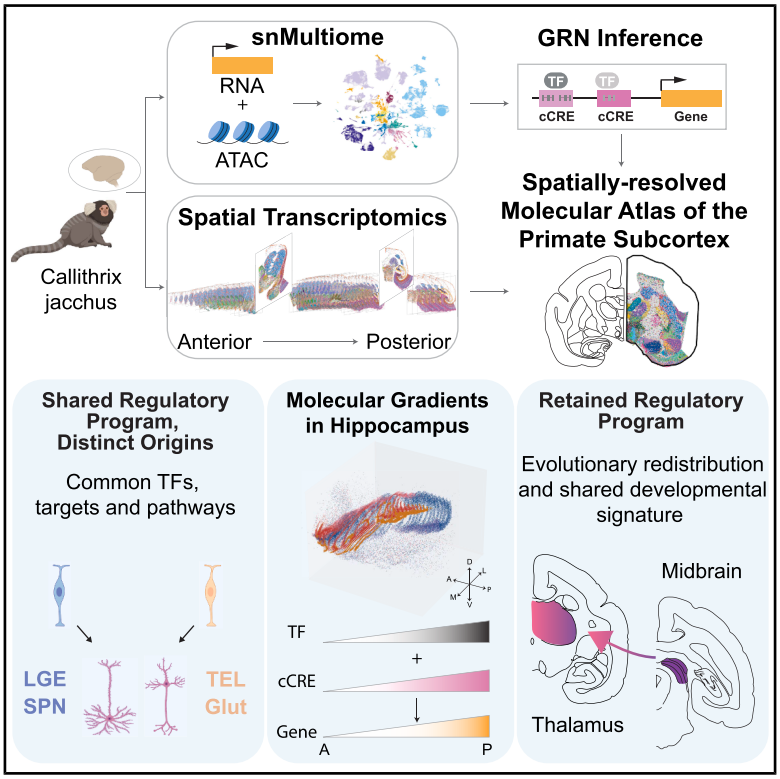
Cell

Spatial and regulatory organization of cell identity across the primate subcortex

Graphical abstract



Highlights

- A marmoset subcortical atlas couples snRNA-seq, snATAC-seq, and spatial transcriptomics
- Graded TF and enhancer activity shape transcriptomic gradients in hippocampal subfields
- Striatal SPNs and cortical glutamatergic neurons converge on a shared regulatory program
- Enhancers carry both cell-type- and region-restricted fingerprints across the subcortex

Authors

Shu Dan, Meghan A. Turner, Michael DeBerardine, ..., Guoping Feng, Brian Long, Fenna M. Krienen

Correspondence

brianl@alleninstitute.org (B.L.), fkrienen@princeton.edu (F.M.K.)

In brief

Dense multimodal profiling of the marmoset subcortex reveals how spatial context and gene regulatory programs jointly shape cell identity across the primate brain. By linking transcriptional states, chromatin accessibility, and anatomy, this atlas uncovers regulatory principles underlying spatial gradients, neuronal convergence, and evolutionary specialization.

Dan et al., 2026, Cell *189*, 7471–7491
November 12, 2026 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2026.10.007

CellPress

Cell

CellPress
OPEN ACCESS

Article

Spatial and regulatory organization of cell identity across the primate subcortex

Shu Dan,[1,8] Meghan A. Turner,[2,8] Michael DeBerardine,[1,8] Lakme Caceres,[1] Victor E. Nieto-Caballero,[1] Jessica Schembri,[1] Kirsten Levandowski,[3,4] Christopher Cardenas,[1] Ning Wang,[1] Chang Li,[5] Soumyadeep Basu,[5,6] Hongyu Zheng,[7] Xinhao Liu,[7] Thomas Höllt,[6] Jessica Feliciano,[1] Qiangge Zhang,[3,4] Reilly Nakamoto,[1] Delissa A. McMillen,[2] Naomi Martin,[2] Nasmil Valera Cuevas,[2] Paul Olsen,[2] Josh Nagra,[2] Jazmin Campos,[2] Marshall M. VanNess,[2] Jack Waters,[2] Shea Ransford,[2] Zoe Juneau,[2] Sam Hastings,[2] Stuard Barta,[2] Augustin Ruiz,[2] Jeanelle Ariza,[2] Benjamin J. Raphael,[7] Hongkui Zeng,[2] Ed S. Lein,[2] Boudewijn Lelieveldt,[5,6] Guoping Feng,[3,4] Brian Long,[2,9,*] and Fenna M. Krienen[1,9,10,*]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
[2]Allen Institute, Brain Science, Seattle, WA, USA
[3]Yang Tan Collective and Govern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[5]Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands
[6]Faculty of EEMCS, Delft University of Technology, Delft, the Netherlands
[7]Department of Computer Science, Princeton University, Princeton, NJ, USA
[8]These authors contributed equally
[9]These authors contributed equally
[10]Lead contact
*Correspondence: brianl@alleninstitute.org (B.L.), fkrienen@princeton.edu (F.M.K.)
https://doi.org/10.1016/j.cell.2026.10.007

SUMMARY

Mammalian brain cell identity is shaped by intrinsic factors and external context. We present a spatially resolved transcriptomic and gene regulatory atlas of cell types across all subcortical regions in a primate, the common marmoset. Dense sampling and cross-species integration revealed spatially precise neuronal assemblies, including ones in complex midbrain and diencephalic structures. Chromatin accessibility and transcriptional identity are spatially tuned within and across subcortical structures; spatial gradients within hippocampal subfields are orchestrated by graded transcription factors acting through graded enhancers. The primate-expanded thalamic γ-aminobutyric acid (GABAergic) population shares transcriptional and regulatory syntax with conserved midbrain populations, reflecting an evolutionary adaptation compared with rodents. Similar regional expression across cell types can arise by distinct regulatory architectures, as observed in telencephalic astrocytes and neurons. Conversely, distant cell types can share regulatory programs despite divergent identities: striatal GABAergic spiny projection neurons and telencephalic glutamatergic neurons share a postsynaptic regulatory program despite divergent lineages, regions, and neurotransmitter identities.

INTRODUCTION

Single-cell genomics studies have transformed our understanding of cellular diversity; the brain is a particularly fertile ground for these efforts. Recent data indicate that thousands of transcriptomically distinct cell populations exist in the adult mouse brain,[1,2] which has the most comprehensively characterized brain cell taxonomy of any mammal. Although the mouse remains foundational for understanding the mammalian brain, the extent to which mouse cell types and their spatial organization generalize to primates remains an open question.

The human brain is ~3,000-fold larger by mass than the mouse brain and contains more than 86 billion neurons. This scale makes comprehensive study a formidable challenge. The most extensive human brain atlas to date[3] used single-nucleus RNA sequencing (snRNA-seq) to sample over 100 anatomical locations in the adult human postmortem brain, but anatomical assignments are relatively coarse and based on microdissection labels. More generally, the size of the human brain precludes complete spatial transcriptomic coverage with current methods. These limitations obscure rare cell types, the precise anatomical localization of individual populations, and complex spatial

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

topographies, such as gradients and discontinuities. In turn, incomplete sampling complicates cross-species comparisons, particularly as cell migration is a driver of evolutionary change.[4–7]

The marmoset *Callithrix jacchus* presents a compelling solution to the issue of scale. The marmoset is a New World monkey whose brain weighs only 8 g, which is comparable to that of a rat.[8,9] Despite its evolutionary distance to humans (35–40 million years from great apes), the adult marmoset brain retains primate-specific anatomical,[10–14] connectivity,[15–17] and cell type specializations.[18–21] With its compact brain and complete anatomical accessibility via current single-cell and spatial transcriptomic platforms, the marmoset is uniquely suited to fill the gaps in our understanding between mouse and primate cell type diversity.

Although transcriptomic atlases of the mouse brain are increasingly detailed, the gene regulatory architecture supporting this diversity remains largely unexplored at a brain-wide scale. Single-region studies link chromatin accessibility and *cis*-regulatory elements (CREs) to cell identity and spatial context,[22–26] but no spatially resolved regulatory atlas spans numerous brain regions in a mammalian species. Such an atlas would inform how cell lineage and regional context shape cell identity and the regulatory logic of spatial gradients, expand the search space for enhancer-based genetic tools across species,[20,27–31] and provide a foundation for interpreting noncoding disease variants.

Here, as part of the BRAIN Initiative Cell Atlas Network (BICAN, RRID: SCR_022794), we generated a hierarchical cell taxonomy based on anatomically comprehensive multi-omic (snRNA-seq and single-nucleus assay for transposase-accessible chromatin using sequencing [snATAC-seq]) sampling of the marmoset subcortex. In parallel, we performed spatial transcriptomic profiling of a marmoset cerebral hemisphere, enabling the anatomical localization of cell populations at each level of the taxonomy. All three modalities can be explored interactively through our Cytosplore Viewer (https://viewer.cytosplore.org/). With this atlas, we characterize gene regulatory networks (GRNs) in both neurons and non-neurons that distinguish cell types and brain regions, dissect convergent and divergent regulatory programs across cell types, characterize spatial gene expression gradients in the hippocampal formation (HiF), and investigate the evolutionary origins of primate-expanded thalamic γ-aminobutyric acid (GABAergic) neurons. Our study analyzes the complete subcortex in the marmoset; companion BICAN papers focus on basal ganglia (BG) cell types across human and nonhuman primates.[32,33]

RESULTS

Generation of the MSCA

The marmoset subcortical cell atlas (MSCA) encompasses the non-isocortical telencephalon, including the HiF, olfactory areas, and cerebral nuclei, as well as the diencephalon (thalamus [THM] and hypothalamus [HTH]), midbrain, and major hindbrain structures (excluding the cerebellar cortex). We integrated the MSCA with existing human, nonhuman primate, and mouse brain atlases to support interoperable annotations and cross-species comparisons. MSCA data that mapped to BG structures were also used in the companion study describing a spatial atlas of marmoset, macaque, and human BG,[33] as well as in a consensus cross-species, multi-omic BG taxonomy.[32]

An anatomically comprehensive cell type taxonomy of the marmoset subcortex

The entire subcortex of four marmoset donors was tiled into contiguous dissections (Figure 1A; Table S1; STAR Methods). Dissociated nuclei from each tile were fluorescence-activated nuclei sorting (FANS)-sorted to enrich for neurons and deplete the oligodendrocyte lineage. Sorted fractions were repooled to target proportions (70% NeuN+, 10% Olig2+, and 20% NeuN−/Olig2−), then profiled by 10x Multiome. In total, we sequenced 1,299,000 nuclei from 76 unique tiles across the four donors, targeting a sequencing depth of 120,000 reads/cell.

After stringent quality control (STAR Methods; Figure S1), 678,917 nuclei remained. Batch effects were corrected by scVI[34] across donors, and iterative clustering[35] in the donor-aligned latent space resolved 1,737 transcriptomic "clusters." We organized clusters into a three-level hierarchy by aggregating transcriptomically and spatially similar clusters into 212 groups and 17 classes (Figures 1A and 1B). Groups are generally named for the anatomical structures where most cells are localized. Clusters belonging to BG structures inherited their class and group labels from the consensus cross-species BG taxonomy.[32]

We hierarchically clustered group-level pseudobulked "metacells" using the top 4,000 highly variable genes (HVGs) (STAR Methods; Figure 1B). Within neurons, clades of excitatory telencephalic (cortex [CTX], amygdala [AMY], and HiF) groups align with major anatomical boundaries. Notably, the striatal spiny projection neurons (SPNs), although inhibitory, clustered with excitatory telencephalic groups rather than other telencephalic GABAergic populations, consistent with prior mouse and primate reports.[20,36,37]

In the diencephalon, excitatory THM groups formed a distinct clade; GABAergic thalamic groups were more like GABAergic groups from other structures. By contrast, hypothalamic groups intermingled regardless of neurotransmitter identity. Other major branches reflected neurotransmitter (e.g., dopaminergic and cholinergic groups) or anatomical (e.g., lateral septum [LSX] and superior colliculus [SC]) identity.

Spatially resolving subcortical cell type diversity and mapping performance

Spatial transcriptomic data (10x Xenium) were collected from one hemisphere at ∼200 μm section spacing across three 5-mm slabs (STAR Methods; Figure 1C). We designed a 300-gene panel (Table S2) from previously described cell-type markers, differentially expressed genes within the caudate and putamen,[12,38] and variable genes from prior marmoset BG and thalamic snRNA-seq.[20] Spatial cells were mapped to the MSCA taxonomy using correlation-based mapping (MapMyCells,[39] RRID: SCR_024672) at the cluster level, with group and class labels inherited.

To evaluate panel performance, we self-mapped the multi-omic RNA back to their own taxonomy using only the 300 spatial-panel genes (STAR Methods; Figure S2). This recovered the "dominant region" for 97.5% of neurons in localized clusters (93% of all taxonomy neurons) (Figure S2C). The panel

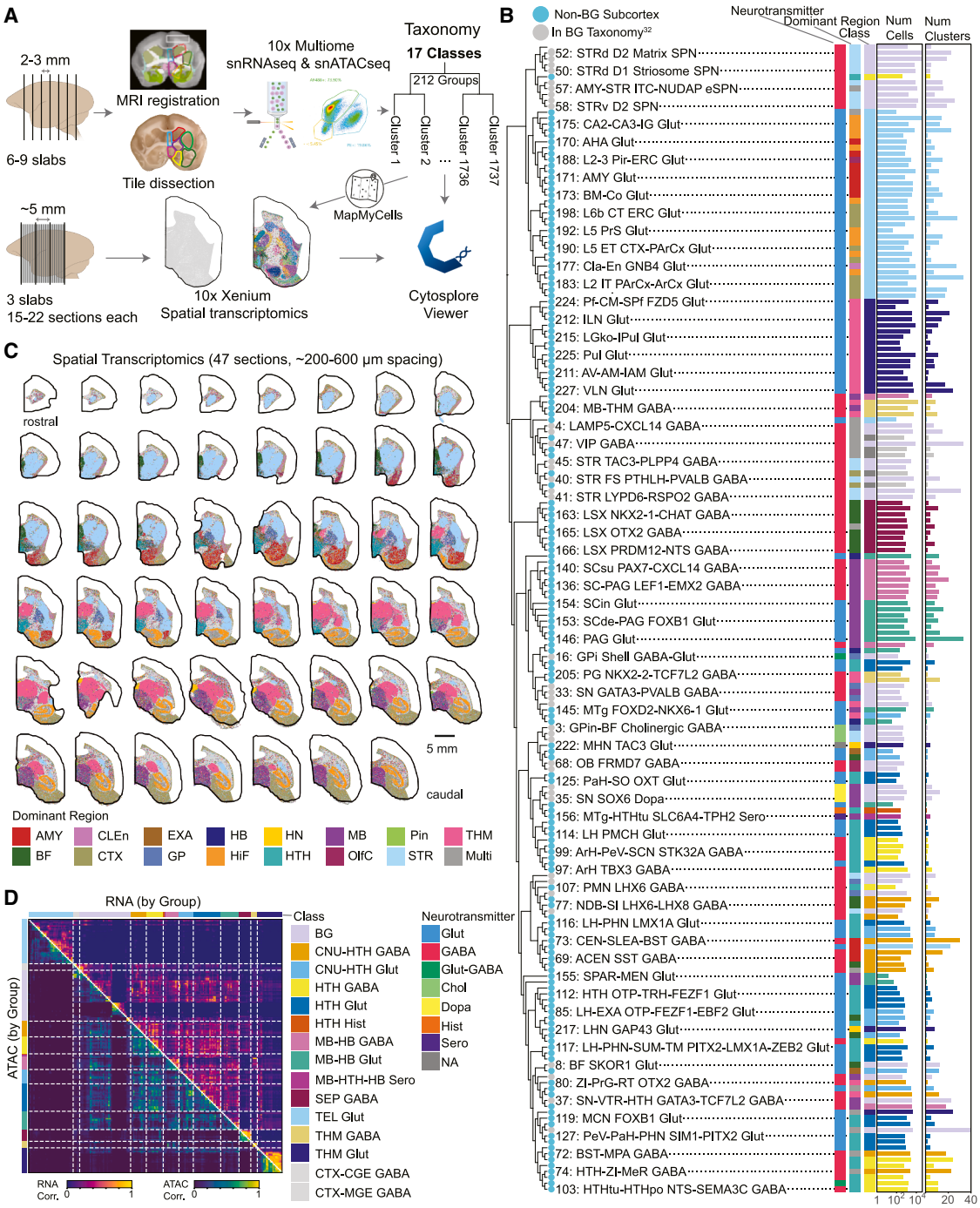


Figure 1. Spatial, transcriptomic, and regulatory atlas of the marmoset subcortex

(A) Schematic of sampling and analysis workflows. FANS-sorted nuclei from MRI-guided dissection tiles were profiled by 10x Multiome, yielding a three-level taxonomy (class, group, and cluster). Xenium spatial transcriptomics data were mapped to the taxonomy to obtain spatially resolved clusters. Cytosplore Viewer links multi-omics and spatial transcriptomics for interactive data exploration.

(B) A total of 194 neuronal transcriptomic groups hierarchically clustered by top 4,000 HVG expression (names labeled every third branch). Companion cross-species consensus BG groups[32] are indicated. Color tiles (left to right) show the neurotransmitter, dominant region, and class. Bar plots indicate cell and cluster counts.

(C) Forty-seven Xenium spatial sections (rostral to caudal); cells are colored by the dominant region of their assigned cluster. Scale bar, 5 mm.

(D) Pairwise Spearman correlation of transcriptomic and chromatin accessibility similarity between all groups, ordered by class (Figure S8).

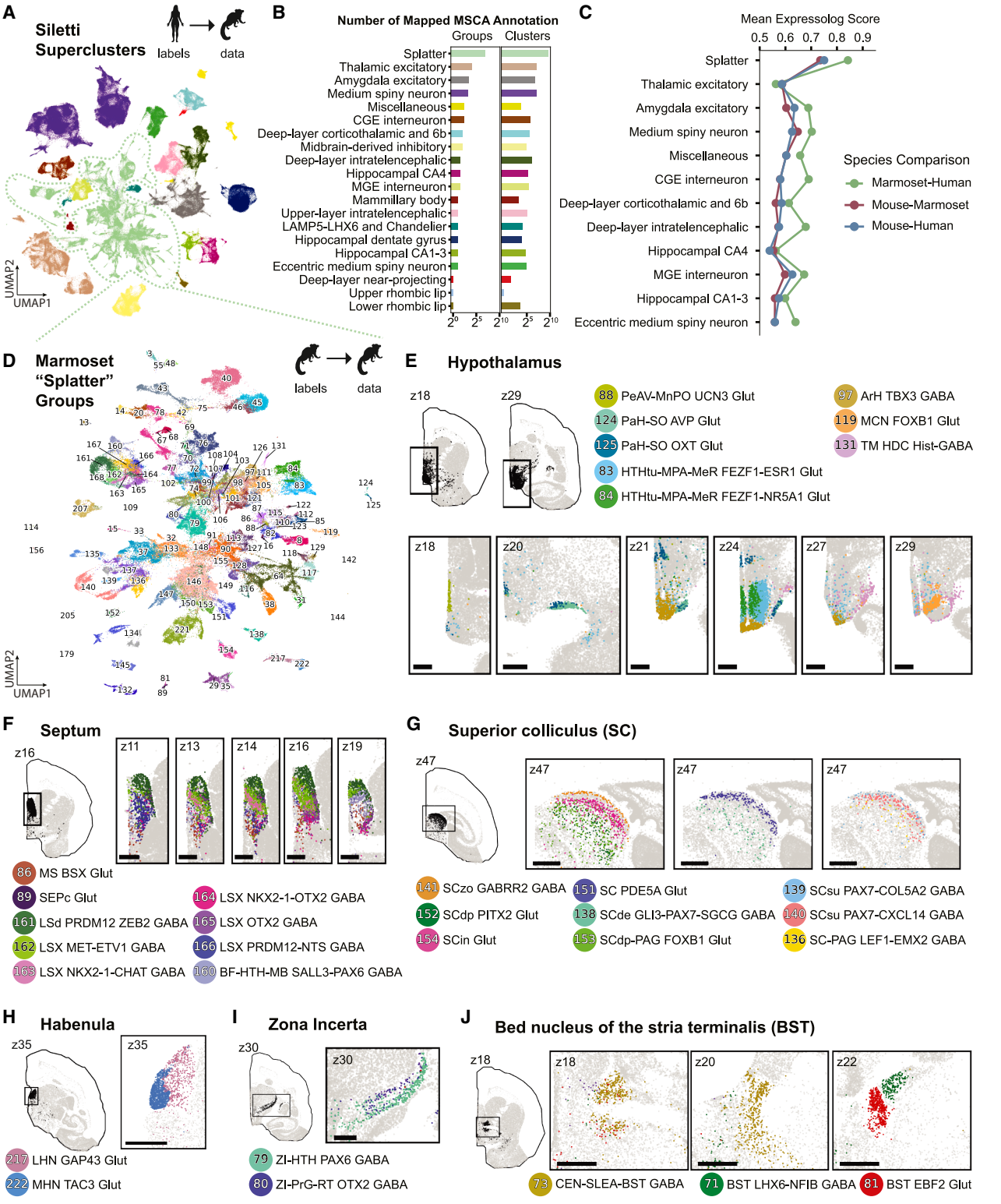


Figure 2. Cross-species spatial mapping of heterogeneous neuron populations

(A) UMAP of marmoset subcortical neurons colored by predicted human supercluster identity from Siletti et al.[3]

(B) Number of MSCA groups and clusters mapped per human supercluster.

(legend continued on next page)

performed well for its size and was balanced across the taxonomy (Figures S2D–S2F).

In this self-mapping, ~99% of cells received a correct class assignment and >90% of cells received a correct group assignment (Figures S2G–S2I). Cluster-level accuracy varied across the taxonomy (60% overall, with confusions concentrated among transcriptomically similar neighbors in the same structure) (Figure S2J; Table S3). After pseudobulking the spatial data by assigned cluster, 71.5% of clusters were most correlated with their multi-omic counterpart, and 87.0% fell within the top two matches. Class- and group-level cell abundances were broadly consistent between modalities (Figures S2K and S2L). Per-cluster and per-group performance is reported in Table S3.

Spatial mapping reveals anatomically coherent cell type organization

Spatial mapping revealed detailed localization of neuronal and non-neuronal populations and allowed for the anatomical annotation of transcriptomic groups. Within the AMY, groups mapped to individual anatomical nuclei via discrete markers, and clusters resolved finer positions through combinatorial and graded expression (Figure S3). Comparable anatomical specificity was observed in thalamic glutamatergic populations (Figure S4) and non-neuronal populations of neuroglia and mesodermal lineages (Figure S5).

Integrating cross-species references into marmoset subcortical cell type annotations

To further support our cell type annotations, we cross-referenced our taxonomy against additional publicly available mouse[1] and human[3] whole-brain atlases through MapMyCells[39] (STAR Methods; Tables S4 and S5). We also integrated regional atlases for the telencephalon,[20] AMY,[40] HTH,[41,42] LSX,[43] HiF,[44] and lateral geniculate nucleus (LG)[45] using scVI[34] (STAR Methods; Figures S6A and S6B; Tables S4 and S5). Despite species and assay modality differences, MSCA cell type divisions and spatial annotations were largely concordant with the transcriptomic groupings and anatomical dissections in these references.

Companion BICAN papers use MSCA BG cells in a cross-species, spatial BG atlas[33] and a consensus cross-species, multi-omic BG taxonomy.[32] The contiguous sampling of the MSCA beyond BG structures reveals the relative similarities of cell populations within and between BG and non-BG structures (Figure 1B). For example, MSCA mapping shows that the evolutionarily divergent striatal *TAC3*+ interneuron population[4,5,18,20,46] includes subtypes distributed outside the striatum—in the septum, ventral pallidum, and HTH—indicating greater molecular and anatomical diversity than previously appreciated (Figure S6C).

Chromatin accessibility recapitulates the transcriptional hierarchy

Gene expression is controlled by a vast diversity of CREs (including enhancers). Millions of putative enhancers decorate mammalian genomes[47,48]; any one gene may have dozens of enhancers, and a given cell type may have thousands of enhancers that are differentially accessible compared to other cell types. From group-level ATAC peaks (SnapATAC2[49]; STAR Methods), we identified ~462,000 CREs robustly accessible in at least one group. Topic modeling (pycisTopic[50]) grouped peaks into modules and yielded a low-dimensional accessibility embedding. The ATAC-based embedding closely mirrored the RNA-based embedding, suggesting that the regulatory syntax captured by the topics recapitulates the transcriptional hierarchy (Figures S7A and S7B). Group-level marker genes showed concomitant open chromatin along their gene bodies in the ATAC data (Figure S7C), indicating that group-level differences were preserved.

Hierarchical clustering of pycisTopic ATAC loadings across groups (Figure 1D, bottom triangle; Figure S8) closely matched group-level RNA correlations (Figure 1D, top triangle; Figure S8) (Mantel test with 5,000 permutations, $r = 0.88$, $p < 0.001$). This indicates that the gene regulatory and gene expression landscapes converge on a globally concordant view of cell identity (Figure S7D).

Cross-species mapping resolves heterogeneous subcortical populations

We leveraged the MSCA to provide greater resolution to the most anatomically complete human brain cell taxonomy,[3] which sampled 100 locations by snRNA-seq but lacked precise spatial information. The human cell taxonomy contains 31 superclusters, 461 clusters, and 3,313 subclusters. We used MapMyCells to transfer human supercluster and cluster labels onto marmoset cells (Figure 2A; STAR Methods). On average, 98% of cells in each MSCA cluster mapped to the same supercluster. Homologous MSCA groups were identified for each of the 20 neuronal superclusters (excluding the cerebellar inhibitory supercluster due to the lack of cerebellar sampling) (Figure S9). The "*Splatter*" supercluster mapped to the most MSCA groups ($n = 117$), clusters ($n = 974$), and cells ($n = 126{,}941$) (Figure 2B; Figure S9), spanning nearly a dozen anatomical regions (Figure S10 inset).

To test conservation across mammals, we defined homologous neuronal cell types as mouse[1] or marmoset clusters with >70% of cells mapped to a human cluster, yielding 114/382 human clusters with conserved counterparts in both species. Pairwise cross-species expression similarity, quantified by expressolog,[51] confirmed that marmoset cells are more similar to human cells than to mouse cells (Figure 2C). Notably, the *Splatter* supercluster is both the most broadly mapped supercluster and the most conserved across all species pairs, despite its transcriptional heterogeneity.

The *Splatter* supercluster was named for its complex topography in low-dimensional embeddings.[3] Siletti et al.[3] attributed this to neurons from diverse regions (basal forebrain, HTH, THM, midbrain, pons, and medulla) clustering together due to

(C) Mean expressolog scores of homologous clusters across human, marmoset, and mouse within each human supercluster; higher scores indicate stronger transcriptomic similarity.

(D) UMAP of marmoset neurons mapped to the *Splatter* supercluster, colored by MSCA group; group IDs are overlaid (Figure S10 and Table S4 for full names).

(E–J) Representative spatial z-planes showing MSCA group localization; whole-hemisphere panels provide anatomical context. Scale bars, 1 mm. (E) HTH groups. (F) Medial and LSX groups. (G) Layers of the SC. (H) MH and LH. (I) ZI subdivisions. (J) BST subdivisions.

undersampling; our dense sampling and spatial data now resolve their anatomical origins (Figure 2D; Figure S10). One of the largest *Splatter* subpopulations localized to the marmoset HTH (Figure 2E), which contains many small and functionally distinct neuronal populations.[41,42,52,53] Consistent with the integration of mouse and human "Hypomap" atlases,[41,42] many marmoset groups are localized to specific hypothalamic nuclei, including the arcuate nucleus (ArH) and magnocellular nucleus (MCN) (Figure 2E; Figure S6B). Other groups resolved subdivisions within nuclei, such as the distinct medial and lateral groups within the ventromedial HTH of the tuberal nucleus (HTHtu) (Figure 2E). The mammillary body (MM)—its own supercluster in the Siletti human atlas—is subdivided by marmoset clusters into six spatially distinct partitions along the medial-lateral and rostral-caudal axes (Figure S11A).

Where regional reference annotations were sparse (basal forebrain and midbrain), we leveraged the spatial transcriptomics data directly. In the basal forebrain, the septum is patterned by seven GABAergic and two glutamatergic groups, anatomically anchoring homologous cell types from an integrated human septum atlas[43] (Figure 2E; Figure S6B). Groups mapping to the midbrain SC highlighted its internal laminar organization (Figure 2G). Distinct groups partitioned the medial (MH) and lateral habenula (LH) (Figure 2H) and the zona incerta (ZI) (Figure 2I); in the extended AMY, three groups delineated the rostral-caudal axis of the bed nucleus of the stria terminalis (BST) (Figure 2J). Further spatial subdivisions are present at the cluster level within the LH/MH (Figure S11B), ZI (Figure S11C), and BST (Figure S11D).

Collectively, these mappings resolve the human *Splatter* supercluster into hundreds of anatomically discrete, evolutionarily conserved populations distributed across the HTH, basal forebrain, midbrain, habenula, ZI, and extended AMY. It reframes what had appeared as a single heterogeneous group in human data[3] as a spatially organized collection of conserved subcortical cell types.

Gene regulatory programs across the primate subcortex

A key benefit of paired RNA and ATAC multi-omic data is the ability to identify enhancers that link transcription factors (TFs) to target genes, forming GRNs. We inferred GRNs using SCENIC+,[54] which integrates chromatin accessibility, gene expression, and transcription factor binding site (TFBS) motif analysis to identify eRegulons linking TFs to binding sites and target genes (Figure 3A). ATAC peaks within eRegulons are candidate CREs (cCREs), predominantly enhancers with fewer repressors. We identified 550 neuronal and 299 non-neuronal eRegulons (STAR Methods; Table S6).

Conserved transcription factor OTX2 illustrates cell-type-specific regulatory logic

Although some GRNs show broad activity, most show cell-type specificity (Figure 3B; FigureS12A). One striking case is the *OTX2* eRegulon, which is highly active in the "THM GABA" group (Figure 3B). *Otx2* is a developmental regulator of the telencephalon, diencephalon, and mesencephalon and specifies the midbrain-hindbrain boundary.[55–57] In neurons, *OTX2* has 142 putative direct targets, including *GRIK1* (GluK1 kainate receptor subunit) (Figure 3C). SCENIC+ predicts *OTX2* activity at multiple peaks associated with *GRIK1* expression, including an intronic peak with a consensus *OTX2* motif whose accessibility correlates with both *OTX2* and *GRIK1* expression (Figure 3D).

Projection of eRegulon activity onto spatial data confirmed that *GRIK1* and other neuronal *OTX2* targets are concentrated in the THM (Figure 3E). Other groups with high *OTX2* expression include midbrain GABAergic types (especially in the SC) as well as ependymal glia and tanycytes (Figure 3F), paralleling the adult mouse brain.[1] We also detected a non-neuronal *OTX2* eRegulon, with 69 additional targets. Many *OTX2* targets show cell-type-divergent expression, including the TFs *DMBX1* and *GATA3* (neuron-specific) and *LMX1B* and *PDLIM5* (non-neuron-specific) (Figure 3G).

How can *OTX2* be expressed in two cell populations yet have different targets in each? We hypothesized that cell-type-specific co-TFs may be acting with *OTX2* at cell-type-specific enhancers. cCREs for cell-type-specific targets show cell-type-specific accessibility (Figures S12B and S12C), consistent with co-TF-gated accessibility. Searching for TFs co-enriched in OTX2 cCREs ($n = 152$ neuronal; $n = 46$ non-neuronal) identified strong DMBX1 and GATA3 enrichment in neurons ($n = 36$ and $n = 20$ shared gene targets, respectively; Figure S12D). DMBX1 and OTX2 share a motif in their shared peaks, and they physically interact,[58] supporting DMBX1 as a potential neuronal co-factor—though at only a minority of neuronal OTX2 sites. In non-neurons, *NFIA* was most co-enriched, albeit with few shared targets ($n = 4$).

Convergent regional expression in neurons and astrocytes arises through distinct regulatory programs

The OTX2 example shows that a single TF can direct distinct targets across cell types. We next examined the converse: can the same regionally biased gene be expressed convergently in different cell types through distinct regulatory architectures? Astrocytes and neurons provide a natural test bed because both are regionally heterogeneous in the mammalian brain.[59–61] We identified three major astrocyte subdivisions: the telencephalic astrocytes (Astro-TE NN), non-telencephalic astrocytes (Astro-NT NN), and white-matter astrocytes (Astro-WM NN) groups (Figure 4A; Figure S12E), paralleling their regional identities reported in mouse[1] and the primate-conserved gray matter striatum (GM-STR), gray matter excluding striatum (GM-exSTR), and white matter (WM) subgroups in Fu et al.[62] Principal component analysis (PCA) on pseudobulked astrocyte clusters (Figure S12F) separated the Astro-NT NN clusters from Astro-TE NN and Astro-WM NN clusters along the first principal component (PC1). Top PC1 genes included *FOXG1*—a known telencephalic regulator in both astrocytes and neurons[63,64]—and *MOXD1*, a copper-dependent monooxygenase, which was enriched in Astro-TE NN and Astro-WM NN groups (Figure 4B). *MOXD1* showed the same regional signature in neurons: enriched in telencephalic neurons, particularly in the AMY and septum (though at lower levels than in Astro-TE and Astro-WM), and largely absent from non-telencephalic neurons (Figure 4C).

MOXD1 was linked to distinct regulatory programs in neurons and astrocytes. In astrocytes, two eRegulons driven by *NR2F1* and *POU3F2* were active across all astrocyte groups, with cell-type specificity achieved through the selective accessibility of

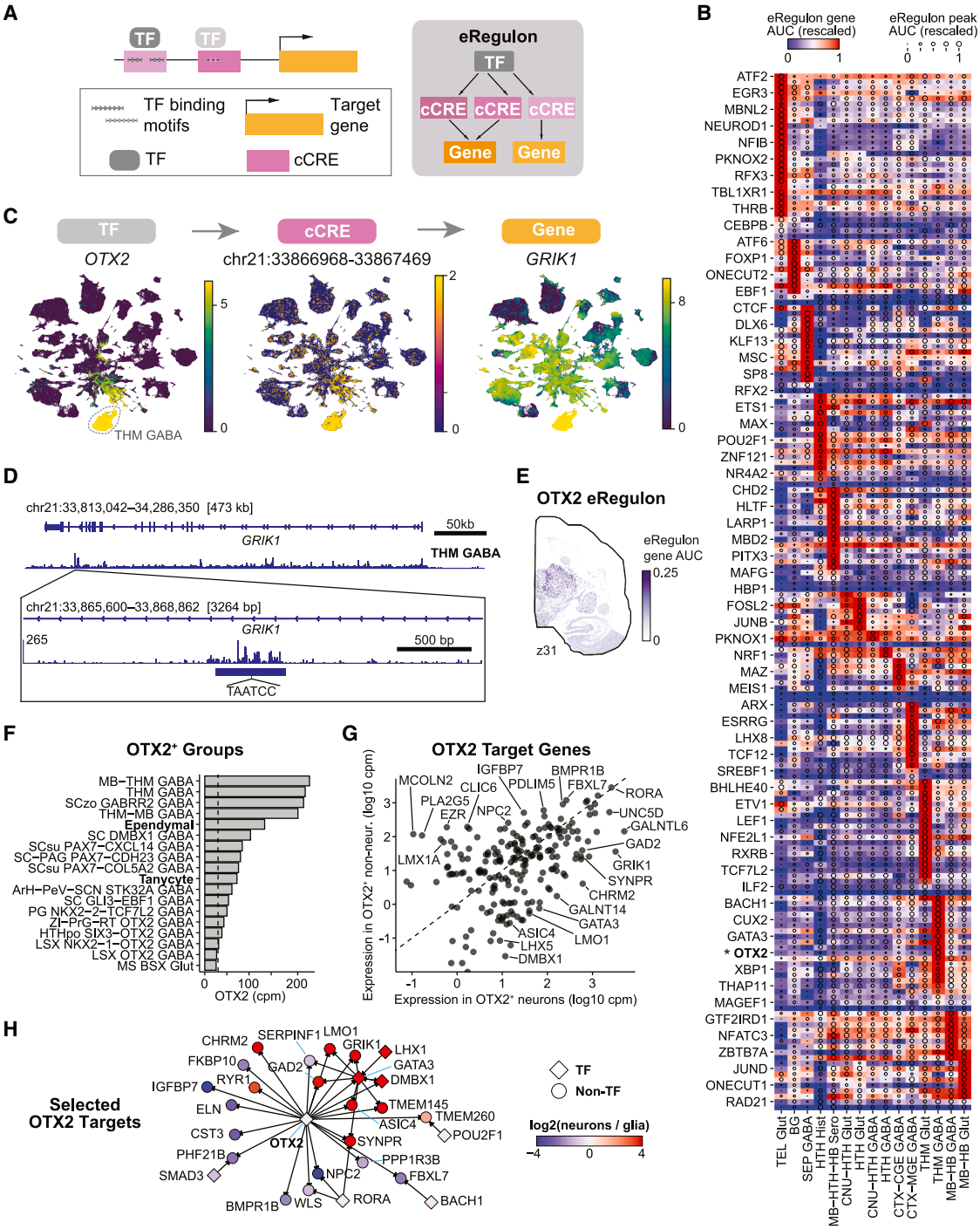


Figure 3. GRNs across the primate subcortex

(A) Schematic of eRegulon components (adapted from Bravo González-Blas et al.[54]).

(B) Direct activator ("+/+") eRegulons for each neuronal class; every third eRegulon is labeled. The *OTX2* eRegulon is denoted by an asterisk.

(C) UMAP plots for an eRegulon triplet; peak accessibility and target gene expression correlate with *OTX2* expression ($r = 0.77$ and 0.57, respectively).

(D) Within the class THM GABA, ATAC-seq insertion events near the peak and target gene from (B); the intronic peak contains a canonical *OTX2* motif.

(E) Representative spatial section showing high *OTX2* eRegulon activity in the THM.

(F) Top groups by *OTX2* expression; non-neurons are in bold.

(G) Expression of *OTX2* target genes in neuronal vs. non-neuronal groups, with select shared and differentially enriched genes highlighted.

(H) Network plot for selected neuron- or non-neuron-specific OTX2 targets (STAR Methods).

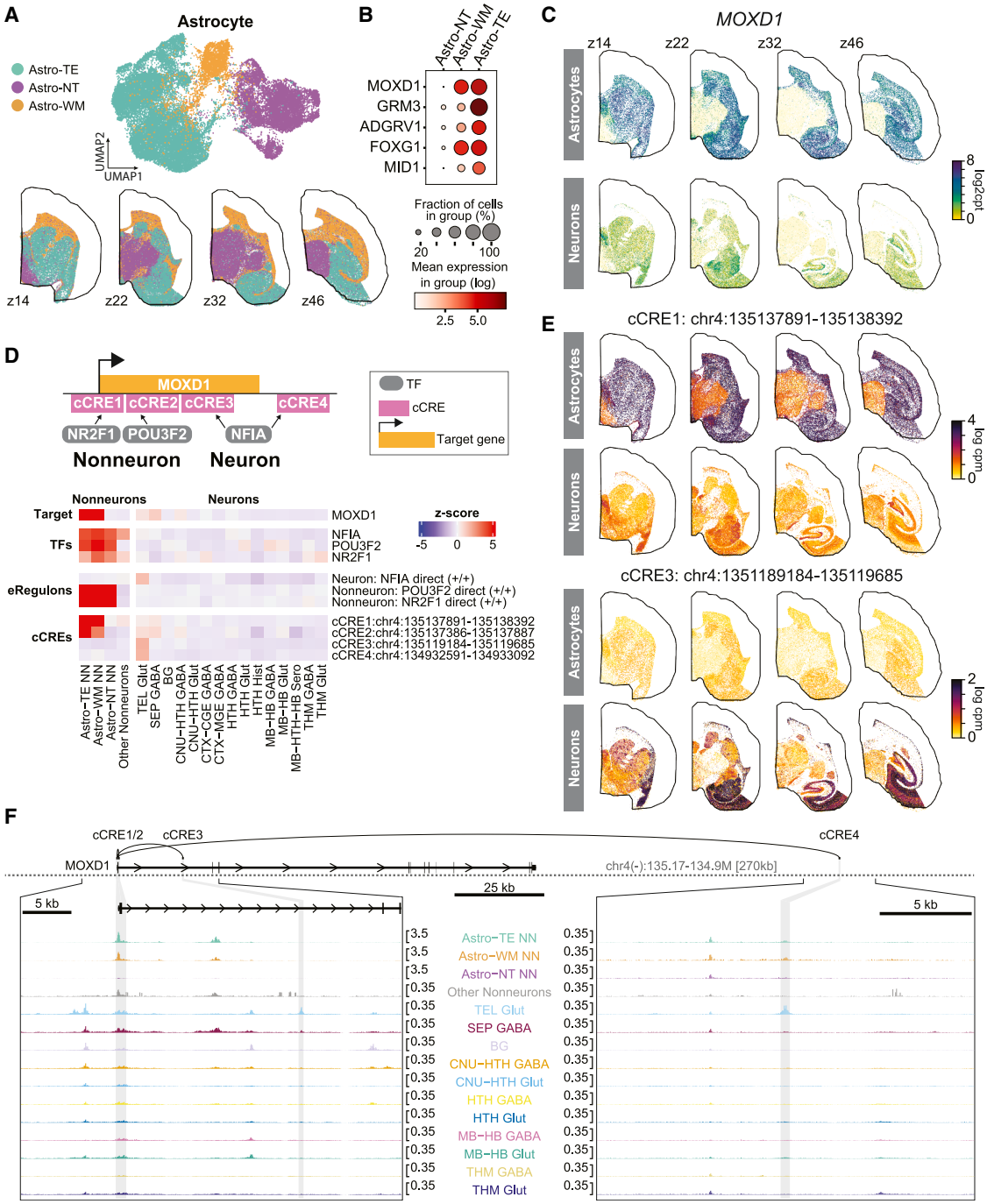


Figure 4. Regional zonation of astrocytes

(A) Regional zonation of the three astrocyte groups (Astro-NT NN, Astro-TE NN, and Astro-WM NN); UMAP plots (top) and representative spatial sections (bottom) are colored by group.

(B) Dot plot of top PC1 loading gene expression across astrocyte groups.

(C) *MOXD1* expression (log_2 counts per thousands [log_2cpt]) in all astrocytes (top) and all neurons (bottom) on representative spatial sections.

(D) *MOXD1*-related eRegulons in non-neurons and neurons. Top: schematic summarizing the TFs and cCREs driving *MOXD1* expression. Bottom: *Z* scored heatmaps of each component.

(E) Representative spatial sections showing cluster-averaged chromatin accessibility (log_2cpm) for an astrocyte-specific cCRE(1) and a neuron-specific cCRE(3).

(F) Browser tracks of *MOXD1*-related cCREs grouped by class. Scale bar, 5 kb.

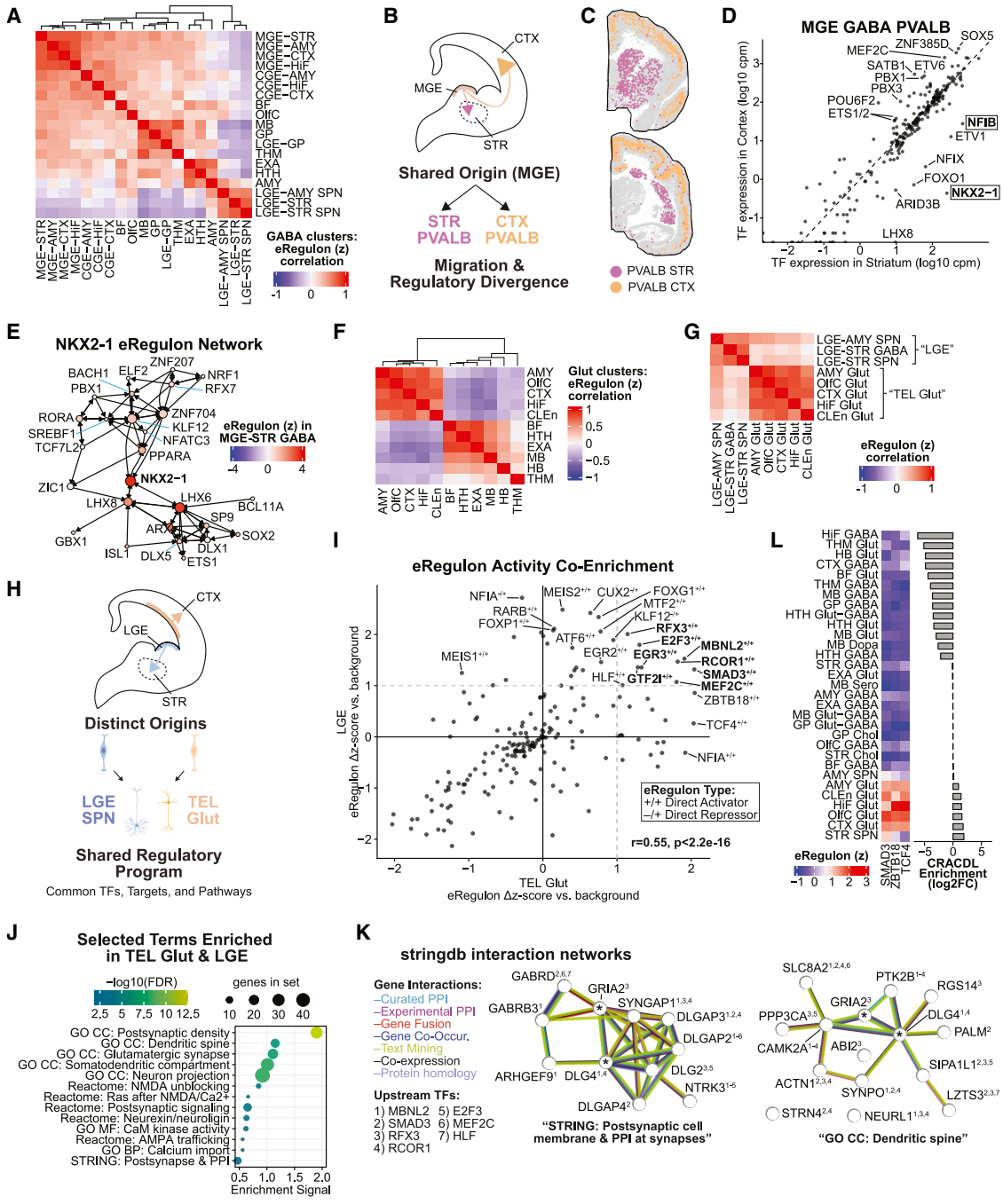


Figure 5. Regional dynamics of GRN activity

(A) GABAergic cluster-level eRegulon scores averaged by region, with SPNs and GE-derived clusters separated; Pearson correlations are shown.

(B) Schematic: MGE-derived GABAergic neurons migrate to both the CTX and striatum.

(C) Spatial mapping of STR- and CTX-localized *PVALB*$^{+}$ interneurons.

(D) TF expression comparison in STR- vs. CTX-localized PVALB clusters; similar NKX2-1 and NFIB patterns in SST neurons (Figure S13B).

(E) TF-TF eRegulon network centered on NKX2-1, colored by eRegulon activity averaged across MGE-STR GABA clusters.

(F) Comparison of glutamatergic clusters as in (A).

(G) LGE-derived GABAergic neurons and telencephalic glutamatergic neurons are highly similar yet distinct from other neurons (see also Figure S13A).

(H) Schematic of regulatory convergence between TEL Glut and LGE GABA.

(legend continued on next page)

promoter cCREs (cCRE1 and cCRE2) only in Astro-TE and Astro-WM. In neurons, *MOXD1* was a target of an *NFIA*-driven eRegulon acting at intronic (cCRE3) and intergenic (cCRE4) sites (Figures 4D–4F). Thus, the convergent regional expression of *MOXD1* in telencephalic neurons and astrocytes arises through distinct regulatory architectures.

Cross-species comparison revealed species differences: mouse astrocytes lack *Moxd1* despite its retention in telencephalic neurons (Figure S12H); human astrocytes show telencephalic enrichment like the marmoset (Figure S12I). This pattern is consistent with prior developmental studies showing *MOXD1* enrichment in glial progenitors of human outer radial glia during cortical development, whereas *Moxd1* is absent in comparable mouse radial glial populations,[65,66] suggesting that telencephalic and white-matter astrocyte *MOXD1* expression may reflect the retention of a primate-associated developmental program.

Developmental origin and regional context shape neuronal regulatory identity

We next examined how regulatory programs track anatomical location and developmental origin in neurons. Some progenitors—notably those that produce telencephalic GABAergic neurons—can disperse lineage-related daughters across regions.[67–70] This provides opportunities for regulatory divergence with a shared lineage or for regulatory convergence across distinct lineages.

To dissect these contributions, we regrouped localized neuron clusters by anatomical region (from spatial mapping) and major neurotransmitter, subdividing ganglionic eminence (GE)-derived neurons (GABAergic or cholinergic) by progenitor origin (medial [MGE], caudal [CGE], or lateral [LGE] ganglionic eminences), and separating SPNs from other GABAergic neurons. We then compared eRegulon usage across groups (Figure 5; Figure S13).

Regionally dispersed MGE-derived interneurons use distinct GRNs

Telencephalic interneurons derive from GE progenitors and migrate widely across the mammalian forebrain.[67,68,71,72] Consistent with their shared origin, MGE- and CGE-derived interneurons cluster by progenitor class rather than by final location at both the RNA (Figure 1B) and eRegulon level (Figure 5A; Figure S13A). MGE-derived interneurons primarily consist of parvalbumin-positive (*PVALB*$^{+}$) or somatostatin-positive (*SST*$^{+}$) subtypes that retain a shared transcriptomic identity across the CTX, hippocampus, and striatum. Yet even within subtypes, they show region-associated regulatory divergence (Figures 5B and 5C).

NKX2-1 emerged as a consistent striatum-enriched regulator for both the PVALB and SST subtypes (Figure 5D; Figure S13B). *NKX2-1* marks MGE progenitors and newborn neurons and is downregulated in CTX-bound interneurons.[73,74] In its TF-TF network (Figure 5E), *NKX2-1* activates downstream effectors (*LHX6*, *LHX8*, *ISL1*, *ARX*, *DLX1/6*, and *PPARA*), creating a self-reinforcing circuit for striatal identity. Some effectors (*LHX6/8* and *ISL1*) are established *NKX2-1* targets.[75,76] Conversely, known cortical TFs (*SATB1*, *SATB2*, ETS-family, *MEF2C*, and *TCF/NR3C1*) are enriched in CTX-localized cells and lie outside the NKX2-1 activator network.[77–80]

We also detected striatum-enriched *NFIB* in both PVALB and SST interneurons (Figure 5D; Figure S13B). *NFIB* activates the chemorepellents *SLIT3* and *EFNB2*, which are co-enriched in striatal interneuron subtypes. Slit-Robo and Ephrin signaling have previously been implicated in MGE interneuron migration,[68,81] and NFI TFs regulate GABAergic maturation[82]; these data support a role for *NFIB* in interneuron migration via Slit-Robo/Ephrin signaling.

Striatal SPNs and telencephalic glutamatergic neurons converge on a shared postsynaptic regulatory program

LGE-derived SPNs are striking among GABAergic neurons: in eRegulon usage, they cluster with telencephalic glutamatergic projection neurons rather than other GABAergic types (Figures 5F and 5G; Figure S13A), mirroring the RNA similarity between these two cell types (Figure 1B) observed in both mice and primates.[20,36,37] Despite arising from distinct progenitor domains,[83,84] SPNs and telencephalic glutamatergic neurons share eRegulons (Figures 5H and 5I; Figure S13C), including TFs implicated in both types (*EGR3*, *E2F3*, *RCOR1*, and *FOXG1*)[85–89] and TFs previously linked to telencephalic excitatory neurons only (*RFX3* and *MBNL2*).[90–93]

Given their divergent lineages, regions, and neurotransmitter identities, the LGE SPN-telencephalic glutamatergic neuron convergence points to a shared molecular function (Figure 5H). STRING[94] analysis of co-enriched target genes from shared eRegulons (Figures 5I and 5J) revealed heavily overlapping, co-regulated gene sets encoding postsynaptic densities, dendritic spines, and glutamatergic synapse components with known protein-protein interactions (Figure 5K; Figure S13D): postsynaptic density protein 95 (PSD-95)/discs large (DLG)-family scaffolds, SH3 and multiple ankyrin repeat domains (SHANK)-associated DLG associated protein (DLGAP)/SAPAP proteins, the Ras/Rap regulator synaptic Ras GTPase-activating protein 1 (SYNGAP1), and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)/N-methyl-D-aspartate (NMDA) calcium signaling components. This shared postsynaptic program may contribute to the frequent co-association of these neuron types in disorders such as autism and Huntington's disease.[95–99]

Beyond recovering known regulatory mechanisms, the joint analysis of gene expression and regulatory elements highlights less-studied biology. *CRACDL*, which we previously identified as jointly enriched in striatal SPNs and telencephalic glutamatergic neurons,[20] encodes a brain-enriched, vertebrate-specific protein that localizes to the nucleus, microtubules, and centrosomes.[100] Why it is restricted to these two neuronal populations has been unclear. We identified *CRACDL* in three direct-activator eRegulons: two glutamatergic-specific (TCF4 and ZBTB18) and one (SMAD3, transforming growth factor β [TGF-

(I) eRegulon activity enrichment for LGE GABA vs. TEL Glut (STAR Methods).
(J) Representative STRING-db-enriched gene sets from (I); set names are paraphrased.
(K) Two STRING enrichment networks; genes are annotated with upstream TFs (Figure S13D). Asterisks mark genes in both sets.
(L) CRACDL enrichment (log_2 fold change vs. neuronal clusters from other regions; STAR Methods).

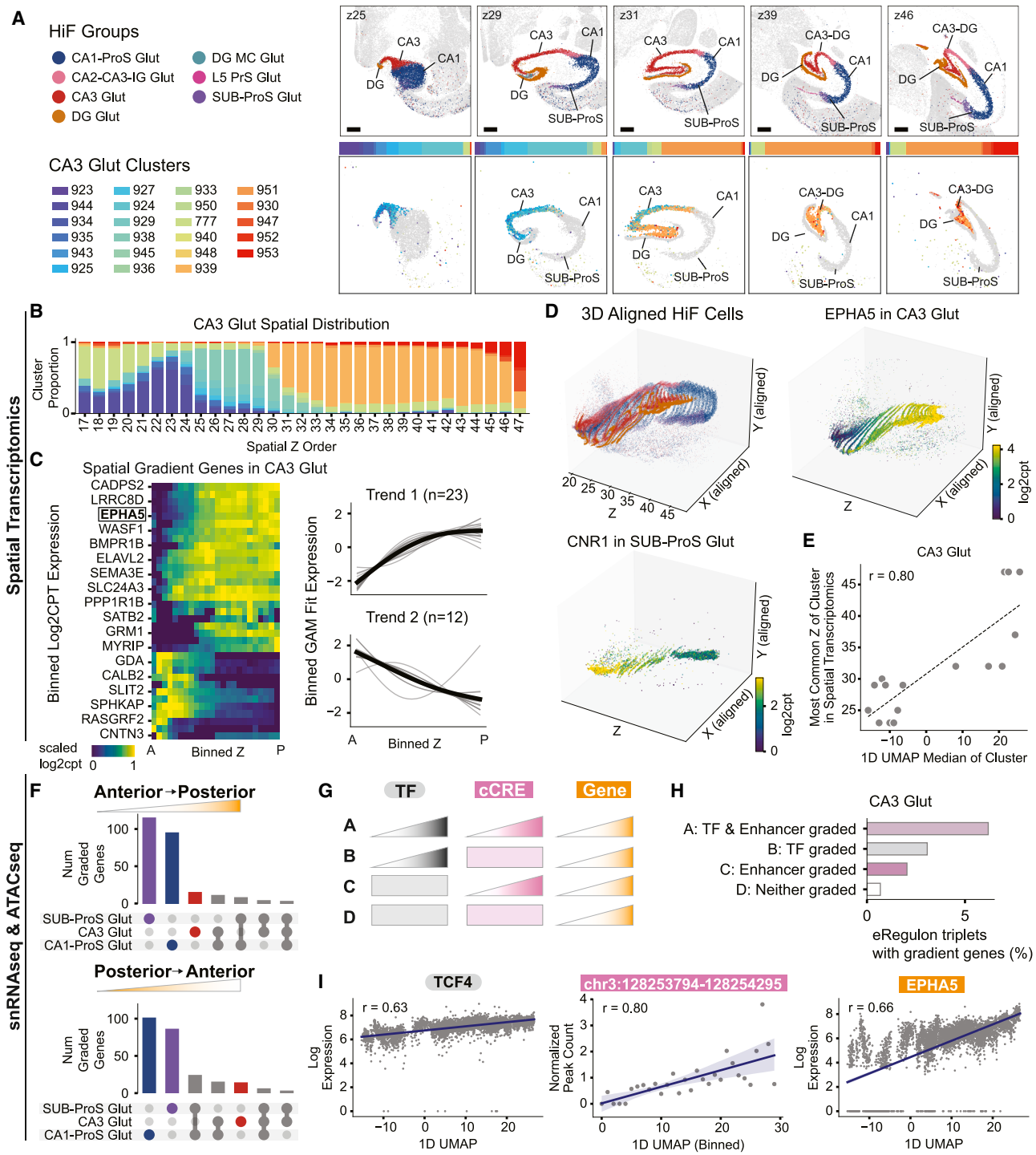


Figure 6. Epigenetic regulation of spatial gradients in primate HiF

(A) Representative spatial sections of groups in hippocampal subfields (top) and clusters in CA3 (bottom).

(B) Distribution of cluster proportions in CA3 Glut across the spatial *z* axis; cluster colors are as in (A).

(C) Left: $\log_2$cpt expression of CA3 Glut gradient genes across 30 A-P spatial bins; rows are ordered by hierarchically clustered GAM fit. Right: *Z* scored GAM-fitted expression separated by trend 1 (n = 23, increasing posteriorly) and trend 2 (n = 12, decreasing posteriorly).

(D) 3D aligned hippocampal cells colored by group, and $\log_2$cpt of *EPHA5* (CA3 Glut) and *CNR1* (SUB-ProS Glut) in aligned space.

(E) Correlation of median 1D UMAP with modal spatial z-section for each cluster in CA3 Glut.

(F) UpSet plots showing overlaps 1D-UMAP-correlated genes across hippocampal glutamatergic groups, partitioned into A-P (top) and P-A (bottom) gene sets.

(legend continued on next page)

β] pathway) also active in striatal SPNs (Figure 5L). In glutamatergic neurons, these TFs form an interconnected, self-reinforcing activator network upstream of *CRACDL*; only part of this network is active in striatal SPNs (Figure S13E). This implicates SMAD3/TGF-β signaling in both populations, suggesting that *CRACDL* is co-regulated with the spiny postsynaptic program.

Hippocampal subfields are patterned by spatially graded GRNs

Spatial mapping of the marmoset HiF along its longitudinal axis

In most mammals, the HiF lies dorsomedially below the corpus callosum—its ancestral position from cortical hem patterning.[67,101] In primates, the HiF shifts ventrally with neocortical and particularly temporal lobe expansion. We examined the marmoset hippocampus along its longitudinal (anterior-posterior [A-P]) extent using our spatial data. HiF structures were detected in 25/47 MSCA spatial sections. Eleven groups and 104 clusters mapped to the HiF, with six groups reflecting subfield boundaries (Figure 6A).

Cross-species mapping to the whole mouse brain atlas[1] corroborated group-level spatial mappings and suggested broad evolutionary conservation with some modifications. For example, the indusium griseum (IG)—a limbic structure adjacent to the corpus callosum linked to the dentate gyrus (DG) via the fasciola cinereum (FC)—remains dorsal in primates despite the hippocampus having shifted ventrally. Reports variously associate the IG with the cornu ammonis (CA)[102] or the DG.[103,104] In the MSCA, Cluster_955 (CA2-CA3-IG group) mapped spatially to the marmoset IG and to a restricted posterior CA3 location near the CA4 border, corresponding to mouse clusters 0400–0401 of the "025 CA2-FC-IG" supertype[1] (Figure S14A). In the marmoset, these cells spanned the A-P extent of the callosum and were posterior-restricted in CA3; in the mouse, they occupied the anterior callosum and dorsal hippocampus. This pattern is consistent with the known rodent-primate rotation of the hippocampal longitudinal axis and suggests that the IG is more closely related to CA subfields than to the DG.

Discrete cell types and continuous gene expression gradients pattern hippocampal subfields

Dorsal-ventral (D-V) gradients shape the topography of the mouse hippocampus.[1,105,106] The mouse D-V axis corresponds to the longitudinal (P-A) hippocampal axis in primates, and A-P gradients have been described in the human hippocampus.[107] We observed marked spatial localization of glutamatergic neuronal clusters along the marmoset A-P axis (Figures 6A and 6B; Figure S14B). A-P-restricted clusters were found in the DG, CA1, CA2, CA3, and prosubiculum/subiculum (ProS/SUB); for example, separate CA3 clusters occupied the head (Clusters_943–945), head/body (Clusters_924–925), body/tail (Cluster_930), and tail (Cluster_949) positions (Figures 6A and 6B). Similarly, A-P-restricted clusters were found in the DG, CA1, and ProS (Figure 6A). DG mossy cells (*CALB2+/PROX1−*) included two A-P-spanning clusters (Cluster_953 and Cluster_957) and one head-restricted cluster (Cluster_956) (Figure S14C). Consistent with the primate rotation of the hippocampal longitudinal axis, anterior marmoset clusters mapped to ventral mouse hippocampal types (e.g., the marmoset CA3 head/anterior-body Cluster_926 mapped to the mouse ventral-CA3 supertype 0076), while posterior marmoset clusters mapped to dorsal mouse types (marmoset CA3 posterior Cluster_951 to mouse dorsal-CA3 supertype 0078 CA3 Glut_4) (Figure S14D).

To characterize A-P gradients, we performed pairwise *z* axis alignment using PASTE2,[108] which leverages spatial proximity and gene expression similarity without using cluster identity. Post-alignment results show that cells from the same cluster occupy comparable spatial locations across sections, supporting continuous A-P transitions (Figure 6D). We then fit generalized additive models to the variable genes of each subfield group within the spatial panel (STAR Methods; Figure 6C; Figure S15A). Across CA3 Glut, CA1-ProS Glut, and SUB-ProS Glut, most graded genes showed monotonic A-P increases or decreases. As examples, EPHA5 (CA3 Glut) and CNR1 (SUB-ProS Glut) (Figure 6D) are validated by *in situ* hybridization (ISH) in the Marmoset Gene Atlas[12,38] (Figure S15B).

We leveraged the multiome data to extend beyond the 300-gene spatial panel. We embedded snRNA-seq clusters from each group in a one-dimensional uniform manifold approximation and projection (UMAP) and compared each cluster's median 1D coordinate to its modal Xenium *z*-section (STAR Methods). Median 1D UMAP positions strongly correlated with longitudinal *z* coordinates in all three subfield groups, indicating that the 1D UMAP axis captures the A-P position (Figure 6E; Figures S15C and S15D). We then identified genes with significant A-P gradients by correlating cell-level gene expression with 1D UMAP coordinates (STAR Methods). Across the three groups, 244 genes were anteriorly and 115 genes posteriorly biased (|Spearman r|> 0.5, p adj < 0.05). Strikingly, 318/359 (89%) graded genes were private to a single group (single hippocampal subfield) (Figure 6F). Comparison with the human hippocampus A-P snRNA-seq data at an anterior and posterior CA1 location[109] showed significant overlap between human CA1 differentially expressed genes and marmoset CA1-ProS gradient genes (p adj < 0.05; Figure S15E), indicating the conservation of subfield-specific A-P programs across primates.

Graded transcription factors and enhancers jointly maintain A-P expression gradients

Classical morphogen gradients (such as Bicoid in *Drosophila*) rely on concentration-dependent enhancer activity with the enhancer itself uniformly competent, while other patterning systems (such as the spinal cord[110]) use spatially graded enhancer accessibility. Distinct enhancers are active at the rostrodorsal and caudoventral poles of the mouse hippocampal primordium,[111] but it is unknown how adult primate

(G) Schematic of possible TF/enhancer gradient scenarios driving target genes.
(H) Percentage of eRegulon triplets per scenario in (G).
(I) Example TCF4 +/+ eRegulon triplet in CA3 Glut: TF *TCF4* expression, enhancer peak accessibility (chromosome 3 [chr3]:128253794–128254295), and target *EPHA5* expression plotted against 1D UMAP coordinates.

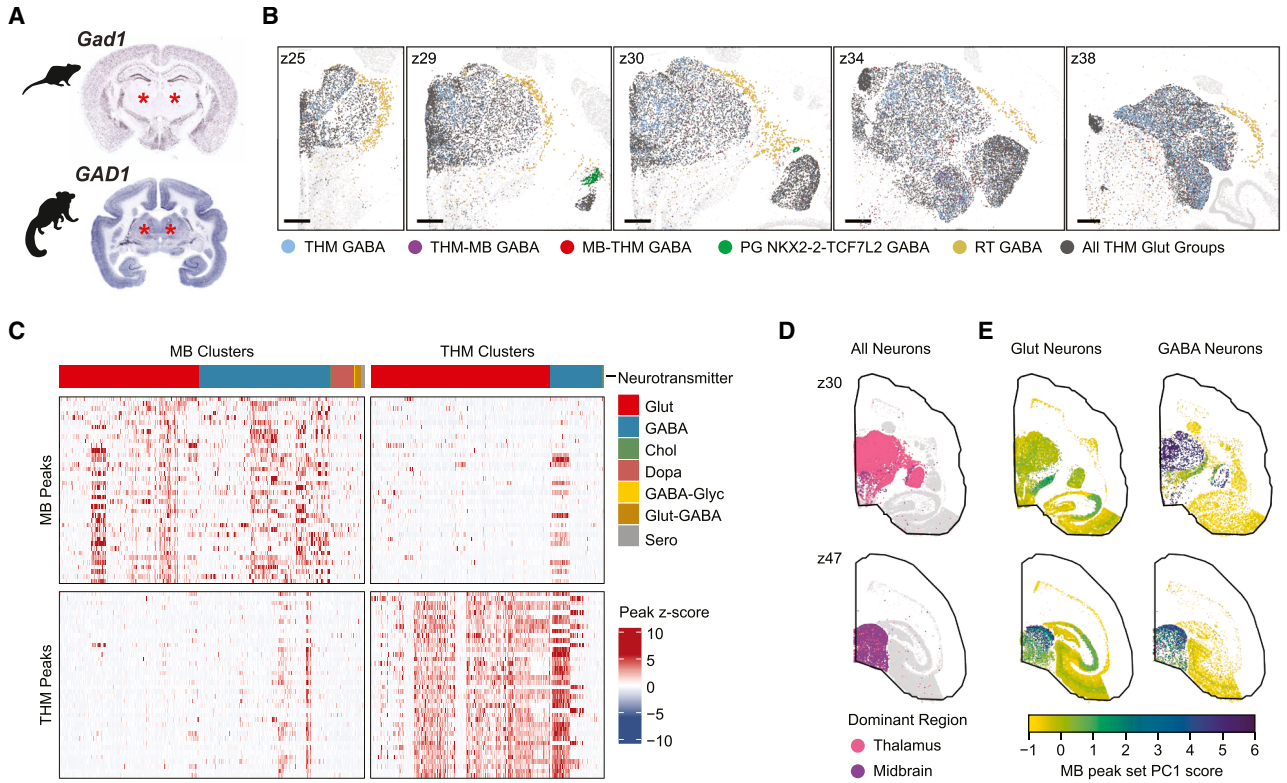


Figure 7. Evolutionary modifications to primate THM

(A) ISH with *Gad1* (mouse) (RRID: SCR_002978) and *GAD1* (marmoset)[12,38] probes showing the absence and presence of thalamic GABAergic cells, respectively; the THM is marked by red asterisks.

(B) Representative spatial z-planes of neurons in the five THM GABA groups (colored) and THM Glut groups (gray).

(C) Heatmap of regional marker peaks (shared in localized Glut and GABA neurons) across MB- and THM-localized clusters.

(D) Representative spatial z-planes of all THM (top) and MB (bottom) neurons.

(E) Same z-planes as (D), showing the PC1 score of the MB peak set in glutamatergic (left) and GABAergic (right) neurons.

hippocampal expression gradients are maintained. We considered four TF-cCRE-target scenarios (Figure 6G): (1) both TF and cCRE graded (reinforced gradient), (2) only TF graded, (3) TF uniform but cCRE accessibility graded, and (4) neither graded (implying other mechanisms).

We tested these by extending the gradient analysis to cCRE accessibility across all eRegulons (cell-level correlation with a 1D UMAP; STAR Methods). Using TF-cCRE-target triplets, we classified each gradient target gene by scenario. In CA3 Glut, most gradient genes fell under scenario A (both TF and enhancer graded; Figure 6H), with the same pattern in other subfields (Figure S15F); scenario D (neither graded) accounted for only a small minority.

As a concrete example, we examined the TCF4+/+ activator eRegulon in CA3 Glut. *TCF4* itself is significantly A-P graded, increasing posteriorly. Its downstream target EPHA5 (ephrin receptor) shows a matched posterior gradient (Figures 6D and 6I), and the TCF4-motif-containing cCRE—whose accessibility correlates with EPHA5 expression—is similarly A-P graded. Thus, longitudinal gradients in excitatory hippocampal neurons are predominantly generated by the coordinated gradations of TFs and their cCREs, with additional mechanisms likely contributing.

Marmoset thalamic GABAergic neurons share transcriptomic and regulatory signatures with midbrain populations

Cross-species mapping links thalamic GABAergic neurons to a conserved midbrain population

Aside from the thalamic reticular nucleus (RT), the rodent THM is almost entirely glutamatergic. In the mouse, non-RT thalamic inhibitory neurons comprise only ~6% of all thalamic neurons, most of which are found in the lateral geniculate nucleus (LG).[7,112] Primates harbor abundant GABAergic populations in both sensory and higher-order thalamic nuclei (30%–50% in some nuclei).[113,114] From our spatial data, 20% of marmoset thalamic neurons (excluding the RT, habenular nuclei, and dorsal LG) are GABAergic (Figures 7A and 7B).

Prior work suggests that the expanded primate thalamic GABAergic population may arise from multiple origins: the diencephalon and telencephalon,[115,116] the midbrain and diencephalon,[7] or the midbrain, diencephalon, and telencephalon (in humans).[117] Because adult transcriptomes strongly retain lineage information,[1,3,20] we reasoned that transcriptional similarity between thalamic GABAergic neurons and other cell types may indicate their origins in the MSCA dataset.

Non-RT thalamic GABAergic cells comprise three groups: THM GABA, THM-MB GABA, and MB-THM GABA (Figure 7B, MB: midbrain). Some clusters were nucleus specific (e.g., the LG-restricted Cluster_38; Figure S16A), but many were distributed across both the THM and midbrain (Figure S16B): 26% of cells in these groups were found outside the THM. This thalamic-midbrain sharing raised the possibility that primate-expanded thalamic GABAergic populations derive substantially from a conserved midbrain population. Supporting this, all three groups aligned to a single mouse[1] subclass, 213 SCsg Gabrr2 Gaba (Table S5; STAR Methods), which predominates in the mouse SC with a sparse LG extension (Figure S16C). SAMap[118] supported this homology (Figure S16D).

We also examined the direct mapping between the human[3] and marmoset atlases, plus an additional Slide-seq mouse atlas.[2] All three marmoset THM GABA groups mapped to the human *midbrain-derived inhibitory* supercluster[3] (containing samples from the human THM, midbrain, and pons; Figure 2A; Table S5), and both mouse atlases mapped them predominantly to midbrain and tegmental populations, reinforcing a conserved midbrain-like transcriptional identity for primate thalamic GABAergic neurons.

Shared CREs support a midbrain origin for marmoset thalamic inhibitory neurons

We next asked whether the midbrain-like identity of THM GABA neurons is also reflected at the regulatory level. By eRegulon similarity, midbrain GABAergic neurons were most similar to thalamic GABAergic neurons (Figures S13A and S16E). Co-enriched eRegulons included those with the TFs *GATA3*, *OTX2*, *BACH1*, and *ESRRG* (Figure S16F).

We further hypothesized that the chromatin landscape might retain evidence of cell lineage if some accessibility patterns are carried from an earlier developmental state. As noted earlier, midbrain glutamatergic and GABAergic neurons co-enrich for midbrain-specific peaks (Figure S7E). If primate-expanded thalamic GABAergic neurons originate from the midbrain, they should retain some of these peaks. We ranked ATAC peaks by their shared specificity to midbrain glutamatergic and GABAergic cell types, defining them as midbrain-marker peaks (STAR Methods). This analysis produced distinct, regionally specific peak sets for most brain regions (Figure S16G), yet many midbrain-specific peaks were also accessible in thalamic GABAergic neurons (Figure 7E). Projecting PC1 of this peak set onto the spatial data showed that these peaks are shared among midbrain glutamatergic/GABAergic and thalamic GABAergic neurons and are absent elsewhere (Figure 7F).

Mouse developmental data suggest that excitatory and inhibitory SC neurons arise from shared multipotent progenitors.[119] In humans, *OTX2*, a shared marker of the non-RT thalamic GABAergic and SC populations, is also expressed in the midbrain and dorsal diencephalon (future THM) as early as gestational week 10.[120] Together with the cross-species transcriptomic and regulatory concordance, these findings supports a midbrain origin for most marmoset non-RT thalamic GABAergic neurons,[7] with their expansion representing a primate adaptation.

Interactive multimodal exploration with Cytosplore Viewer

To enable the integrated exploration of the MSCA, we developed the Cytosplore ATAC Viewer. Complementing web-based taxonomy viewers (e.g., Cirrocumulus [RRID: SCR_021646]), Cytosplore Viewer provides an offline, fast, and interactive plugin-based platform for exploring all three data modalities in depth. The viewer leverages the common taxonomy across modalities to transfer gene expression or chromatin features onto spatial data. In addition, we developed a bidirectional search engine to link spatial RNA gene expression and spatial ATAC peak accessibility patterns (STAR Methods). Starting from a seed gene's expression pattern across the spatial atlas, a search can be performed for ATAC peaks with a similar spatial pattern. Conversely, a spatial ATAC profile can be used as a seed to find similarly distributed genes. These searches can be confined to manually selected spatial regions of interest. The Cytosplore ATAC Viewer also enables interactive PCA computation from user-defined sets of peaks or genes, followed by spatial visualization of PC1 (Video S1). Cytosplore Viewer can be downloaded from https://viewer.cytosplore.org and is an integral part of the BICAN data tooling ecosystem.

DISCUSSION

A spatially resolved multimodal atlas of the primate subcortex

Brain cell atlases now describe cellular diversity at high resolution, but variable sampling strategies across studies mean that gaps remain. Here, the combination of anatomically comprehensive tiling, spatial localization, and paired chromatin accessibility in a primate brain creates a distinct resource that is well suited for linking molecular taxonomy to anatomical context and the candidate regulatory architecture. The analyses presented highlight several ways this framework reveals broad principles of molecular organization but are only a snapshot of what the MSCA supports, given the breadth of the taxonomy and its spatial and chromatin modalities.

Notably, groups and even clusters of the snRNA-seq-defined taxonomy mapped to anatomically precise locations. This suggests that non-optimized gene panels with modest numbers of genes can still capture substantial cellular diversity. More broadly, defining cell types through quantitative variation across many genes—rather than through highly specific individual markers—aligns with the concept of cell types as peaks in a joint expression probability distribution,[121] a framework that is especially powerful for cross-species comparisons of divergent cell populations.[4,122]

Distinct regulatory architectures can produce similar transcriptional results

By comprehensively surveying broad anatomical structures, our multimodal atlas clarifies not only where transcriptionally related cell types occur but also how those identities are implemented through gene regulatory programs. Although chromatin accessibility and gene expression are broadly aligned across the taxonomy (Figure 1D), their occasional discordance is informative: similar transcriptional outputs can arise through distinct

regulatory architectures, as illustrated by the shared expression of *MOXD1* in neurons and astrocytes. This case shows that transcriptional similarity does not always predict enhancer activity across cell types—a consideration for the design of cell-type-specific genetic tools.

Regulatory models reveal how spatial identity is preserved or reshaped

Because postmitotic brain cells usually remain within their developmental compartment of origin, regional identity and cell type identity are typically intertwined.[123,124] But cellular identity can also be reshaped: either by migration (MGE-derived interneurons) or by cell-type-specific functional requirements (telencephalic glutamatergic neurons and SPNs). Our analyses nominate TFs that may act as master regulators of these shifts, though adult data alone cannot resolve the underlying developmental mechanisms.

Our analysis of the hippocampus extends this framework to continuous regulatory variation within an anatomically defined population: discrete, subfield-restricted cell types are superimposed on A-P programs marked by graded transcription factors and coordinately graded cCREs. Whether these subfield-specific gradients are coordinated by a shared upstream regulator, and how they are established during development, remain open questions.

Adult regulatory states generate testable hypotheses about developmental origin

Differential migration of conserved cell types is a potent route by which evolution reorganizes brain circuits[5,18]; our atlas enables predictions about developmental origins from the adult regulatory state, where migration or anatomical redistribution obscures lineage relationships in single-modality analyses. The primate-expanded non-RT thalamic GABAergic population illustrates this, aligning with conserved midbrain populations across transcriptomic mapping, eRegulon usage, and chromatin accessibility. Developmental validation will be required to confirm the proposed migratory origin; more work is needed to determine how and when such migration occurs and to determine its functional consequences. More generally, similar peak- and gene-set-based analyses can interrogate candidate lineage relationships, regionally patterned cell states, or disease-relevant regulatory features across the atlas. Users can project their own feature sets in our Cytosplore Viewer for interactive exploration.

Conclusions

The MSCA links high-resolution molecular taxonomies to three-dimensional anatomy across the primate subcortex. Through cross-species embedding, we disambiguate heterogeneous human cell populations and refine spatial annotations in complex anatomical subdivisions. The concurrent cross-species BG taxonomy (marmoset, macaque, and human)[32,33] using parallel nomenclature and annotation indicates that much of the MSCA will prove conserved across primates. Ongoing work in BICAN to define cell types across the whole brain in humans and nonhuman primates will benefit from MSCA annotations as data collection and processing proceed across other species. A powerful future application will be to map human genetic risk variants for neuropsychiatric and neurodegenerative diseases onto spatially resolved primate cell types,[125] leveraging the spatial context to interpret noncoding variation. As such, this cellular-resolution atlas of the compact marmoset subcortex is well situated to bridge our understanding of rodent and human cell-type organization.

Limitations of the study

Although comprehensive in the subcortex, the present study is anatomically incomplete. Coverage of the cerebellum and neocortex will be required for a complete taxonomy. Our limited donor number and the absence of developmental time points preclude the analyses of interindividual variation, sex differences, or developmental trajectories between cell types. Future whole-brain and developmental primate atlases, including those from BICAN, will extend this work.

Furthermore, GRN inference remains technically limited. A limitation of SCENIC+ is its insensitivity to repressive relationships[54]; accordingly, we recovered relatively few repressors. SCENIC+ also relies on conventional motif matching for TFBS prediction; these predictions may be improved by emerging deep-learning approaches.[126,127]

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Fenna Krienen (fkrienen@princeton.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw 10x Multiome data have been deposited at NeMO (RRID: SCR_016152, https://assets.nemoarchive.org/dat-kkuxhz4) and all processed data are publicly available at (https://brain-map.org/consortia/hmba and https://allen-hmba-releases.s3.us-west-2.amazonaws.com/index.html#marmoset_subcortex/).
- All original code used in the manuscript has been deposited at https://github.com/KrienenLab/MSCA.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This publication was supported by and coordinated through the Brain Initiative Cell Atlas Network (BICAN), RRID: SCR_022794. This research was supported by the National Institutes of Health award UM1MH130981. The work was additionally supported by National Institutes of Health award DP2MH140136 and the Klingenstein-Simons Fellowship (F.M.K.), the Dutch Research Council (NWO) award 024.004.012 (B. Lelieveldt), the McDonnell Fellowship in Neuroscience at Princeton University (S.D. and N.W.), and National Institutes of Health award 5T32MH065214 (N.W.). We thank the MIT DCM veterinarian staff and animal care technicians who assisted in anesthesia, postoperative care, clinical support, and animal husbandry. We thank Christina DeCoste, Gabriel Palmieri, Jailene Garcia, and the Molecular Biology Flow Cytometry Resource Facility, which is partially supported by the Rutgers Cancer Institute of New Jersey NCI-CCSG P30CA072720-5921. Illustrations in the graphical abstract and Figure 1A were created in part using BioRender.[128,129]

AUTHOR CONTRIBUTIONS

Sample preparation and multi-omic data generation: C.C., F.M.K., J.F., J.S., K.L., L.C., Q.Z., and R.N. Spatial transcriptomic data generation: A.R., B. Long, D.A.M., J.A., J.C., J.N., J.W., M.M.V., M.A.T., N.M., N.V.C., P.O., S. Barta, S.H., S.R., and Z.J. Data analysis and interpretation: B.J.R., B. Lelieveldt, B. Long, C.L., F.M.K., H. Zheng, M.A.T., M.D., N.W., S. Basu, S.D., T.H., V.E.N.-C., and X.L. Manuscript writing: B. Lelieveldt, B. Long, C.L., F.M.K., H. Zheng, J.S., K.L., L.C., M.A.T., M.D., N.W., S.D., V.E.N.-C., and X.L. Funding acquisition: H. Zeng, E.S.L., B. Lelieveldt, G.F., B. Long, and F.M.K.

DECLARATION OF INTERESTS

H. Zeng is on the scientific advisory board of MapLight Therapeutics, Inc.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE MANUSCRIPT PREPARATION PROCESS

During the preparation of this work, the authors used ChatGPT in order to improve readability of original drafts. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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
- METHOD DETAILS
  - Tissue processing
  - Nuclei isolation
  - Fluorescence-activated nuclei sorting (FANS)
  - Sample preparation for multiome sequencing
  - Xenium spatial transcriptomics data collection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Marmoset genome reference
  - Multiome sequence alignment and library QC
  - Multiome RNA processing pipeline
  - Marker gene identification and differential gene expression analysis
  - Cross-species mapping against reference dataset
  - Xenium spatial transcriptomics processing and QC
  - Cross-modality cell-type mapping between spatial transcriptomics and multiome RNA-seq
  - In silico validation of gene panel performance for spatial RNA-seq mapping
  - Generation of MSCA Taxonomy
  - Comparative analysis across mouse, marmoset and human neurons
  - Multiome ATAC processing pipeline
  - Multiome ATAC-seq browser shots
  - Gene Regulatory Network (GRN/eRegulon) calling and scoring
  - Cytosplore Viewer
  - RNA/ATAC co-enrichment analysis by region and neurotransmitter
  - Principal component analysis of astrocyte Clusters
  - TEL Glut vs. LGE GABA functional convergence analysis
  - Alignment of Xenium Spatial Transcriptomics data in three-dimensions
  - Hippocampal Gradient Analysis
  - Cross-species mapping of primate non-RT Thalamic GABAergic population
  - Regional ATAC marker analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.10.007.

Received: November 24, 2025
Revised: April 27, 2026
Accepted: October 3, 2026
Published: November 12, 2026

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. 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.
4. Corrigan, E.K., DeBerardine, M., Poddar, A., Turrero García, M., de la O, S., He, S., Sen, H., Duhne, M., Lindberg, S., Song, M., et al. (2025). Conservation and alteration of mammalian striatal interneurons. Nature *647*, 187–193. https://doi.org/10.1038/s41586-025-09592-w.
5. Schmitz, M.T., Sandoval, K., Chen, C.P., Mostajo-Radji, M.A., Seeley, W.W., Nowakowski, T.J., Ye, C.J., Paredes, M.F., and Pollen, A.A. (2022). The development and evolution of inhibitory neurons in primate cerebrum. Nature *603*, 871–877. https://doi.org/10.1038/s41586-022-04510-w.
6. Letinic, K., and Rakic, P. (2001). Telencephalic origin of human thalamic GABAergic neurons. Nat. Neurosci. *4*, 931–936. https://doi.org/10.1038/nn0901-931.
7. Jager, P., Moore, G., Calpin, P., Durmishi, X., Salgarella, I., Menage, L., Kita, Y., Wang, Y., Kim, D.W., Blackshaw, S., et al. (2021). Dual midbrain and forebrain origins of thalamic inhibitory interneurons. eLife *10*, e59272. https://doi.org/10.7554/eLife.59272.
8. Okano, H. (2021). Current Status of and Perspectives on the Application of Marmosets in Neurobiology. Annu. Rev. Neurosci. *44*, 27–48. https://doi.org/10.1146/annurev-neuro-030520-101844.
9. Chen, Y., Menegas, W., Zhang, Q., and Feng, G. (2025). Common marmoset: An emerging non-human primate model for translational applications in brain disorders. Curr. Opin. Neurobiol. *92*, 102998. https://doi.org/10.1016/j.conb.2025.102998.
10. Bridge, H., Leopold, D.A., and Bourne, J.A. (2016). Adaptive Pulvinar Circuitry Supports Visual Cognition. Trends Cogn. Sci. *20*, 146–157. https://doi.org/10.1016/j.tics.2015.10.003.
11. Rosa, M.G.P., and Tweedale, R. (2005). Brain maps, great and small: lessons from comparative studies of primate visual cortical organization. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 665–691. https://doi.org/10.1098/rstb.2005.1626.
12. 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.
13. Chaplin, T.A., Yu, H.-H., Soares, J.G.M., Gattass, R., and Rosa, M.G.P. (2013). A Conserved Pattern of Differential Expansion of Cortical Areas

in Simian Primates. J. Neurosci. *33*, 15120–15125. https://doi.org/10.1523/JNEUROSCI.2909-13.2013.

14. Huang, Z., Yang, Q., Li, S., Zhu, X., Wang, H., Lin, J., Zhan, Y., Wu, Y., Wang, Z., Majka, P., et al. (2026). An opposing molecular gradient axis underlies primate cortical organization. Science *392*, eaea2673. https://doi.org/10.1126/science.aea2673.
15. Buckner, R.L., and Margulies, D.S. (2019). Macroscale cortical organization and a default-like apex transmodal network in the marmoset monkey. Nat. Commun. *10*, 1976. https://doi.org/10.1038/s41467-019-09812-8.
16. Tian, X., Chen, Y., Majka, P., Szczupak, D., Perl, Y.S., Yen, C.C.-C., Tong, C., Feng, F., Jiang, H., Glen, D., et al. (2022). An integrated resource for functional and structural connectivity of the marmoset brain. Nat. Commun. *13*, 7416. https://doi.org/10.1038/s41467-022-35197-2.
17. Watakabe, A., Skibbe, H., Nakae, K., Abe, H., Ichinohe, N., Rachmadi, M.F., Wang, J., Takaji, M., Mizukami, H., Woodward, A., et al. (2023). Local and long-distance organization of prefrontal cortex circuits in the marmoset brain. Neuron *111*, 2258–2273.e10. https://doi.org/10.1016/j.neuron.2023.04.028.
18. 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.
19. Kelava, I., Reillo, I., Murayama, A.Y., Kalinka, A.T., Stenzel, D., Tomancak, P., Matsuzaki, F., Lebrand, C., Sasaki, E., Schwamborn, J.C., et al. (2012). Abundant occurrence of basal radial glia in the subventricular zone of embryonic neocortex of a lissencephalic primate, the common marmoset Callithrix jacchus. Cereb. Cortex *22*, 469–481. https://doi.org/10.1093/cercor/bhr301.
20. 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.
21. 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.
22. Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., Su, G., Liu, Y., Qin, X., Rosoklija, G.B., et al. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. Nature *609*, 375–383. https://doi.org/10.1038/s41586-022-05094-1.
23. Kong, D., He, H., Lechner, J., Yang, J., Zhang, A., Li, L., Mathur, R., Darland, D.C., Takaku, M., Lu, L., et al. (2025). Spatial profiling of chromatin accessibility reveals alteration of glial cells in Alzheimer's disease mouse brain. Preprint at bioRxiv. https://doi.org/10.1101/2025.05.01.651759.
24. Ding, W., Klein, A., Báez-Becerra, C.T., Rink, J.A., Bartlett, A., Zeng, Q., Wang, R., Castanon, R.G., Nery, J.R., Osgood, E., et al. (2026). A Multimodal Single-Cell Epigenomic and 3D Genome Atlas of the Human Basal Ganglia. Preprint at bioRxiv. https://doi.org/10.64898/2026.02.12.705594.
25. Lu, T., Ang, C.E., and Zhuang, X. (2022). Spatially resolved epigenomic profiling of single cells in complex tissues. Cell *185*, 4448–4464.e17. https://doi.org/10.1016/j.cell.2022.09.035.
26. Zhang, D., Deng, Y., Kukanja, P., Agirre, E., Bartosovic, M., Dong, M., Ma, C., Ma, S., Su, G., Bao, S., et al. (2023). Spatial epigenome–transcriptome co-profiling of mammalian tissues. Nature *616*, 113–122. https://doi.org/10.1038/s41586-023-05795-1.
27. Hunker, A.C., Wirthlin, M.E., Gill, G., Johansen, N.J., Hooper, M., Omstead, V., Vargas, S., Lerma, M.N., Taskin, N., Weed, N., et al. (2025). Enhancer AAV toolbox for accessing and perturbing striatal cell types and circuits. Neuron *113*, 1507–1524.e17. https://doi.org/10.1016/j.neuron.2025.04.035.
28. Mich, J.K., Graybuck, L.T., Hess, E.E., Mahoney, J.T., Kojima, Y., Ding, Y., Somasundaram, S., Miller, J.A., Kalmbach, B.E., Radaelli, C., et al. (2021). Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. Cell Rep. *34*, 108754. https://doi.org/10.1016/j.celrep.2021.108754.
29. Caplan, K.A., Chen, K., Wienisch, M., Wang, Y., Chai, R., Kang, I.-H., Chen, C.S., Del Rosario, R., Krienen, F.M., Wang, F., et al. (2025). Next-generation enhancer AAVs for selective interspecies targeting of midbrain dopaminergic neurons. Preprint at bioRxiv. https://doi.org/10.1101/2025.11.06.687071.
30. Dimidschstein, J., Chen, Q., Tremblay, R., Rogers, S.L., Saldi, G.-A., Guo, L., Xu, Q., Liu, R., Lu, C., Chu, J., et al. (2016). A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat. Neurosci. *19*, 1743–1749. https://doi.org/10.1038/nn.4430.
31. Graybuck, L.T., Daigle, T.L., Sedeño-Cortés, A.E., Walker, M., Kalmbach, B., Lenz, G.H., Morin, E., Nguyen, T.N., Garren, E., Bendrick, J.L., et al. (2021). Enhancer viruses for combinatorial cell-subclass-specific labeling. Neuron *109*, 1449–1464.e13. https://doi.org/10.1016/j.neuron.2021.03.011.
32. Johansen, N.J., Fu, Y., Schmitz, M., Dubuc, A., Kempynck, N., Wirthlin, M., Garcia, A.D., Hewitt, M., Turner, M.A., Seeman, S.C., et al. (2025). Cross-species consensus atlas of the primate basal ganglia. Preprint at bioRxiv. https://doi.org/10.64898/2025.12.15.694496.
33. Hewitt, M.N., Turner, M.A., Johansen, N., McMillen, D.A., Dan, S., DeBerardine, M., Ruiz, A., Huang, M., Quon, J., Fu, Y., et al. (2025). A cross-species spatial transcriptomic atlas of the human and non-human primate basal ganglia. Preprint at bioRxiv. https://doi.org/10.1101/2025.11.22.688128.
34. 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.
35. Allen Institute for Brain Science. (2025). transcriptomic_clustering: improved clustering pipeline for transcriptomic data (v1.0.0). GitHub. https://github.com/AllenInstitute/transcriptomic_clustering.
36. 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.
37. 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.
38. 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.
39. Daniel, S.F., Lee, C., Mollenkopf, T., Lee, M., Arbuckle, J., Fiabane, E., Gabitto, M.I., Johansen, N., Kapen, I., Kraft, A.W., et al. (2026). MapMyCells: High-performance mapping of unlabeled cell-by-gene data to reference brain taxonomies. Preprint at bioRxiv. https://doi.org/10.64898/2026.03.06.710160.
40. Totty, M.S., Cervera Juanes, R., Bach, S.V., Ben Ameur, L., Valentine, M.R., Simons, E., Romac, M.D., Trinh, H., Henderson, K., Del Rosario, I., et al. (2025). Transcriptomic diversity of amygdalar subdivisions across humans and nonhuman primates. Sci. Adv. *11*, eadw1029. https://doi.org/10.1126/sciadv.adw1029.
41. Steuernagel, L., Lam, B.Y.H., Klemm, P., Dowsett, G.K.C., Bauder, C.A., Tadross, J.A., Hitschfeld, T.S., del Rio Martin, A., Chen, W., de Solis, A.J., et al. (2022). HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. Nat. Metab. *4*, 1402–1419. https://doi.org/10.1038/s42255-022-00657-y.
42. Tadross, J.A., Steuernagel, L., Dowsett, G.K.C., Kentistou, K.A., Lundh, S., Porniece, M., Klemm, P., Rainbow, K., Hvid, H., Kania, K., et al. (2025).

A comprehensive spatio-cellular map of the human hypothalamus. Nature *639*, 708–716. https://doi.org/10.1038/s41586-024-08504-8.

43. Phillips, R.A., Oh, S., Bach, S.V., Du, Y., Miller, R.A., Kleinman, J.E., Hyde, T.M., Hicks, S.C., Page, S.C., and Martinowich, K. (2025). Transcriptomic characterization of human lateral septum neurons reveals conserved and divergent marker genes across species. iScience *28*, 111820. https://doi.org/10.1016/j.isci.2025.111820.

44. Thompson, J.R., Nelson, E.D., Tippani, M., Ramnauth, A.D., Divecha, H.R., Miller, R.A., Eagles, N.J., Pattie, E.A., Kwon, S.H., Bach, S.V., et al. (2025). An integrated single-nucleus and spatial transcriptomics atlas reveals the molecular landscape of the human hippocampus. Nat. Neurosci. *28*, 1990–2004. https://doi.org/10.1038/s41593-025-02022-0.

45. Bakken, T.E., van Velthoven, C.T., Menon, V., Hodge, R.D., Yao, Z., Nguyen, T.N., Graybuck, L.T., Horwitz, G.D., Bertagnolli, D., Goldy, J., et al. (2021). Single-cell and single-nucleus RNA-seq uncovers shared and distinct axes of variation in dorsal LGN neurons in mice, non-human primates, and humans. eLife *10*, e64875. https://doi.org/10.7554/eLife.64875.

46. Liu, X.-P., Dalley, R., Johansen, N., Budzillo, A., Thijssen, J., Miller, J.A., Walling-Bell, S., Sawchuk, S., Alfiler, L., Andrade, J., et al. (2026). Morphoelectric Diversity and Specialization of Neuronal Cell Types in the Primate Striatum. Preprint at bioRxiv. https://doi.org/10.64898/2026.02.26.708019.

47. Kosicki, M., Baltoumas, F.A., Kelman, G., Boverhof, J., Ong, Y., Cook, L.E., Dickel, D.E., Pavlopoulos, G.A., Pennacchio, L.A., and Visel, A. (2025). VISTA Enhancer browser: an updated database of tissue-specific developmental enhancers. Nucleic Acids Res. *53*, D324–D330. https://doi.org/10.1093/nar/gkae940.

48. The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

49. Zhang, K., Zemke, N.R., Armand, E.J., and Ren, B. (2024). A fast, scalable and versatile tool for analysis of single-cell omics data. Nat. Methods *21*, 217–227. https://doi.org/10.1038/s41592-023-02139-9.

50. Hulselmans, G., Aibar, S., Flerin, C.C., Seppe, cbravo, flochlay, swann, Burkhardt, D., and Mannens, C. (2023). aertslab/pycisTopic: pycisTopic at date of publication. Zenodo. https://doi.org/10.5281/zenodo.7857024.

51. Suresh, H., Crow, M., Jorstad, N., Hodge, R., Lein, E., Dobin, A., Bakken, T., and Gillis, J. (2023). Comparative single-cell transcriptomic analysis of primate brains highlights human-specific regulatory evolution. Nat. Ecol. Evol. *7*, 1930–1943. https://doi.org/10.1038/s41559-023-02186-7.

52. Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science *362*, eaau5324. https://doi.org/10.1126/science.aau5324.

53. Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. *20*, 484–496. https://doi.org/10.1038/nn.4495.

54. Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat. Methods *20*, 1355–1367. https://doi.org/10.1038/s41592-023-01938-4.

55. Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A., and Boncinelli, E. (1992). Nested expression domains of four homeobox genes in developing rostral brain. Nature *358*, 687–690. https://doi.org/10.1038/358687a0.

56. Martinez, S., Crossley, P.H., Cobos, I., Rubenstein, J.L.R., and Martin, G.R. (1999). FGF8 induces formation of an ectopic isthmic organizer and isthmocerebellar development via a repressive effect on Otx2 expression. Development *126*, 1189–1200. https://doi.org/10.1242/dev.126.6.1189.

57. Kiecker, C., and Lumsden, A. (2005). Compartments and their boundaries in vertebrate brain development. Nat. Rev. Neurosci. *6*, 553–564. https://doi.org/10.1038/nrn1702.

58. Luo, J., Liu, K., Yao, Y., Sun, Q., Zheng, X., Zhu, B., Zhang, Q., Xu, L., Shen, Y., and Ren, B. (2019). DMBX1 promotes tumor proliferation and regulates cell cycle progression via repressing OTX2-mediated transcription of p21 in lung adenocarcinoma cell. Cancer Lett. *453*, 45–56. https://doi.org/10.1016/j.canlet.2019.03.045.

59. 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.

60. Schroeder, M.E., McCormack, D.M., Metzner, L.R., Kang, J., Li, K.X., Yu, E., Melamed, L., Levandowski, K.M., Zaniewski, H., Zhang, Q., et al. (2025). A transcriptomic atlas of astrocyte heterogeneity across space and time in mouse and marmoset. Neuron *113*, 3942–3965.e19. https://doi.org/10.1016/j.neuron.2025.09.011.

61. Khakh, B.S. (2025). On astrocyte-neuron interactions: Broad insights from the striatum. Neuron *113*, 3079–3107. https://doi.org/10.1016/j.neuron.2025.08.009.

62. Fu, Y., Johansen, N.J., Kempynck, N., Ding, W., Turner, M.A., Garcia, A.D., Schmitz, M.T., Close, J., Kapen, I., Hewitt, M., et al. (2025). Circuit specific specialization of human basal ganglia astrocytes. Preprint at bioRxiv. https://doi.org/10.64898/2025.12.19.695583.

63. 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.

64. Xuan, S., Baptista, C.A., Balas, G., Tao, W., Soares, V.C., and Lai, E. (1995). Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. Neuron *14*, 1141–1152. https://doi.org/10.1016/0896-6273(95)90262-7.

65. Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A., et al. (2015). Molecular Identity of Human Outer Radial Glia during Cortical Development. Cell *163*, 55–67. https://doi.org/10.1016/j.cell.2015.09.004.

66. Fu, Y., Yang, M., Yu, H., Wang, Y., Wu, X., Yong, J., Mao, Y., Cui, Y., Fan, X., Wen, L., et al. (2021). Heterogeneity of glial progenitor cells during the neurogenesis-to-gliogenesis switch in the developing human cerebral cortex. Cell Rep. *34*, 108788. https://doi.org/10.1016/j.celrep.2021.108788.

67. Hébert, J.M., and Fishell, G. (2008). The genetics of early telencephalon patterning: some assembly required. Nat. Rev. Neurosci. *9*, 678–685. https://doi.org/10.1038/nrn2463.

68. Marín, O., Yaron, A., Bagri, A., Tessier-Lavigne, M., and Rubenstein, J.L.R. (2001). Sorting of Striatal and Cortical Interneurons Regulated by Semaphorin-Neuropilin Interactions. Science *293*, 872–875. https://doi.org/10.1126/science.1061891.

69. Mayer, C., Jaglin, X.H., Cobbs, L.V., Bandler, R.C., Streicher, C., Cepko, C.L., Hippenmeyer, S., and Fishell, G. (2015). Clonally Related Forebrain Interneurons Disperse Broadly across Both Functional Areas and Structural Boundaries. Neuron *87*, 989–998. https://doi.org/10.1016/j.neuron.2015.07.011.

70. Bandler, R.C., Vitali, I., Delgado, R.N., Ho, M.C., Dvoretskova, E., Ibarra Molinas, J.S., Frazel, P.W., Mohammadkhani, M., Machold, R., Maedler, S., et al. (2022). Single-cell delineation of lineage and genetic identity in the mouse brain. Nature *601*, 404–409. https://doi.org/10.1038/s41586-021-04237-0.

71. Harwell, C.C., Fuentealba, L.C., Gonzalez-Cerrillo, A., Parker, P.R.L., Gertz, C.C., Mazzola, E., Garcia, M.T., Alvarez-Buylla, A., Cepko, C.L., and Kriegstein, A.R. (2015). Wide Dispersion and Diversity of Clonally Related Inhibitory Interneurons. Neuron *87*, 999–1007. https://doi.org/10.1016/j.neuron.2015.07.030.

72. Marín, O., and Rubenstein, J.L.R. (2003). Cell migration in the forebrain. Annu. Rev. Neurosci. *26*, 441–483. https://doi.org/10.1146/annurev.neuro.26.041002.131058.

73. Nóbrega-Pereira, S., Kessaris, N., Du, T., Kimura, S., Anderson, S.A., and Marín, O. (2008). Postmitotic Nkx2-1 Controls the Migration of Telencephalic Interneurons by Direct Repression of Guidance Receptors. Neuron *59*, 733–745. https://doi.org/10.1016/j.neuron.2008.07.024.

74. Du, T., Xu, Q., Ocbina, P.J., and Anderson, S.A. (2008). NKX2.1 specifies cortical interneuron fate by activating Lhx6. Development *135*, 1559–1567. https://doi.org/10.1242/dev.015123.

75. Zhao, Y., Marín, O., Hermesz, E., Powell, A., Flames, N., Palkovits, M., Rubenstein, J.L.R., and Westphal, H. (2003). The LIM-homeobox gene Lhx8 is required for the development of many cholinergic neurons in the mouse forebrain. Proc. Natl. Acad. Sci. USA *100*, 9005–9010. https://doi.org/10.1073/pnas.1537759100.

76. Liodis, P., Denaxa, M., Grigoriou, M., Akufo-Addo, C., Yanagawa, Y., and Pachnis, V. (2007). Lhx6 activity is required for the normal migration and specification of cortical interneuron subtypes. J. Neurosci. *27*, 3078–3089. https://doi.org/10.1523/JNEUROSCI.3055-06.2007.

77. Close, J., Xu, H., De Marco García, N., Batista-Brito, R., Rossignol, E., Rudy, B., and Fishell, G. (2012). Satb1 is an activity-modulated transcription factor required for the terminal differentiation and connectivity of medial ganglionic eminence-derived cortical interneurons. J. Neurosci. *32*, 17690–17705. https://doi.org/10.1523/JNEUROSCI.3583-12.2012.

78. Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron *57*, 378–392. https://doi.org/10.1016/j.neuron.2007.12.028.

79. Mayer, C., Hafemeister, C., Bandler, R.C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G., and Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. Nature *555*, 457–462. https://doi.org/10.1038/nature25999.

80. Spiegel, I., Mardinly, A.R., Gabel, H.W., Bazinet, J.E., Couch, C.H., Tzeng, C.P., Harmin, D.A., and Greenberg, M.E. (2014). Npas4 Regulates Excitatory-Inhibitory Balance within Neural Circuits through Cell-Type-Specific Gene Programs. Cell *157*, 1216–1229. https://doi.org/10.1016/j.cell.2014.03.058.

81. Andrews, W., Barber, M., Hernadez-Miranda, L.R., Xian, J., Rakic, S., Sundaresan, V., Rabbitts, T.H., Pannell, R., Rabbitts, P., Thompson, H., et al. (2008). The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons. Dev. Biol. *313*, 648–658. https://doi.org/10.1016/j.ydbio.2007.10.052.

82. Bright, A.R., Kotlyarenko, Y., Neuhaus, F., Rodrigues, D., Feng, C., Peters, C., Vitali, I., Dönmez, E., Myoga, M.H., Dvoretskova, E., et al. (2025). Temporal control of progenitor competence shapes maturation in GABAergic neuron development in mice. Nat. Neurosci. *28*, 1663–1675. https://doi.org/10.1038/s41593-025-01999-y.

83. Puelles, L., Kuwana, E., Puelles, E., Bulfone, A., Shimamura, K., Keleher, J., Smiga, S., and Rubenstein, J.L. (2000). Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. J. Comp. Neurol. *424*, 409–438. https://doi.org/10.1002/1096-9861(20000828)424:3<409::AID-CNE3>3.0.CO;2-7.

84. Flames, N., and Marín, O. (2005). Developmental mechanisms underlying the generation of cortical interneuron diversity. Neuron *46*, 377–381. https://doi.org/10.1016/j.neuron.2005.04.020.

85. Li, L., Yun, S.H., Keblesh, J., Trommer, B.L., Xiong, H., Radulovic, J., and Tourtellotte, W.G. (2007). Egr3, a synaptic activity regulated transcription factor that is essential for learning and memory. Mol. Cell. Neurosci. *35*, 76–88. https://doi.org/10.1016/j.mcn.2007.02.004.

86. Yamagata, K., Kaufmann, W.E., Lanahan, A., Papapavlou, M., Barnes, C.A., Andreasson, K.I., and Worley, P.F. (1994). Egr3/Pilot, a zinc finger transcription factor, is rapidly regulated by activity in brain neurons and colocalizes with Egr1/zif268. Learn. Mem. *1*, 140–152.

87. McClellan, K.A., Ruzhynsky, V.A., Douda, D.N., Vanderluit, J.L., Ferguson, K.L., Chen, D., Bremner, R., Park, D.S., Leone, G., and Slack, R.S. (2007). Unique Requirement for Rb/E2F3 in Neuronal Migration: Evidence for Cell Cycle-Independent Functions. Mol. Cell. Biol. *27*, 4825–4843. https://doi.org/10.1128/MCB.02100-06.

88. Maksour, S., Ooi, L., and Dottori, M. (2020). More than a Corepressor: The Role of CoREST Proteins in Neurodevelopment. eNeuro *7*. ENEURO.0337-19.2020. https://doi.org/10.1523/ENEURO.0337-19.2020.

89. Evans, A.E., Kelly, C.M., Precious, S.V., and Rosser, A.E. (2012). Molecular regulation of striatal development: a review. Anat. Res. Int. *2012*, 106529. https://doi.org/10.1155/2012/106529.

90. Harris, H.K., Nakayama, T., Lai, J., Zhao, B., Argyrou, N., Gubbels, C.S., Soucy, A., Genetti, C.A., Suslovitch, V., Rodan, L.H., et al. (2021). Disruption of RFX family transcription factors causes autism, attention-deficit/hyperactivity disorder, intellectual disability, and dysregulated behavior. Genet. Med. *23*, 1028–1040. https://doi.org/10.1038/s41436-021-01114-z.

91. Gray, L.T., Yao, Z., Nguyen, T.N., Kim, T.K., Zeng, H., and Tasic, B. (2017). Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. eLife *6*, e21883. https://doi.org/10.7554/eLife.21883.

92. Palmer, C.R., Song, J., Yang, B., Chen, C.-J., Diep, D., Conklin, K., Plongthongkum, N., Indralingam, H.S., Liu, C.S., Kurtz, J., et al. (2026). Single-cell multiomic human brain atlas reveals regulatory drivers of cortical regionality. Nat. Commun. *17*, 3051. https://doi.org/10.1101/2025.08.07.667992.

93. Charizanis, K., Lee, K.-Y., Batra, R., Goodwin, M., Zhang, C., Yuan, Y., Shiue, L., Cline, M., Scotti, M.M., Xia, G., et al. (2012). Muscleblind-like 2-Mediated Alternative Splicing in the Developing Brain and Dysregulation in Myotonic Dystrophy. Neuron *75*, 437–450. https://doi.org/10.1038/s41467-026-69368-2.

94. Szklarczyk, D., Nastou, K., Koutrouli, M., Kirsch, R., Mehryary, F., Hachilif, R., Hu, D., Peluso, M.E., Huang, Q., Fang, T., et al. (2025). The STRING database in 2025: protein networks with directionality of regulation. Nucleic Acids Res. *53*, D730–D737. https://doi.org/10.1093/nar/gkae1113.

95. Li, W., and Pozzo-Miller, L. (2020). Dysfunction of the corticostriatal pathway in autism spectrum disorders. J. Neurosci. Res. *98*, 2130–2147. https://doi.org/10.1002/jnr.24560.

96. Pressl, C., Mätlik, K., Kus, L., Darnell, P., Luo, J.-D., Paul, M.R., Weiss, A.R., Liguore, W., Carroll, T.S., Davis, D.A., et al. (2024). Selective vulnerability of layer 5a corticostriatal neurons in Huntington's disease. Neuron *112*, 924–941.e10. https://doi.org/10.1016/j.neuron.2023.12.009.

97. Handsaker, R.E., Kashin, S., Reed, N.M., Tan, S., Lee, W.-S., McDonald, T.M., Morris, K., Kamitaki, N., Mullally, C.D., Morakabati, N.R., et al. (2025). Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease. Cell *188*, 623–639.e19. https://doi.org/10.1016/j.cell.2024.11.038.

98. Hedreen, J.C., Peyser, C.E., Folstein, S.E., and Ross, C.A. (1991). Neuronal loss in layers V and VI of cerebral cortex in Huntington's disease. Neurosci. Lett. *133*, 257–261. https://doi.org/10.1016/0304-3940(91)90583-F.

99. Shepherd, G.M.G. (2013). Corticostriatal connectivity and its role in disease. Nat. Rev. Neurosci. *14*, 278–291. https://doi.org/10.1038/nrn3469.

100. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A.,

et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

101. Lee, S.M.K., Tole, S., Grove, E.A., and McMahon, A.P. (2000). A local Wnt-3a signal is required for development of the mammalian hippocampus. Development *127*, 457–467. https://doi.org/10.1242/dev.127.3.457.

102. Cohen, L., Halpern, A., Blosser, T.R., Che, P., Pan, X., and Zhuang, X. (2026). Whole-transcriptome-scale and isoform-resolved spatial imaging of single cells in complex tissues. Cell *189*, 6189–6207.e10. https://doi.org/10.1016/j.cell.2026.06.027.

103. Wyss, J.M., and Sripanidkulchai, K. (1983). The indusium griseum and anterior hippocampal continuation in the rat. J. Comp. Neurol. *219*, 251–272. https://doi.org/10.1002/cne.902190302.

104. Laplante, F., Mnie-Filali, O., and Sullivan, R.M. (2013). A neuroanatomical and neurochemical study of the indusium griseum and anterior hippocampal continuation: Comparison with dentate gyrus. J. Chem. Neuroanat. *50–51*, 39–47. https://doi.org/10.1016/j.jchemneu.2013.03.004.

105. Yang, X., Wan, R., Liu, Z., Feng, S., Yang, J., Jing, N., and Tang, K. (2023). The differentiation and integration of the hippocampal dorsoventral axis are controlled by two nuclear receptor genes. eLife *12*, RP86940. https://doi.org/10.7554/eLife.86940.

106. Thompson, C.L., Pathak, S.D., Jeromin, A., Ng, L.L., MacPherson, C.R., Mortrud, M.T., Cusick, A., Riley, Z.L., Sunkin, S.M., Bernard, A., et al. (2008). Genomic Anatomy of the Hippocampus. Neuron *60*, 1010–1021. https://doi.org/10.1016/j.neuron.2008.12.008.

107. Vogel, J.W., La Joie, R., Grothe, M.J., Diaz-Papkovich, A., Doyle, A., Vachon-Presseau, E., Lepage, C., Vos de Wael, R., Thomas, R.A., Iturria-Medina, Y., et al. (2020). A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems. Nat. Commun. *11*, 960. https://doi.org/10.1038/s41467-020-14518-3.

108. Liu, X., Zeira, R., and Raphael, B.J. (2023). Partial alignment of multislice spatially resolved transcriptomics data. Genome Res. *33*, 1124–1132. https://doi.org/10.1101/gr.277670.123.

109. Ayhan, F., Kulkarni, A., Berto, S., Sivaprakasam, K., Douglas, C., Lega, B.C., and Konopka, G. (2021). Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. Neuron *109*, 2091–2105.e6. https://doi.org/10.1016/j.neuron.2021.05.003.

110. Sagner, A., and Briscoe, J. (2019). Establishing neuronal diversity in the spinal cord: a time and a place. Development *146*, dev182154. https://doi.org/10.1242/dev.182154.

111. Pattabiraman, K., Golonzhka, O., Lindtner, S., Nord, A.S., Taher, L., Hoch, R., Silberberg, S.N., Zhang, D., Chen, B., Zeng, H., et al. (2014). Transcriptional regulation of enhancers active in protodomains of the developing cerebral cortex. Neuron *82*, 989–1003. https://doi.org/10.1016/j.neuron.2014.04.014.

112. Jager, P., Ye, Z., Yu, X., Zagoraiou, L., Prekop, H.-T., Partanen, J., Jessell, T.M., Wisden, W., Brickley, S.G., and Delogu, A. (2016). Tectal-derived interneurons contribute to phasic and tonic inhibition in the visual thalamus. Nat. Commun. *7*, 13579. https://doi.org/10.1038/ncomms13579.

113. Hunt, C.A., Pang, D.Z., and Jones, E.G. (1991). Distribution and density of GABA cells in intralaminar and adjacent nuclei of monkey thalamus. Neuroscience *43*, 185–196. https://doi.org/10.1016/0306-4522(91)90426-O.

114. Dixon, G., and Harper, C.G. (2001). Quantitative analysis of glutamic acid decarboxylase-immunoreactive neurons in the anterior thalamus of the human brain. Brain Res. *923*, 39–44. https://doi.org/10.1016/S0006-8993(01)03194-8.

115. Jones, E.G. (2002). Dichronous appearance and unusual origins of GABA neurons during development of the mammalian thalamus. Thalamus Relat. Syst. *1*, 283–288. https://doi.org/10.1016/S1472-9288(02)00002-X.

116. Petanjek, Z., Kostović, I., and Esclapez, M. (2009). Primate-Specific Origins and Migration of Cortical GABAergic Neurons. Front. Neuroanat. *3*, 26. https://doi.org/10.3389/neuro.05.026.2009.

117. Kim, C.N., Shin, D., Wang, A., and Nowakowski, T.J. (2023). Spatiotemporal molecular dynamics of the developing human thalamus. Science *382*, eadf9941. https://doi.org/10.1126/science.adf9941.

118. Tarashansky, A.J., Musser, J.M., Khariton, M., Li, P., Arendt, D., Quake, S.R., and Wang, B. (2021). Mapping single-cell atlases throughout Metazoa unravels cell type evolution. eLife *10*, e66747. https://doi.org/10.7554/eLife.66747.

119. Guillamón-Vivancos, T., Favaloro, F., Dori, F., and López-Bendito, G. (2024). The superior colliculus: New insights into an evolutionarily ancient structure. Curr. Opin. Neurobiol. *89*, 102926. https://doi.org/10.1016/j.conb.2024.102926.

120. Li, Y., Li, Z., Wang, C., Yang, M., He, Z., Wang, F., Zhang, Y., Li, R., Gong, Y., Wang, B., et al. (2023). Spatiotemporal transcriptome atlas reveals the regional specification of the developing human brain. Cell *186*, 5892–5909.e22. https://doi.org/10.1016/j.cell.2023.11.016.

121. Sarra, C., Sarra, L., Di Carlo, L., GrandPre, T., Zhang, Y., Callan, C.G., and Bialek, W. (2025). Maximum entropy models for patterns of gene expression. Phys. Rev. E *112*, 014408. https://doi.org/10.1103/wjcn-l4ms.

122. Hahn, J., Monavarfeshani, A., Qiao, M., Kao, A.H., Kölsch, Y., Kumar, A., Kunze, V.P., Rasys, A.M., Richardson, R., Wekselblatt, J.B., et al. (2023). Evolution of neuronal cell classes and types in the vertebrate retina. Nature *624*, 415–424. https://doi.org/10.1038/s41586-023-06638-9.

123. Puelles, L., and Rubenstein, J.L.R. (2003). Forebrain gene expression domains and the evolving prosomeric model. Trends Neurosci. *26*, 469–476. https://doi.org/10.1016/S0166-2236(03)00234-0.

124. O'Leary, D.D.M., Chou, S.-J., and Sahara, S. (2007). Area Patterning of the Mammalian Cortex. Neuron *56*, 252–269. https://doi.org/10.1016/j.neuron.2007.10.010.

125. Abdelhady, G., Brull, O.R., He, J., Lin, M.K., Galvan, A., Pfenning, A.R., Bostan, A.C., and Stauffer, W.R. (2025). A Multimodal Atlas Reveals the Anatomical Distribution of Medium Spiny Neuron Subtypes and a Novel RGS6+ Population in the Primate Striatum. Preprint at bioRxiv. https://doi.org/10.1101/2025.11.16.688724.

126. Kempynck, N., De Winter, S., Blaauw, C.H., Konstantakos, V., Ekşi, E.C., Dieltiens, S., Abaffyová, D., Bercier, V., Taskiran, I.I., Hulselmans, G., et al. (2026). CREsted: modeling genomic and synthetic cell-type-specific enhancers across tissues and species. Nat. Methods *23*, 946–959. https://doi.org/10.1038/s41592-026-03057-2.

127. Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366. https://doi.org/10.1038/s41588-021-00782-6.

128. Pena, C. (2026). GA. Created in BioRender. https://BioRender.com/yi16vmv.

129. Pena, C. (2026). 1. Created in BioRender. https://BioRender.com/s8shsrh.

130. 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.

131. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDblFinder. F1000Res *10*, 979. https://doi.org/10.12688/f1000research.73600.2.

132. 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.

133. Aevermann, B.D., Zhang, Y., Novotny, M., Keshk, M., Bakken, T.E., Miller, J.A., Hodge, R.D., Lelieveldt, B., Lein, E.S., and Scheuermann, R.H. (2021 Jun). A machine learning method for the discovery of minimum marker gene combinations for cell-type identification from single-cell RNA sequencing. Genome Research *4*, gt.275569.121. https://doi.org/10.1101/gr.275569.121.

134. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Wieser, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

135. Zhu, A., Ibrahim, J.G., and Love, M.I. (2018). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895.

136. Aibar, S., Bravo Gonzalez-Blas, C., Moerman, T., Huynh-Thu, V., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J., Geurts, P., Aerts, J., et al. (2017). SCENIC: Single-Cell Regulatory Network Inference And Clustering. Nature Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

137. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

138. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. J. Stat. Softw. *25*, 1–18. https://doi.org/10.18637/jss.v025.i01.

139. Servén, D., and Brummitt, C. (2018). pyGAM: Generalized Additive Models in Python. Zenodo. https://doi.org/10.5281/zenodo.1208723.

140. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

141. Paxinos, G. (2012). The Marmoset Brain in Stereotaxic Coordinates, First Edition (Academic Press).

142. Miller, J. (2025). mfishtools: Building Gene Sets and Mapping mFISH Data. GitHub. https://github.com/AllenInstitute/mfishtools.

143. Missarova, A., Jain, J., Butler, A., Ghazanfar, S., Stuart, T., Brusko, M., Wasserfall, C., Nick, H., Brusko, T., Atkinson, M., et al. (2021). geneBasis: an iterative approach for unsupervised selection of targeted gene panels from scRNA-seq. Genome Biol. *22*, 333. https://doi.org/10.1186/s13059-021-02548-z.

144. DeBerardine, M. (2025). mCalJa1.2.pat.X_mitos2. Zenodo. https://doi.org/10.5281/zenodo.16915694.

145. 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.

146. Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., Brown, S., Capodiferro, M.R., Al-Ajli, F.O., Ambrosini, R., et al. (2021). Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biol. *22*, 120. https://doi.org/10.1186/s13059-021-02336-9.

147. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

148. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.

149. Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. Curr. Protoc. Bioinform. *69*, e96. https://doi.org/10.1002/cpbi.96.

150. Campitelli, E. (2019). ggnewscale: Multiple Fill and Color Scales in ggplot2. Zenodo. https://doi.org/10.5281/zenodo.2543772.

151. Liu, A., Peng, B., Pankajam, A.V., Duong, T.E., Pryhuber, G., Scheuermann, R.H., and Zhang, Y. (2024). Discovery of optimal cell type classification marker genes from single cell RNA sequencing data. BMC Methods *1*, 15. https://doi.org/10.1186/s44330-024-00015-2.

152. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084–2092. https://doi.org/10.1093/bioinformatics/bty895.

153. Shen, W.-K., Chen, S.-Y., Gan, Z.-Q., Zhang, Y.-Z., Yue, T., Chen, M.-M., Xue, Y., Hu, H., and Guo, A.-Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. *51*, D39–D45. https://doi.org/10.1093/nar/gkac907.

154. Allen Institute. (2023). HOMBA ontology — Common Coordinate Framework and Multi-Species Atlas Primer (CCF-MAP). https://alleninstitute.github.io/CCF-MAP/docs/HOMBA_ontology_v1.html.

155. Hie, B., Cho, H., DeMeo, B., Bryson, B., and Berger, B. (2019). Geometric Sketching Compactly Summarizes the Single-Cell Transcriptomic Landscape. Cell Syst. *8*, 483–493.e7. https://doi.org/10.1016/j.cels.2019.05.003.

156. Winter, S.D., and cbravo. (2025). aertslab/pycistarget: pycistarget v1.1. Zenodo. https://doi.org/10.5281/zenodo.14627914.

157. Pedersen, T.L. (2025). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. Version R package version 2.2.2.9000. https://cran.r-project.org/web/packages/ggraph/ggraph.pdf.

158. Vieth, A., Kroes, T., Thijssen, J., van Lew, B., Eggermont, J., Basu, S., Eisemann, E., Vilanova, A., Höllt, T., and Lelieveldt, B. (2024). ManiVault: A Flexible and Extensible Visual Analytics Framework for High-Dimensional Data. IEEE Trans. Vis. Comput. Graph. *30*, 175–185. https://doi.org/10.1109/TVCG.2023.3326582.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-NeuN-Phycoerythrin (clone A60)	Millipore Sigma	Cat# FCMAB317PE; RRID: AB_11212465
Rabbit monoclonal anti-Olig2-Alexa Fluor 488	Abcam	Cat# ab225099; RRID: AB_3714884
Mouse IgG1, κ Isotype Control, R-Phycoerythrin	BD Biosciences	Cat# 555749; RRID: AB_396091
Rabbit IgG, monoclonal Isotype Control, Alexa Fluor 488	Abcam	Cat# ab199091; RRID: AB_2637027
Biological samples		
Marmoset adult brain tissue	Massachusetts Institute of Technology; Feng Laboratory	N/A
Chemicals, peptides, and recombinant proteins		
Iodixanol 60% (v/v), OptiPrep™ Density Gradient Medium	MilliporeSigma	Cat# D1556
DAPI	Thermo Fisher Scientific	Cat# D1306
Protector RNase Inhibitor	MilliporeSigma	Cat# 3335402001
Protease Inhibitor Cocktail	Thermo Fisher Scientific	Cat# 78437
Triton X-100	MilliporeSigma	Cat# T8787
Bovine Serum Albumin 30%	MilliporeSigma	Cat# 126625
Critical commercial assays		
Chromium X Instrument	10X Genomics	Cat# PN-1000326; RRID: SCR_024537
Chromium Next GEM Single Cell Multiome ATAC + Gene Expression	10X Genomics	Cat# PN-CG000338
Xenium In Situ Gene Expression assay	10X Genomics	Cat# PN-CG000582
Xenium Cell Segmentation Kit	10X Genomics	Cat# PN-1000661
Xenium Analyzer Instrument, v3.0.2.0	10X Genomics	Cat# PN-1000481; RRID: SCR_023910
MA900 Multi-Application Cell Sorter	Sony	Cat# LE-MA900FP1; RRID: SCR_026300
TapeStation 4150	Agilent	Cat# G2992AA; RRID: SCR_019393
Qubit Flex Fluorometer	Thermo Fisher Scientific	Cat# Q33327; RRID: SCR_020553
NovaSeq X Plus	Illumina	Cat# 20084804; RRID: SCR_024568
Xenium v1 Standalone Custom 101 to 300 Gene Panel (16 rxn)	10X Genomics	Cat# PN-1000563; Panel ID: EU82Q4
Deposited data		
10X Genomics Multiome (Raw)	This paper	NeMO: dat-kkuxhz4
Marmoset reference genome mCalJa1.2.pat.X_mitos2	This paper; DeBerardine	https://zenodo.org/records/16915694
10X Genomics Multiome (Processed)	This paper	https://brain-map.org/consortia/hmba, https://allen-hmba-releases.s3.us-west-2.amazonaws.com/index.html#marmoset_subcortex/
Adult mouse whole brain atlas (ABC Atlas)	Yao et al.[1]	https://allen-brain-cell-atlas.s3.us-west-2.amazonaws.com/index.html
Adult human brain atlas	Siletti et al.[3]	https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Adult mouse whole brain atlas (Macosko Lab)	Langlieb et al.[2]	https://docs.braincelldata.org/downloads/index.html
Human hypothalamus HypoMap atlas	Tadross et al.[42]	https://doi.org/10.17863/CAM.111988
Mouse hypothalamus HypoMap atlas	Steuernagel et al.[41]	https://doi.org/10.17863/CAM.87955
Human Lateral Septum atlas	Phillips et al.[43]	https://research.libd.org/globus/jhpce_ls_molecular-profiling_sce/index.html
Human Hippocampus atlas	Thompson et al.[44]	Bioconductor: humanHippocampus2024
Cross-species Lateral Geniculate atlas	Bakken et al.[45]	https://portal.brain-map.org/atlases-and-data/rnaseq/comparative-lgn
Primate Amygdala atlas	Totty et al.[40]	GEO: GSE281366
Marmoset brain cell census atlas	Krienen et al.[20]	https://cellxgene.cziscience.com/collections/0fd39ad7-5d2d-41c2-bda0-c55bde614bdb
Experimental models: Organisms/strains		
C. jacchus	Massachusetts Institute of Technology; Feng Laboratory, https://fenglaboratory.org/	NCBI taxonomy: txid9483
Software and algorithms		
NeMO: The Neuroscience Multi-omic Data Archive	Institute for Genome Sciences, University of Maryland School of Medicine	https://nemoarchive.org/; RRID: SCR_016152
NIMP: Neuroanatomy-anchored Information Management Platform for Collaborative BICAN Data Generation	BRAIN Initiative Cell Atlas Network	https://brain-specimenportal.org; RRID: SCR_024684
Xenium Analyzer Onboard Analysis, v-xenium-3.0.0.15	10X Genomics	Cat# PN-1000481; RRID: SCR_026158
RAPIDS AI 25.4.0	RAPIDS Development Team	https://rapids.ai
CellRanger-ARC v2.0.2	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
CellBender v0.3.0	Fleming et al.[130]	https://github.com/broadinstitute/CellBender; RRID: SCR_025990
scDblFinder 1.21.0	Germain et al.[131]	https://bioconductor.org/packages/release/bioc/html/scDblFinder.html; RRID: SCR_022700
scanpy 1.10.4	Wolf et al.[132]	https://scanpy.readthedocs.io/; RRID: SCR_018139
scVI v1.2.2.post2	Lopez et al.[34]	https://scvi-tools.org; RRID: SCR_026673
transcriptomic_clustering v1.0.0	Allen Institute for Brain Science	https://github.com/AllenInstitute/transcriptomic_clustering
SnapATAC2 v2.8.0	Zhang et al.[49]	https://scverse.org/SnapATAC2/; RRID: SCR_026622
pycistarget v1.1	Bravo Gonzalez-Blas and De Winter et al.[54]	https://github.com/aertslab/pycistarget; RRID: SCR_026626
SCENIC+ v1.0a2	Bravo González-Blas and De Winter et al.[54]	https://github.com/aertslab/scenicplus; RRID: SCR_026702
NSForest v4.1	Aevermann et al.[133]	https://nsforest.readthedocs.io/en/stable/index.html
MapMyCells	Daniel et al.[39]	https://brain-map.org/bkp/analyze/mapmycells#use-mmc; RRID: SCR_024672
PASTE2	Liu et al.[108]	https://github.com/raphael-group/paste2
Cytosplore Viewer	Allen Institute for Brain Science	https://viewer.cytosplore.org/; RRID: SCR_018330

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
numpy 2.0.2	Harris, C.R., Millman, K.J., van der Walt, S.J. et al.[134]	https://numpy.org/install
pandas 2.2.3	The pandas development team	Zenodo https://doi.org/10.5281/zenodo.23065223
pycisTopic v2.0a0	Gonzalez-Blas and De Winter et al.[54]	https://github.com/aertslab/pycisTopic; RRID: SCR_026618
Python 3.10.18	Python Software Foundation	https://www.python.org
Python 3.13.7	Python Software Foundation	https://www.python.org
scikit-learn v1.5.2	scikit-learn developers	https://scikit-learn.org/
scikit-learn v1.8.0	scikit-learn developers	https://scikit-learn.org/
STRING v12.0	STRING Consortium	https://version-12-0.string-db.org/; RRID: SCR_005223
apeglm v1.28.0	Zhu et al.[135]	https://bioconductor.org/packages/apeglm/; RRID: SCR_026951
AUCell v1.28.0	Aibar et al.[136]	https://bioconductor.org/packages/AUCell/; RRID: SCR_021327
DESeq2 v1.46.0	Love et al.[137]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html; RRID: SCR_015687
Expressolog	Suresh et al.[51]	https://www.nature.com/articles/s41559-023-02186-7#citeas
FactoMineR v2.11	Lê et al.[138]	https://cran.r-project.org/web/packages/FactoMineR/index.html
pyGAM v0.12.0	Servén D., Brummitt C.[139]	https://pygam.readthedocs.io/en/latest/; Zenodo. https://doi.org/10.5281/zenodo.1208723
IGV desktop v2.16.0	Robinson et al.[140]	https://igv.org/doc/desktop/; RRID: SCR_011793
BioRender		https://www.biorender.com; RRID: SCR_018361
Other		
30 μm cell strainer	Miltenyi Biotec	Cat# 130-098-458
70 μm cell strainer	Miltenyi Biotec	Cat# 130-098-462
Disposable hemocytometer (Neubauer Improved)	INCYTO	Cat# DHC-N01
KIMBLE Dounce tissue grinder set	MilliporeSigma	Cat# D8938

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Marmosets

Marmoset experiments were approved, overseen, and conducted in compliance with the Massachusetts Institute of Technology IACUC under protocol number 2303000479. Adult marmosets (*Callithrix jacchus*; 5 individuals, 3 male and 2 female, 2-7.5 years old with a median age of 5.6 years; Table S1) used for the study were generally in good health condition, were experimentally naïve, had not experienced major surgery affecting the brain, and had not received experimental drugs or treatments prior to the study. The animals were group housed in family units consisting of 2 adult animals and offspring up to the age of 15 months. Animals were housed in a 12h/12h light/dark cycle. The MIT Division of Comparative Medicine (DCM) routinely monitored animal welfare and provided veterinary care as needed.

METHOD DETAILS

Tissue processing

Marmosets (Table S1) were initially sedated with Alfaxalone (12mg/kg, 10 mg/ml) and Midazolam (0.3 mg/kg, 5mg/ml) via intramuscular injection. They were further sedated with a secondary dose of Alfaxalone (4mg/kg, 10mg/ml) to ensure deep sedation followed by an intravenous injection of Euthasol (>120mg/kg, 390 mg/ml). When respiration and the pedal withdrawal reflex were eliminated,

the marmosets were transcardially perfused with ice-cold carbogenated *N*-methyl-D-glucamine (NMDG) artificial cerebrospinal fluid (aCSF) (92 mM NMDG, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 30 mM $NaHCO_3$, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM sodium L-ascorbate, 3 mM sodium pyruvate, 0.5 mM $CaCl_2 \cdot 2H_2O$, and 10 mM $MgSO_4 \cdot 7H_2O$; pH 7.3-7.4 adjusted with HCl). Brains were extracted into ice-cold NMDG aCSF. The brainstem, cerebellum, and olfactory bulb were manually separated using a sterile dissecting knife. The brainstem was bisected sagittally and then cut into 3-4 mm sections at an angle perpendicular to the long axis. The cerebellum was bisected sagittally and then cut into 3-4mm sagittal sections, ensuring to leave the deep cerebellar nuclei, vermis, and cerebellar hemispheres intact.

The remaining cerebrum was placed dorsal-side-down into a custom-made steel marmoset brain matrix for sectioning in ice-cold HEPES-sucrose solution (110 mM NaCl, 10 mM HEPES, 25 mM glucose, 75 mM sucrose, 7.5 mM $MgCl_2 \cdot 6H_2O$, 2.5 mM KCl; pH 7.4 adjusted with NaOH; ~350 mOsm). Coronal sections were made using a razor blade at 2mm intervals from posterior to anterior for multiomic samples and 5mm intervals from posterior to anterior for spatial transcriptomic samples. The slabs were arranged in order on a teflon coated metal plate, ensuring that the posterior surface of each slab remained face-up. The metal plate with the brain samples was submerged into a shallow tray of isopentane chilled on dry ice, freezing the samples in less than a minute. Samples that became submerged were briefly air-dried on the chilled plate before further handling. From this point, all tissue was maintained and transported on the metal plate in contact with dry ice. Smaller samples (e.g., blood, olfactory bulb, dura, meninges, choroid plexus) were frozen by floating 1.5 mL tubes in the chilled isopentane. Each brain slab was vacuum sealed on dry ice and then stored at -80°C.

Subcortical dissections from each slab were performed by outlining prospective tile boundaries on each frozen slab. We intentionally sampled the following subcortical structures: the non-isocortical telencephalon (HiF, OlfC, CNU), diencephalon (THM, HTH), midbrain, and majority of the hindbrain (excluding cerebellum). One to eight contiguous tiles (14-60 mg each) were dissected from each slab. Microdissected tiles were stored in microcentrifuge tubes at -80°C or used immediately for nuclei isolation. When possible, tile boundaries respected visible anatomical boundaries on the slab face. However, each tile invariably encloses multiple anatomical structures and some structures from outside the regions above are inadvertently sampled. Tiles were generally sampled from the left hemisphere of each slab; the right hemisphere was reserved for resampling due to sample or library preparation failures.

Nuclei isolation

Single nuclei isolation of marmoset samples for 10X sequencing was conducted as described in (https://dx.doi.org/10.17504/protocols.io.5qpvok1exl4o/v2). Briefly, frozen marmoset neural tissue microdissections were placed into a 3 mL dounce homogenizer containing 2 mL homogenization buffer (250 mM sucrose, 25 mM KCl, 10 mM Tris pH 8.0, 5 mM $MgCl_2$, 0.1% Triton X-100, 1X Protease Inhibitor Cocktail (Halt™), Protector RNase Inhibitor, and 0.1 mM DL-Dithiothreitol (DTT)). Tissue was homogenized with approximately 10 to 20 strokes of the pestle and passed through two 15 mL conical tubes with 70 and 30 μm cell strainers. To pellet the nuclei, the sample was centrifuged at 900 x g for 10 minutes at 4°C. Nuclei were resuspended in 3% blocking buffer (1X phosphate-buffered saline (PBS), 30% bovine serum albumin (BSA), and Protector RNase inhibitor). To remove myelin from the sample, an equivalent volume of 42% OptiPrep™ density gradient medium (iodixanol) solution was added to the sample. This resulting 21% iodixanol mixture was layered on top of an equivalent amount of 25% iodixanol solution. Samples were centrifuged at 8,000 x g for 15 minutes at 4°C followed by removal of the myelin at the surface and the remaining supernatant. Samples were resuspended in 1 mL of blocking buffer and nuclei were incubated in mouse anti-NeuN PE and rabbit anti-Olig2 Alexa Fluor 488 for 30 minutes. After staining, the samples were spun down at 400 x g for 5 minutes at 4°C. The supernatant was carefully removed so as to not disturb the pellet and resuspended with 300 μL blocking buffer and DAPI (4′,6-diamidino-2-phenylindole) at a final concentration of 0.1 μg/mL.

Fluorescence-activated nuclei sorting (FANS)

Nuclei suspensions were sorted with the Sony MA900 cell sorter using a 130 μm sorting chip. The gating strategy included first gating on their size and scatter properties. Nuclei were then gated based on DAPI signal to label DNA-containing nuclei and exclude debris and doublets. Finally, nuclei were sorted three ways based on the NeuN positive, Olig2 positive, and double negative signals. It took approximately 25 minutes to sort through each 300 μL nuclei suspension. Typically, Olig2 positive nuclei were the most abundant population, followed by slightly fewer NeuN positives, and a much smaller double negative count.

Sample preparation for multiome sequencing

In order to enrich the neuronal nuclei populations for sequencing, the sorted nuclei were pooled at an ideal ratio of 70% NeuN positive, 10% Olig2 positive, and 20% double negative. If the resulting pool contained less than 40,000 total nuclei, the proportions of each population were manipulated to allow for a greater sample size while attempting to maximize NeuN positive input. This usually meant decreasing the proportion of double negative nuclei in the pool. After pooling, BSA was added to a final concentration of 3% followed by centrifugation at 600 x g for 10 minutes at 4°C. The supernatant was decanted using a P1000 pipette leaving approximately 250 μL above the invisible pellet of nuclei. An equal amount of 2X nuclei buffer (10X Genomics PN2000153; with Protector RNase inhibitor, 0.2 mM DTT) was added to the sample. The sample was finally centrifuged at 600 x g for 5 minutes at 4°C. The

supernatant was again decanted carefully leaving just 10 μL at the bottom per desired 10X Multiome reaction. Nuclei were quantified using a hemocytometer on an EVOS M5000 fluorescent microscope. A sample concentration of at least 3,230 nuclei per μL was ideal for the 10X Chromium for a recovery of 10,000 cells/μl.

GEM generation and library preparation were performed following the manufacturer's protocol (10x Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits user guide, CG000338). Libraries were validated and normalized for pooling using the 4150 TapeStation System (Agilent, #G2992AA) and the Qubit Flex Fluorometer (Thermo Fisher, #Q33327). Libraries were shipped to the Broad Institute and sequenced using a NovaSeq X Plus 25B flow cell (Illumina) at a depth of 120,000 reads/nucleus.

Xenium spatial transcriptomics data collection

Experimental details for this dataset are also described in Hewitt et al.[33] A fresh whole brain from a 5-year-old male marmoset (Table S1) had its brainstem and cerebellum resected at an oblique angle, and the remaining hemispheres were slabbed into six, 5-mm thick, coronal slabs from rostral-to-caudal, which were then flash frozen. Slabs were then divided into hemispheres, and slab face images were used to identify the three right hemisphere slabs (slab 42, slab 43, and slab 44) that contained the rostral-caudal extent of all basal ganglia structures. Specifically, the rostral pole and caudal-most portion of the caudate (approximately interaural +13.80 mm and +0.80 mm, respectively[141]) were used as anatomical landmarks for sectioning. Selected slabs were sectioned on a Leica cryostat at 10 μm onto 10X Xenium slides and collected at 200 μm intervals. During sectioning, tissue sectioned from the front of each slab was discarded until a near-complete hemisphere was obtained; similarly, incomplete sections at the end of each slab were discarded. 22 sections were collected from slab 42, 21 sections from slab 43, and 16 sections from slab 44 for a total of 59 sections. 12 sections were discarded due to Xenium instrument failure or per-section QC failure. These factors led to a final number of 47 sections (19, 14, and 14 sections per slab, respectively), which span the extent of the basal ganglia with a finest inter-section interval of 200 μm, some inter-section intervals of 400-600 μm, and larger intervals between slabs. Spatial transcriptomic sampling ended rostral to the caudal-most multiome slabs and did not include the medulla, pons, or cerebellum.

The 300 gene panel was constructed from a combination of computationally derived and manually selected marker genes. Literature derived marker genes, including genes chosen for variation between caudate and putamen,[12,38] along with shared genes from macaque and human spatial gene panels, formed a starting gene list. This list was expanded using mFISHtools[142] and GeneBasis[143] applied to variable genes in BG and thalamus cells from a previous marmoset census dataset.[20] Final gene panel selection was made iteratively with 10X to achieve compatibility with Xenium probe and detection constraints. Sections were fixed, probe hybridized, amplified, and imaged according to standard 10X Xenium protocols. The 10X Xenium Cell Segmentation Kit was used; the on-instrument cell segmentation algorithm was run with default settings.

QUANTIFICATION AND STATISTICAL ANALYSIS

Marmoset genome reference

All sequenced libraries were aligned to "mCalJa1.2.pat.X_mitos2"[144], a version of NCBI GenBank/RefSeq GCF_011100555.1/GCA_011100555.2 modified to contain mitochondrial gene annotations generated using the mitos2 webserver[145] and mitochondrial chromosome CM021961.1.[146]

Multiome sequence alignment and library QC

10x Multiome libraries were aligned using 10X cellranger-arc (v2.02). For library QC, we used cellranger-arc output metrics and MultiQC[147] (v1.30). Out of 159 total alignments, we considered 12 RNA libraries to be QC fails for either failing to align (n=1), having <50% of reads map confidently to transcriptome (n=3), <60% of reads in cells (n=4), or >55% of reads contain the TSO sequence (n=8). We additionally failed one more library after determining a large fraction of its cells were found in low quality ("junk") clusters. We considered 15 ATAC libraries to be QC fails for having >10% of read pairs unmapped (n=1), <50% of high-quality fragments in cells (n=7), or having <20% of transposition events in peaks be in cells (n=7). In total, 133/159 libraries passed all QC, however we maintained an additional 13 RNA libraries whose ATAC counterparts were called QC failures.

Multiome RNA processing pipeline

RNA count matrices from cellranger-arc were processed using cellbender[130] (v0.3.0) to remove ambient RNA. scDblFinder[131] (v1.21.0) was used to annotate doublets using an expected doublet rate of 0.01 per 1000 cells (after removing cells with <200 UMIs). After doublet removal, we kept cells if they had 2000-50000 UMIs, 1000-13000 genes, and <3% mitochondrial gene content. Basic cell curation was done in Seurat[148] (v5.0.3) before exporting to h5ad.

Scanpy[132] (v1.10.4) was used for the calculation of top 4000 highly variable genes (HVGs) with "seurat_v3" as flavor and donor name as batch key. Using the HVG subset, we corrected for donor batch effects by training a scVI[34] (v1.2.2.post2) model with donor name as the batch key and n_hidden=256, n_latent=64, n_layers=3 as the network parameters for 300 epochs to integrate cells across all donors. In the donor-integrated scVI latent space, cells were clustered using the transcriptomic clustering package[35] (v1.0.0) with the same clustering parameters as described in the companion cross-species BG taxonomy paper.[32] Initial iterative

clustering yielded 1811 clusters and additional manual curation was performed to remove 74 low quality clusters with high mean doublet scores. For each cell, raw gene expression counts were normalized with a target sum of 1e6.

Pseudobulked "metacells" for each Cluster and Group were generated by summing raw reads, which were then CPM-normalized. Normalized metacells for all Groups were reduced to only include HVGs as input for hierarchical clustering using FactoMineR::HCPC[138] (v2.11). The resulting dendrogram was split into neurons and nonneurons, and visualized using ggtree[149] (v3.6.2) and ggnewscale[150] (v0.4.10). The Cluster-level metacells are used throughout other analyses. When gene expression is shown with z-scores in heatmaps, the z-scores are calculated across all Clusters, and when shown for some grouping of Clusters (for instance, for a Class, Group, Dominant Region, etc.), the Cluster-level z-scores are averaged (arithmetic mean).

Marker gene identification and differential gene expression analysis

Marker genes for every Cluster and Group were computed using NSForest[151] (v4.1) with default parameters. When directly comparing between 2 clusters, we used scanpy.tl.rank_genes_groups with wilcoxon rank sum tests. When comparing groupings of multiple Clusters, differentially expressed genes were calculated using DESeq2[137] (v1.46.0) using the individual Cluster metacells (pseudobulks) as pseudo-replicates. For DESeq2, pseudobulk count matrices were subset to contain only the clusters being compared. Where noted, shrunken log2FoldChange estimates were produced using apeglm[152] (v1.28.0).

Cross-species mapping against reference dataset

To generate the taxonomy, we mapped our RNA dataset to various published atlas-level scRNA-seq and snRNA-seq datasets in human, mouse and other primates. MSCA dataset was restricted to 1:1 orthologous genes in the reference species. For large reference datasets with over 1 million cells, we used MapMyCells[39] (RRID: SCR_024672) to predict labels from reference datasets to each cell in MSCA due to memory constraints. All other reference datasets were integrated with corresponding MSCA subsets using scVI with custom tuned network parameters and donor labels as covariate. KNN classifiers were then used to predict labels for each cell in MSCA based on the integrated latent space. Detailed algorithms and parameters used for each reference dataset and the associated mapping results from each reference dataset are summarized per Cluster and Group in Tables S4 and S5.

Additional pairwise mapping among whole brain human,[3] mouse scRNA-seq[1] and mouse Slide-seq[2] atlases was performed using MapMyCells to identify homologous clusters across species.

Xenium spatial transcriptomics processing and QC

Data processing and quality control for the Xenium spatial transcriptomics data were done as described in Hewitt et al.[33] Briefly, on-instrument cell segmentation was run with default settings. Segmented cells were deemed low quality and removed if they had fewer than 20 total transcripts and fewer than three unique genes (Figure S1G). Gene counts were normalized to log2 counts per thousands (log2cpt).

Cross-modality cell-type mapping between spatial transcriptomics and multiome RNA-seq

Cells from the Xenium spatial transcriptomics datasets were mapped to the snRNA-seq "v4 Clusters" using the MapMyCells[39] (RRID: SCR_024672) flat mapping algorithm with bootstrapping (100 iterations, bootstrap factor 0.95) and using all 300 genes from the gene panel (Table S2) as marker genes. The higher levels of the MSCA taxonomy (Class, Group) were assigned to each Xenium cell using label transfer via its mapped Cluster identity.

In silico validation of gene panel performance for spatial RNA-seq mapping

MapMyCells was used to self-map the multiome RNA to its own taxonomy using the same settings as was used for the Xenium spatial RNA-seq mapping. The gene sets used for the mapping task were: original 300 gene spatial panel; the 4000 HVGs used for multiome RNA-seq scVI integration and clustering; Group- and Cluster-level NSForest markers from the taxonomy (n=398 and n=2916, respectively); all marmoset transcription factors, taken from AnimalTFDB v4.0[153] (n=1329); the mouse ABC atlas MERFISH panel genes (n=500); and all genes used for labelling in the mouse ABC WMB atlas taxonomy (n=194) (Table S2). For assessing performance, the original cell labels in the multiome taxonomy were used as ground truth for calculating precision, recall, and F1 scores.

For the spatial gene panel feature importance analysis, a Random Forest classifier (sklearn v1.8.0) was trained on the multiome snRNA-seq data using the union set of the 4000 HVGs and the 300 Gene Spatial Panel (n=4104). The model was trained with 1000 decision trees, and impurity-based feature importance scores were extracted to quantify the extent to which each gene contributed to splitting decisions across the forest. Training was parallelized in 40 batches of 25 trees each.

Generation of MSCA Taxonomy

Groups in the initial draft taxonomy were generated by aggregating clusters with similar transcriptomic signatures and consistent mapping to the same reference label in all relevant reference datasets. After mapping snRNAseq defined Clusters to Xenium spatial transcriptomics data as described above, we examined each Cluster's spatial distributions and refined Cluster-to-Group assignments. We merged Clusters with similar transcriptomic profiles and contiguous spatial localization, as well as split transcriptomically similar Groups into subsets with clearly distinct spatial territories. All acronyms used in Group names are provided in the Harmonized Ontology of Mammalian Brain Anatomy (HOMBA, RRID: SCR_027628).[154] Groups were then assigned to the corresponding Class

derived from the mouse taxonomy.[1] Marker genes that are consistently expressed in MSCA and reference datasets are included as part of the Group name. Clusters identified to be in the basal ganglia were assigned a Group based on the cross-species consensus taxonomy generated in Johansen et al.[32] In the present manuscript, two BG Groups, STRv D1 NUDAP eSPN and AMY-SLEA-BNST GABA, were renamed to AMY-STR ITC-NUDAP eSPN and CEN-SLEA-BST GABA, respectively, to reflect their anatomical distributions when considering locations outside of the BG.

Two Groups (OB Dopa-GABA and OB FRMD7 GABA) received olfactory bulb (OB) anatomical labels due to their shared gene expression programs with OB periglomerular interneurons. However, cells from these Groups localized to the white-matter boundary of the striatum in the spatial transcriptomics donor, and these cells are unlikely to have originated from the OB itself because we did not sample the rostral-most slab, where the OB is located, in the multiome donors. These Group labels were inherited from the BG cross-species consensus taxonomy, and the characterization of these two peri-striatal Groups in marmoset, as well as macaque and human, is described in greater detail in the companion paper.[32]

Additionally, we used our spatial mapping to manually assign a Dominant Region to each neuronal Group in our taxonomy where possible. For Groups with representation across multiple anatomical structures, we refined these assignments at the Cluster level.

Comparative analysis across mouse, marmoset and human neurons

To assess the transcriptional similarity and expression variability of cell types across mammalian species, we first identified homologous cell types. We took the cross-species mapping results of mouse,[1] marmoset MSCA, and human[3] atlases generated above and used human clusters as the common anchor. A mouse or marmoset cluster was considered homologous to a given human cluster if >70% of its cells mapped to that human cluster. Using this criterion, 114 of 382 human clusters were identified as having homologs in both mouse and marmoset. Each dataset was then restricted to the corresponding set of 1:1 orthologous genes, and these subsets were used as input to quantify expression similarity per cluster using expressolog.[51]

Multiome ATAC processing pipeline

ATAC processing was done following RNA analysis and the generation of a draft taxonomy, and using only libraries and cells that passed RNA QC and cell curation (for a final 632920 ATAC cells, vs. 678917 in RNA). All ATAC preprocessing was done using SnapATAC2[49] (v2.8.0). After generating fragment files based on our RNA-curated cells, a draft version of our subcortical atlas "Group" level annotation was used for MACS3 peak calling, and peaks were merged using "merge_peaks" (half_width=250), generating >1.7 million peaks. We generated a full cell-by-peak matrix using "make_peak_matrix" (counting_strategy="insertion") as well as a pseudobulked cluster-by-peak matrix. We focused subsequent analysis on a reduced peak set that contains only peaks with non-zero signal in >50% of the cluster-level pseudobulks in at least one Group (n=462,698 peaks).

When peak accessibility is shown in heatmaps (z-scored), they are treated the same as the RNA: z-scores are calculated across all Clusters and then averaged (arithmetic mean) as needed. bigWig files were generated for our taxonomic Groups using "export_coverage" (normalization="CPM", counting_strategy="insertion", bin_size=1). We used "topic modeling" (with pycisTopic[50]) to group peaks into modules and to produce a low-dimensional representation of chromatin accessibility.

Multiome ATAC-seq browser shots

The browser shot in Figure 3D was generated using IGV desktop[140] (v2.16.0) and a single raw Class-level bigWig file. The browser shots in Figure 4F were generated in R using github.com/mdeber/browser_shot_plotter with CPM-normalized bigWig files and a bin-size of 50bp. Because no bigWig track was generated for "All Nonneurons besides Astrocytes", that track was generated by merging all other nonneuron Group CPM-normalized bigWig files and renormalizing to sum to one million.

Gene Regulatory Network (GRN/eRegulon) calling and scoring

Due to performance constraints, SCENIC+[54] (v1.0a2) was run separately on 50000 neuronal and 50000 non-neuronal cells sampled using geometric sketching[155] based on our RNA scVI latent space. We performed Topic modelling with pycisTopic[50] (v2.0a0) using default parameters, and we chose the optimal number of topics (200 for neurons, 175 for non-neurons) based on log-likelihood. Three sets of candidate enhancers were created by taking (1) top 3000 peaks per topic, (2) differentially accessible peaks on the imputed matrix at the Group level (Wilcoxon rank sum test logFC>0.5 and Benjamini–Hochberg-adjusted $P < 0.05$) and (3) selection by topic binarization using the Otsu method. A custom database using the reduced peak set was created and motif enrichment analysis was implemented with pycistarget[156] (v1.1). Finally, we ran the SCENIC+ workflow with default settings. SCENIC+ eRegulons are called either "direct" in the case of an exact within-species TFBS motif match, or "extended" if based on motif similarity or orthology. In neurons we produced 321 direct (307 activator + 14 repressor) and 229 extended (214+15) eRegulons; and in non-neurons 173 direct (158+15) and 126 extended (115+11) regulons. All eRegulons shown and discussed are "direct", with activators being classified by SCENIC+ as "+/+" and repressors "-/+".

eRegulon activity was scored at the cluster-level using RNA or ATAC pseudobulks with the R package AUCell[54] (v1.28.0) using default parameters. When eRegulon z-scores are used, these are generated at the Cluster-level using all clusters, and unless otherwise specified, these are based on gene AUC. When eRegulon z-scores are generalized across a set of clusters, the arithmetic mean of the Cluster z-scores is used. For Group-level threshold-based analyses of OTX2 expression (e.g., defining OTX2+ Groups in Figure 3F), a 30 CPM cutoff was applied. eRegulon network plots were generated in R using ggraph[157] (v2.2.2); in network

visualizations, TF-gene edges are displayed only if the SCENIC+ importance score exceeds 1.5 (Figure 3H); for TF-TF networks centered on a focal TF, connectivity is restricted to two hops upstream or downstream (Figure 5E).

When multiome data (e.g. eRegulon activity, which is gene AUC z-score) is "projected onto spatial data" (as in Figure 3E), each spatial cell's mapped Cluster is taken as the cell identity, and the eRegulon gene AUC z-score for that Cluster is plotted. This means all cells with the same mapped Cluster will have the same values projected.

Cytosplore Viewer

Cytosplore Viewer is a plugin-based visual analytics system designed to support comparative analysis of multi-modal brain atlases across species. It is built on Manivault,[158] an application building toolkit for visual analytics applications for multi-modal high-dimensional datasets. By integrating coordinated visualizations of embeddings, gene expression, spatial data, and regulatory signals, the system links local compute to the underlying single-cell measurements and enables iterative exploration of complex biological datasets. Cytosplore Viewer leverages local, on-the-fly computation in combination with the fact that ATAC-seq and RNA-seq are annotated with the same, fine-grained clustering. This enables average feature profiles for ATAC peaks and gene expression to be projected on the spatial data. By approximating features with cluster averages, real-time interaction, searches and visualization for data exploration can be achieved on standard compute hardware. This allows for spatial visualization of any of the ATAC peaks, as well as of transcriptome-wide gene expression, in two linked scatterplots.

The spatial mapping module offers a search engine for genome coordinates in the ATAC data, as well as gene symbols in the RNA-seq data, enabling the spatial visualization of these single-cell features. These views are complemented with a bar chart showing the top cell types (and proportions) in which these features are predominantly expressed. In addition, the corresponding peaks and genes are projected onto the RNA UMAP. The module also supports PCA on user-defined feature sets, where PC1 is calculated and mapped onto spatial coordinates, with optional filtering to restrict visualization to selected cell types. By integrating these operations into a single environment with low-latency user interaction, the ATAC Viewer enables interactive exploration of spatial regulatory and gene expression features, enabling user-drive hypothesis generation in a seamless analytical workflow.

For a detailed description of the design considerations and process of Cytosplore Viewer, see Chapter 5 of Basu, 2026 (https://www.dropbox.com/scl/fi/10e9lw7kjdkvye9uczbzz/Soumya_PhD_Thesis_Final.pdf?rlkey=az9v4khlvo3h5sn63aanwniem&dl=0), and viewer.cytosplore.org.

RNA/ATAC co-enrichment analysis by region and neurotransmitter

To look for the most conspicuous overlaps of specific genes (RNA) and peaks (ATAC) across neurons by region and neurotransmitter (as in Figure S3E), we regrouped our neuronal clusters by dominant region and neurotransmitter. Then, using the raw Cluster-level pseudobulked metacells (using all genes for RNA, or the reduced peak set for ATAC), we ran DESeq2, treating each cluster as a pseudo-replicate observation of each region+neurotransmitter pairing, and as a background we used all neuron clusters from other brain regions. We then used apeglm to shrink the log2FoldChange estimates, which reduces the magnitude of log2FoldChanges for features with higher variance, thus producing single adjusted parameter estimates for enrichment. To look for overlaps among highly-specific features, for RNA we used the top 2% of genes by shrunken log2FoldChange, and for ATAC, the top 1% of peaks. Then, for every pairwise comparison between a region+neurotransmitter (which could be between neurons that share a region, or share a neurotransmitter, or share neither), we scored co-enrichment as the ratio of observed overlaps vs. that expected by random chance. High co-enrichment indicates that the neurons being compared share a large fraction of their most-specific features. The same DESeq2/apeglm pipeline was used to compute per-gene enrichment scores for individual groupings vs. neurons from other regions; CRACDL enrichment in Figure 5L shows the resulting apeglm-shrunken log2FoldChange.

Principal component analysis of astrocyte Clusters

Metacells matrix with top 4000 HVGs (see 'multiome RNA processing pipeline' above) for each astrocyte Cluster was used as input for the principal component analysis (sklearn.decomposition.PCA (v1.5.2), n_components=10). PC1 and PC2 captured 27.3% and 18.2% variation respectively.

TEL Glut vs. LGE GABA functional convergence analysis

"TEL Glut" and "LGE" are defined in Figure 5G, which is based on regrouping neuronal Clusters by Dominant Region, Neurotransmitter, and GE origins (see main text). The "background" set includes all other neuronal groupings. All enrichment is based on z-score differences vs. that background set, with difference >1 in both TEL Glut & LGE being used to call co-enriched direct eRegulons. Co-enriched genes within those co-enriched eRegulons were also selected using the same criteria (Cluster-level z-scores are averaged for each grouping, and differences >1 are again considered enriched). That set of 279 genes was exported, and the 242 coding genes were used for enrichment analysis in string-db (v12.0, permalink: https://version-12-0.string-db.org/cgi/network?networkId=bzt5qpNsbveL). Selected network plots were downloaded from string-db. Dot plot was recreated for selected terms in ggplot2 (v4.0.2). Enrichment values reported in Figure 5J are the harmonic mean of observed/expected ratio and -log(FDR). UpSet plots were generated with R package ComplexHeatmap (v2.25.2). Given the number and redundancy of significant terms, terms selected for use in final plots were chosen manually following both a clustering analysis and size/clarity considerations.

Abbreviations used in STRING gene set enrichment analysis are ''GO'' (Gene Ontology), ''CC'' (Cellular Component), ''MF'' (Molecular Function), ''BP'' (Biological Process), ''NMDA'' (N-methyl-D-aspartate), ''AMPA'' (α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid), ''CaM'' (calmodulin), and ''PPI'' (protein-protein interactions). Abridged term names are ''Reactome: NMDA unblocking'' (''Unblocking of NMDA receptors, glutamate binding and activation''), ''Reactome: AMPA trafficking'' (''Trafficking of AMPA receptors''), ''Reactome: Ras after NMDA/Ca2+'' (''Ras activation upon Ca2+ influx through NMDA receptor''), ''Reactome: Postsynaptic signaling'' (''Neurotransmitter receptors and postsynaptic signal transmission''), ''Reactome: Neurexin/neuroligin'' (''Neurexins and neuroligins''), ''GO MF: CaM kinase activity'' (''Calmodulin-dependent protein kinase activity''), ''GO BP: Calcium import'' (''Calcium ion import across plasma membrane''), and ''STRING: Postsynapse & PPI'' (''Postsynaptic cell membrane, and Protein-protein interactions at synapses'').

Alignment of Xenium Spatial Transcriptomics data in three-dimensions

To align consecutive Xenium slices, we used PASTE2,[108] which leverages the spatial location and gene expression of all cells simultaneously. Due to the large number of individual cells, cells are first aggregated to alignment spots with a radius of 120um. For each alignment spot, raw gene expression is computed as the sum of expression profiles from its constituent cells. The spatial location of each spot is set to the UMI-weighted centroid of these cells, and its weight is defined as the number of constituent cells. Following standard procedures, gene expression of alignment spots are log-normalized and reduced to 30 principal components. Adjacent pairs of slices are then aligned with PASTE2 with minimum overlap fraction of 0.9, Wasserstein (spatial) loss fraction of 0.1, and marginal distribution weighted by cell counts. After computing the transport map between each adjacent pair, an affine transformation is trained to fit the transport map, minimizing squared distance between the barycentric projection and affine projection of all spots in the source slice. Finally, to register all slices into a common space, adjacent affine transformations are composed such that each slice can be projected onto the first one by a single affine transformation.

Hippocampal Gradient Analysis

Cells from hippocampal Groups CA3 Glut, CA1-ProS Glut and SUB-ProS Glut Group in the spatial transcriptomics and multiomics datasets were included for this analysis. In the Xenium spatial transcriptomics dataset, because the hippocampus was only present from sections z=17 onwards, any clusters mapped to anterior sections were excluded as low-confidence. All subsequent analyses are conducted for each Group separately.

For examining longitudinally graded genes out of the 300 genes included in the spatial transcriptomics data, we excluded non-neuronal marker genes from this screen (*AQP4*, *GFAP*, *MEGF10*, *BCAS1*, *VCAN*, *CLDN5*, *PECAM1*, *MFSD2A*, *RGS5*, *LRP2*, *CD3D*). To reduce sensitivity to uneven sampling density across sections, cells were binned (n_bins=30) along the z-axis using quantile binning and we restricted our screening to only genes with log2cpt that vary more than 0.5 standard deviations across the spatial bins. Bins containing fewer than 10 cells were dropped from downstream analysis. For each bin, we computed its mean z coordinate and the median expression value for each gene. To cluster continuous expression trends across spatial bins, we fit a generalized additive model (pyGAM, v0.12.0) for each gene as a smooth function of z position using 6 splines, with smoothing parameter λ = 0.6 and no intercept. Fitted expressions of each gene were used for hierarchical clustering (seaborn.clustermap v0.13.2, metric='euclidean', method='ward'). Across the three subfield Groups tested (CA3 Glut, CA1-ProS Glut, SUB-ProS Glut), two dominant trends were evident, with majority genes showing monotonic increases or decreases along the A-P axis.

We then used the multiomics dataset to look for transcriptome-wide longitudinally graded genes. For each cluster, spatial position along the hippocampal longitudinal axis was assigned as the section containing the greatest number of mapped Xenium cells. Clusters with fewer than 10 cells in their maximally represented section were excluded to avoid assigning spatial positions from sparsely sampled clusters. One-dimensional UMAP embeddings (1D UMAP) were computed from snRNAseq data separately for each Group using scanpy.tl.umap with n_components = 1. The median 1D UMAP coordinates of each cluster were then correlated with the modal spatial *z*-order to which that cluster mapped in the Xenium spatial transcriptomics data. To identify genes with A-P gradients in the hippocampus, we computed correlations between gene expression and the 1D UMAP coordinates in the multiome dataset. P-values for all correlations were adjusted for multiple testing using Benjamini-Hochberg; features with |Spearman r| >0.5 and Benjamini-Hochberg-adjusted $p < 0.05$ were considered significantly graded. For cross-subfield comparisons of graded genes (Figure 6F), significantly graded genes from each hippocampal glutamatergic Group were partitioned by the sign of their correlation with modal spatial z-position, yielding anterior-posterior and posterior-anterior sets. Overlaps were visualized as UpSet plots.

Similarly, we identified peaks with A-P gradients by correlating the binned chromatin accessibility (n_bins=30) against the 1D UMAP coordinates due to the sparsity of the ATAC data. Significance of peaks are assessed similarly as genes above (|Spearman r| >0.5, p-adj < 0.05). To test the frequencies of each hypothesis in Figure 6H, we subset the eRegulon triplets identified above based on graded TFs and/or peaks and then calculated the percentage of target genes that were significantly graded within each subset.

Cross-species mapping of primate non-RT Thalamic GABAergic population

In addition to methods described in 'cross-species mapping against reference dataset' section above to find homologous rodent cell types in the mouse brain atlas using MapMyCells, we also ran SAMap[118] following the same methods and parameters described in Corrigan et al.[4]

Regional ATAC marker analysis

Related to Figures 7E, 7F, and S7E, we defined regional markers in chromatin accessibility to be ATAC peaks that are highly specific to a single region but also shared among lineage-separated neurons. For this, we restricted analysis to glutamatergic and GABAergic neurons, and to regions that contained >2 clusters of each neurotransmitter (AMY, BF, CTX, EXA, HiF, HTH, MB, THM). Using our Cluster-level CPM-normalized ATAC metacell matrix, we defined region+neurotransmitter specificity as the fraction of a peak's normalized reads that are found within that region+neurotransmitter. This naturally had the tendency to favor peaks with at least moderate accessibility. Then, to find peaks with shared specificity across glutamatergic and GABAergic clusters, we first ranked peaks by region+neurotransmitter specificity (with 1 being the least specific), and summed these ranks for the two neurotransmitters. We took the 40 highest combined ranks as a set of regional ATAC markers.