

Spatial patterning of transcriptional and regulatory programs in the primate subcortex

Shu Dan^{1*}, Meghan A. Turner^{2*}, Michael DeBerardine^{1*}, Lakme Caceres¹, Victor E. Nieto-Caballero¹, Jessica Schembri¹, Kirsten Levandowski^{3,4}, Christopher Cardenas¹, Ning Wang¹, Chang Li⁵, Soumyadeep Basu^{5,6}, Hongyu Zheng⁷, Xinhao Liu⁷, Thomas Höllt⁶, Jessica Feliciano¹, Qiangge Zhang^{3,4}, Reilly Nakamoto¹, Delissa A. McMillen², Naomi Martin², Nasmil Valera Cuevas², Paul Olsen², Josh Nagra², Jazmin Campos², Marshall M VanNess², Jack Waters², Shea Ransford², Zoe Juneau², Sam Hastings², Stuard Barta², Augustin Ruiz², Jeanelle Ariza², Benjamin J. Raphael⁷, Hongkui Zeng², Ed S. Lein², Boudewijn Lelieveldt^{5,6}, Guoping Feng^{3,4}, Brian Long^{2†}, Fenna M. Krienen^{1†}

^{*}equal contribution

[†]equal contribution; corresponding author

¹Princeton Neuroscience Institute, Princeton University, Princeton NJ

²Allen Institute for Brain Science, Seattle, WA

³Yang Tan Collective and Govern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; Cambridge, MA

⁴Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA

⁵Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands

⁶Faculty of EEMCS, Delft University of Technology, Delft, the Netherlands

⁷Department of Computer Science, Princeton University, Princeton NJ

Abstract

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 found across all subcortical brain regions in a primate - the marmoset monkey. Dense sampling and cross-species integrations revealed spatially precise neuronal assemblies, including in complex structures such as hypothalamus. We find chromatin accessibility and transcriptional identity are spatially tuned within and across subcortical structures. Spatial gradients within subfields of the hippocampal formation are predominantly orchestrated by graded transcription factors coupled with graded enhancers. The primate-expanded population of GABAergic neurons in the thalamus shares transcriptional and regulatory syntax with neurons in superior colliculus, reflecting an evolutionary adaptation compared with rodents. We show that unexpected transcriptional convergence, such as between striatal GABAergic medium spiny neurons and telencephalic glutamatergic neurons, can arise when distinct gene regulatory networks impinge on the same downstream genes.

Introduction

Mammalian brain cell identity is patterned by intrinsic factors as well as by experience, including from cell-cell interactions and the tissue microenvironment. Recent estimates indicate there may be thousands of transcriptomically-distinct cell populations in the adult mouse brain^{1,2}, which has the most comprehensively sampled brain cell taxonomy of any mammal. Focused comparisons of anatomical structures indicate that cell types can diverge between species in their proportions, molecular details, location, or, in rare cases, by complete gains or losses³⁻⁶. A key question is how these cell types relate to the diversity present in other mammals, specifically humans or other primates.

The sheer size of the human brain, which contains over 86 billion neurons and is ~3.6 orders of magnitude larger (by mass) than the mouse brain, presents a formidable challenge for determining the extent of conservation and divergence of cell populations and their topography. Currently, only partial sampling of large brained species is feasible, which limits the detection of rare cell types and obscures complex spatial topographies such as regional gradients and other discontinuities. Incomplete sampling also complicates cross-species comparisons, particularly since cell migration can be a potent driver of evolutionary change^{4,7-9}.

The small size of the marmoset (*Callithrix jacchus*) – a New World monkey whose brain weighs 8g (similar to rats) – could serve as a “bridging model” to understand divergence between rodent and human brains^{10,11}. Despite its small size and evolutionary divergence (35-40 million years to great apes), the adult marmoset brain retains primate-specific anatomical¹²⁻¹⁵, connectivity¹⁶⁻¹⁸ and cell type specializations^{3,19-21}.

As part of the BRAIN Initiative Cell Atlas Network (BICAN), we generated a hierarchical cell taxonomy based on anatomically comprehensive multiomic (snRNA-seq and snATAC-seq) sampling of the marmoset subcortex. In parallel, we performed spatial transcriptomic profiling of an intact marmoset cerebral hemisphere using the 10x Xenium platform, enabling precise anatomical localization of cell populations at each level of the taxonomy. We integrated our atlas with existing human, non-human primate, and mouse brain atlases to facilitate interoperable annotations across datasets and to provide a foundation for cross-species comparisons. The dense and comprehensive sampling of the marmoset atlas enabled mapping of neuronal taxonomies with high spatial precision, including in complex structures such as the hypothalamus and basal forebrain. We also observed considerable spatial zonation within non-neuronal populations, including spatially-restricted subtypes of ependyma, astrocytes, and microglia.

Regulatory elements such as enhancers give rise to spatially and temporally controlled gene expression patterns during brain development. These patterns establish regional boundaries and cell fates. Spatial co-mapping of ATAC-seq and RNA-seq indicate that broad cell classes retain distinct epigenetic and transcriptomic profiles in adulthood²²⁻²⁴. Leveraging our spatially-resolved taxonomy, we show that chromatin accessibility is precisely tuned by cell type identity as well as by a cell’s anatomical location. In general, the degree of chromatin accessibility similarity between a pair of neuronal populations mirrors their degree of transcriptional similarity; to a large extent, both modalities reflect the cells’ developmental histories. However, we also find that spatially co-

localized populations share an excess of accessible chromatin features, suggesting that adult neurons maintain an anatomically “fingerprinted” epigenetic identity beyond what is conferred by their lineage. We show that cases of unexpected transcriptional convergence – such as between distantly related GABAergic and glutamatergic neurons – can arise when cell types recruit distinct regulatory programs that control the same downstream genes.

The hippocampal formation – an elaborate structure that subserves memory and spatial navigation – has strong functional and connectivity gradients along its longitudinal axis. We find that the primate hippocampal formation is patterned along the anterior-posterior axis by both positionally-discrete neuronal populations as well as broader gene expression gradients that transcend individual cell types. These gradients are driven by spatially graded gene regulatory networks private to each hippocampal subfield. In principle, gene expression gradients can be driven by gradients in transcription factors (TFs), gradients in enhancer activities, or both. We find that in the primate hippocampus, gradients are generally maintained by graded enhancers and graded TFs acting together.

Differential migration of conserved cell types is a powerful way to modify the structure, function, and circuitry of otherwise conserved anatomical structures. Through cross-species mapping, we determined that while primate and mouse thalamic glutamatergic neuron populations are broadly conserved, the massive expansion of GABAergic neurons in the primate thalamus likely derives from a population that originates outside of the diencephalon. Our analysis indicates conservation between primate thalamic GABAergic neurons and cells in the rodent superior colliculus of the brainstem. Consistent with a non-diencephalic origin, primate-expanded thalamic GABAergic neuron populations share chromatin accessibility features in common with conserved brainstem populations. We provide an interactive Cytosplore Viewer app (<https://viewer.cytosplore.org/>) that enables users to flexibly explore spatial correlations between data modalities.

Results

Generation of the marmoset subcortical cell atlas (MSCA)

Multiome dataset.

Our marmoset subcortical cell atlas (MSCA) encompasses the non-isocortical telencephalon-including the hippocampal formation (HiF), olfactory areas (OlfC), and cerebral nuclei (CNU)-as well as the diencephalon (thalamus and hypothalamus), midbrain, and major hindbrain structures (excluding cerebellar cortex).

The subcortex of four marmoset donors was sampled using a tiling approach whereby small, contiguous pieces of tissue were dissected from each slab to cover the entire subcortex (**Fig 1A**). Between one to eight contiguous tiles weighing 14-60 mg each were dissected from each slab. When possible, tile boundaries respected visible anatomical boundaries on the slab face. However, each tile invariably encloses multiple anatomical structures. Generally, tiles were sampled from one hemisphere of each slab (usually the left hemisphere); the right hemisphere was reserved for resampling due to sample or library preparation failures. Dissociated nuclei from each tile were fluorescence-activated nuclei sorting (FANS)-sorted to enrich for neurons and deplete oligodendrocyte lineage populations following protocol in Johansen, Fu, Schmitz et al²⁵. Sorted fractions were repooled to target proportions (70% NeuN+, 10% Olig2+, 20% NeuN-/Olig2-) and used for multiome sequencing with the 10x Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression reagent kit (PN1000283). 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.

Construction of the cell type taxonomy.

We applied stringent quality control criteria to remove low-quality cells, yielding 678,917 nuclei in the final dataset (Methods, **Supp Fig S1**). Batch effects were corrected by scVI²⁶ integration across donors and followed by iterative clustering in the donor-aligned latent space using transcriptomic clustering²⁷ (Methods). In total, we resolved 1737 transcriptomic Clusters as the basis for examining cell type diversity in the marmoset subcortex. To facilitate analysis at multiple levels of resolution, we organized the Clusters into a nested hierarchical taxonomy of three levels by aggregating transcriptomically similar Clusters with overlapping marker gene expression and spatial localization determined through mapping to our spatial transcriptomics data into 212 “Groups” and these subsequently into 17 “Classes” (Methods, **Fig 1A-B, Table S1**). Clusters belonging to the basal ganglia nuclei (BG) inherited their Class and Group labels from the Basal Ganglia Cross-Species Consensus Taxonomy²⁵. In the present manuscript, two BG Groups, STRv D1 NUDAP MSN and AMY-SLEA-BNST GABA, were renamed to AMY-STR ITC-NUDAP MSN and CEN-SLEA-BST GABA, respectively, to reflect their anatomical distributions when considering locations outside of the BG.

To understand global similarities between the Groups, we hierarchically clustered Group-level pseudobulked ‘metacells’ restricted to the top 4000 highly variable genes (HVGs) (Methods, **Fig**

1B). Within neurons, clades of excitatory telencephalic (CTX, AMY, and HiF) Groups align with major anatomical boundaries. Notably, the striatal medium spiny neurons (MSNs), which are inhibitory GABAergic cell types, clustered within the same major branch as excitatory telencephalic Groups rather than other telencephalic GABAergic cell types, consistent with prior observations from mouse and primates^{20,28,29}. In the diencephalon, excitatory THM Groups formed a distinct clade whereas inhibitory THM Groups are more similar to diverse inhibitory Groups. On the contrary, Groups in the HTH are not segregated by neurotransmitter identity. Instead, excitatory and inhibitory HTH Groups intermingled in shared clades. Other major branches are organized either by neurotransmitter identity (e.g. dopaminergic and cholinergic Groups) or by anatomical divisions (e.g. Lateral Septum (LSX) and superior colliculus (SC) Groups).

Our spatial transcriptomic data was collected using the Xenium platform (10x Genomics) with a 300 gene panel, from one hemisphere sampled at 200 μ m spacing of coronal sections across three 5 mm slabs (**Fig 1C**). The gene panel was drawn from analysis of variable genes from a prior marmoset single-nucleus RNA sequencing (snRNA-seq) dataset²⁰, which contained partial sampling of ten cortical areas and eight subcortical locations. Spatial cells were mapped to the MSCA taxonomy using correlation-based mapping (MapMyCells RRID:SCR_024672) at the Cluster level, with Group and Class labels inherited from the Cluster. Aggregation of Clusters into Groups was an iterative process that incorporated the spatial localization to refine Group membership. Groups are generally named for the anatomical structures where the preponderance of cells are found.

To determine whether the spatial gene panel contained sufficient information to discriminate between Clusters in the MSCA snRNA-seq data, we trained K-nearest neighbor (KNN) classifiers on the complete snRNA-seq atlas data using 4000 HVGs or the 300 genes in the spatial panel subsets (Methods). We used 5-fold cross validation to evaluate prediction accuracy relative to the original MSCA taxonomy at the Group and Cluster level. The 300 gene set performed similarly to the larger HVG gene set (**Supp Fig S2A**), suggesting that despite representing only a small fraction of expressed/HVG genes, the spatial gene panel in principle contains comparable information to predict a cell's Cluster membership as the much larger gene sets. When classification errors arose, they most frequently occurred between transcriptionally similar Clusters, as indicated by closer distance to the predicted Clusters compared to distances to a uniform (background) distribution over all Clusters (**Supp Fig S2A**). 772/1811 Clusters (~43%) and 1031/1811 (~57%) had classification accuracy greater than 0.5 when using the 300 gene panel genes and the HVGs in the snRNA-seq data, respectively (random chance under a uniform distribution is 0.00055). Moreover, the Cluster level prediction accuracy between the two gene sets was strongly correlated ($r = 0.89$, **Supp Table S2**).

Clusters manually annotated as low quality and excluded for analysis (due to non-specific spatial mapping coupled with high doublet scores) had a median accuracy score (0.07) well below the first quartile of scores. Clusters that had bottom quartile accuracy predictions for both HVG and 300 gene sets also had median Cluster sizes in the snRNA-seq data in the bottom quartile of cluster sizes. Prediction scores were similar at the Group level when using the expression from the Xenium spatial transcriptomics data directly, though were lower at the Cluster level compared

to predictions on the RNA data directly (**Supp Fig S2A**). At the Group level, classification accuracy is high in both neurons (mean accuracy = 84%) and non-neurons (mean accuracy = 82%) (**Supp Fig S2**). Transcription factors (TFs) – particularly TF classes such as homeobox genes – are thought to fundamentally control neuronal identity, such that a combinatorial code of TFs can be used to specify all neuron classes in *C. elegans*³⁰. While being statistically enriched in the gene panel, as indicated by a hypergeometric test ($p = 0.000546$, $\alpha = 0.001$), only 23/300 genes were TFs.

We next explored consistency between cell type abundance estimates measured from the spatial transcriptomics and the snRNA-seq data. Across the entire atlas, we compared population abundances (numbers of cells) from our most complete RNA-seq donor, CJ23.56.003, to abundances from the spatial donor (CJ23.56.004). Population abundances were highly correlated across modalities at the Class level (Pearson $r = 0.98$) (**Supp Fig S2B**). Group-level populations were also correlated for nonneuronal (0.88) and neuronal (0.82) Groups (**Supp Fig S2C,D**). Classification performance based on genes in the 300 gene panel, coupled with the correlation of cellular abundances, supports the anatomical localization of most Groups in our atlas, and also suggests that analyses based on Cluster-level localization should be approached with caution. Cluster-level prediction accuracy scores (**Table S2**) can be used as an indication of confidence in the spatial mapping of individual Clusters.

Comprehensive analysis of BG structures and cell types can be found in a companion study describing a spatial atlas of marmoset, macaque and human BG³¹ as well as in a consensus cross-species multiomic BG taxonomy²⁵. The contiguous sampling of the MSCA provides additional insight into relative similarities of cell populations within and between BG and non-BG structures (**Fig 1B**). For example, subtypes of a recently characterized striatal TAC3+ interneuron population^{3,4,7,20} are found distributed across the striatum, septum, ventral pallidum, and hypothalamus (**Supp Fig S3**).

Concordance between gene expression and regulatory syntax.

Gene expression is controlled by a vast diversity of cis-regulatory elements (CREs, including enhancers). Millions of putative enhancers decorate mammalian genomes^{32,33} 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. To identify CREs, we called MACS2/3 narrow peaks on ATAC cells divided by their RNA Group identities, which we then merged into a consensus set of 500bp peaks using SnapATAC³⁴. As this approach produced ~1.7 million unique peaks, we focused all subsequent analysis on a reduced peak set that includes only those that are robustly accessible in at least one Group (see Methods). We used “topic modeling” (with pycisTopic³⁵) to group peaks into modules and to produce a low-dimensional representation of chromatin accessibility. The resulting ATAC-based embedding closely mirrored the RNA-based embedding, suggesting that the regulatory syntax captured by the topics recapitulates the transcriptional hierarchy (**Supp Fig S4A,B**). Similarly, marker genes that were differentially expressed between Groups across the taxonomy showed concomitant open chromatin along their gene bodies in the ATAC data (**Supp Fig S4C**), indicating that Group-level differences were preserved in the downsampled ATAC data.

We performed hierarchical clustering on correlations of pycisTopic ATAC loadings across Groups (**Fig 1D**, bottom triangle) and compared them to Group-level RNA correlations (based on 4000 HVGs, **Fig 1D**, top triangle), which revealed a significant association between the two modalities (Mantel test with 5000 permutations, $r=0.88$, $p<0.001$). This indicates that gene regulatory and gene expression landscapes converge on a globally similar picture of cell identity, such that the distance between any two Groups in one modality is generally concordant with distance in the other (**Supp Fig S4D**).

Interactive multimodal explorer with Cytosplore Viewer

To enable integrated, linked exploration of the MSCA, we developed a Cytosplore ATAC Viewer as part of the Cytosplore Viewer data visualizers. Cytosplore Viewer leverages the fact that all data modalities are mapped to the same standardized cell type reference taxonomy, enabling transfer of gene expression or chromatin features through their common annotations, with the spatial data as a spatial reference frame. In addition, we developed a bi-directional search engine between spatial RNA-seq and spatial ATAC peak accessibility patterns (Methods). Starting from a seed gene's expression pattern across the spatial atlas, a search can be performed for ATAC peaks that exhibit a similar spatial pattern. Conversely, a spatial ATAC profile can be used as a seed to search for genes exhibiting a similar spatial expression pattern. These searches can be confined to manually selected spatial regions of interest. Altogether, this enables an interactive, integrated exploration highlighting the specific properties of each modality in a tri-modal manner (**Supp Fig S5**). Cytosplore Viewer can be downloaded from <https://viewer.cytosplore.org>.

Assigning spatial provenance to human subcortical cell types

Mapping to whole brain human and mouse atlases

To facilitate species comparisons and interoperable analyses, we mapped the MSCA to major existing mouse and human brain cell atlases (Methods, **Supp Table S2-3**). The mouse transcriptomic atlas¹ identified over 5,000 transcriptomically distinct clusters across the entire adult mouse brain along with a spatial transcriptomic atlas based on MERFISH. This atlas offers an unprecedented view into the spatial organization of cell types throughout the mouse brain. The most comprehensive human brain dataset to date is the Siletti et al. atlas³⁶, which used snRNA-seq to sample over 100 anatomical locations of the adult human postmortem brain. The resulting taxonomy of human cells contained 31 Superclusters, 461 Clusters, and 3313 Subclusters. As spatial transcriptomics were not available in the original paper, Siletti et al³⁶ focused on mapping the taxonomy to coarse anatomical subdivisions (based on regions annotated by microdissections).

We sought to leverage our spatially-resolved marmoset taxonomy to provide greater spatial and cluster resolution to major human Superclusters found in the human brain atlas³⁶. To identify homologous cell types across species, we used MapMyCells to predict Supercluster and Cluster labels from the human brain atlas on marmoset cells (Methods). Although Supercluster labels were assigned on a per-cell basis in the MSCA, marmoset cells within the same Cluster showed

highly coherent mapping. On average, 98% of cells in a given marmoset Cluster were assigned to the same human Supercluster. We further visualized these predictions by coloring a UMAP embedding of marmoset cells with their predicted human Supercluster labels (**Fig 2A**), revealing robust mapping of transcriptomically similar cells. Of the 20 neuronal Superclusters (excluding cerebellar inhibitory Supercluster due to lack of cerebellar sampling in the MSCA), we identified homologous marmoset Groups for each Supercluster by summarizing the co-clustering proportions and quantifying the Spearman correlation of each corresponding cell type (**Supp Fig S6**). The top 3 Superclusters with the most numerous marmoset Groups are Splatter, Thalamic excitatory and Amygdala excitatory (**Fig 2B**).

To increase confidence in our marmoset taxonomy, we cross-referenced our annotations against additional publicly available regional atlases for the HTH^{37,38}, lateral septum (LSX)³⁹, hippocampal formation (HiF)⁴⁰, lateral geniculate nucleus (LG)⁴¹, and habenula (HN)^{42,43} through integration using scVI (Methods, **Supp Table S4**).

Amygdala

Using the spatial data, we resolved the human Amygdala Excitatory Supercluster into nine marmoset glutamatergic Groups (**Fig 2C**). Two additional GABAergic Groups in the amygdala came from the Splatter Supercluster, and a final Group, AMY-STR ITC-NUDAP MSN, is linked to the BG Consensus Taxonomy. These Groups localize to each of the major nuclei of the amygdala (**Fig 2D**, top), including the medial (Me), anterior (AAA), paralaminar (PL), and basolateral (BL) nuclei, and highlight discrete cell types aligning to classical histochemical divisions of the amygdala. Our Groups also identify unique transcriptomic identities associated with transition structures such as the amygdalocortical transition area (ACTA), cortical nucleus of amygdala (Co), and amygdalohippocampal area (AHA). The intercalated nucleus of the amygdala (InA) weaves between the other amygdala nuclei. The intercalated cells (ITCs) that populate this region are hypothesized to have a shared developmental origin with the eccentric medium spiny neurons found in the striatum⁷. In our marmoset data, the AMY-STR ITC-NUDAP MSN Group is localized to the InA, as well as the NUDAP compartments, supporting a shared transcriptomic identity between ITCs and STRv D1 NUDAP MSNs.

At the Cluster level, we also find subdivisions of select amygdala nuclei. Two sets of Clusters delineate the basomedial nucleus (BM) into its magnocellular (BMmc) and parvocellular (BMpc) parts (**Fig 2D**, bottom left). In the central nucleus (CEN), Clusters localize to CeM and CeL subdivisions, with a specific set of Clusters marking the boundary between the two (**Fig 2D**, bottom center). Additionally, we find extensive spatial localization of Clusters within the lateral nucleus (La), which subdivides this structure along its dorsal-ventral, medial-lateral, and anterior-posterior axes (**Fig 2D**, bottom right)

Splatter

The human brain atlas from human brain atlas³⁶ contains an enigmatic population known as the “Splatter” Supercluster, originally named for its complex topography in UMAP embeddings (**Fig 2A**). Splatter neurons originate from diverse, largely subcortical origins, including the basal forebrain, hypothalamus, thalamus, midbrain, pons, and medulla. The human Splatter

Supercluster mapped to 117 marmoset Groups (**Fig 2B, Supp Fig S6**) and spans nearly a dozen anatomical regions (**Fig 2E, Supp Fig S7A**).

One of the largest populations of Splatter neurons maps to the hypothalamus (HTH) (**Fig 2E,F**), which contains a high diversity of small and functionally distinct neuronal populations^{37,38,44,45}. In addition to the human hypothalamus cells from the human whole-brain atlas³⁶, we also integrated our dataset with recent mouse³⁷ and human³⁸ hypothalamus atlases (**Supp Fig S7B**). Many marmoset Groups mapped to specific hypothalamic nuclei, such as the arcuate nucleus (ArH), magnocellular nucleus (MCN), and mammillary bodies (MM) (**Fig 2F**). Some Groups resolved spatial subdivisions within hypothalamic structures, such as the distinct medial and lateral Groups within the ventromedial hypothalamus of the tuberal nucleus (HTHtu) (**Fig 2F**).

In the midbrain, Groups mapping to the superior colliculus (SC) clearly highlighted this structure's laminar organization (**Fig 2G**). Additionally, the septum is patterned with 7 spatially distinct GABAergic and 2 spatially distinct glutamatergic Groups (**Supp Fig S8B**). Other anatomical regions show clear subdivisions at the Cluster level, including the MM (**Supp Fig S8A**), the lateral and medial habenula (HN) (**Supp Fig S8C**), and bed nucleus of the stria terminalis (BST) (**Supp Fig S8F**).

Siletti et al³⁶ concluded that the human Splatter supercluster is composed of highly diverse but small populations that, due to insufficient sampling, appear more similar to each other relative to other populations and hence did not form more segregated superclusters on the embedding. Cell populations derived from the same anatomical structures in the mouse brain, particularly the midbrain and hindbrain, also show a highly complex topography in UMAP embeddings¹. We sought to further clarify features of Splatter populations through cross-species comparisons by examining homologous “Splatter” cell types across human, marmoset and mouse. We defined homologous cell types as mouse or marmoset clusters for which >70% of cells mapped to a given human cluster, yielding 114 of 382 human neuronal clusters with conserved counterparts in both species. For these homologous clusters, pairwise expression similarity across species was quantified using *expressolog*⁴⁶, which computes an AUROC score that measures how similarly 1:1 orthologues vary in expression across matched cell types. Splatter clusters showed the highest mean *expressolog* scores for all species pairs compared to other Superclusters, suggesting they are transcriptionally more similar and potentially more evolutionarily conserved across mammals, despite their complex transcriptional profiles (**Supp Fig S7C**).

Primate non-neuronal types are regionally specialized

Non-neurons match or exceed the abundance of neurons in primate brains⁴⁷. Our spatial dataset resolved 3.4 M non-neuronal cells (65% of all spatial cells). As we enriched for neurons (Methods) prior to multiome sequencing, the final multiome dataset comprised a lower proportion of non-neurons (93,500, 28% of all sequenced cells post-QC). We resolved 17 Groups and 113 Clusters of non-neurons (**Fig 3A, Supp Fig 9A**). Hierarchical clustering of RNA expression in non-neuronal Groups revealed largely expected groupings of populations, such as the distinction between cells of neuroglial origin (astrocyte, oligodendrocyte and ependymal cells) and those of mesodermal origin (vascular cells and microglia) (**Fig 3A**).

The non-neuronal hierarchical clustering also revealed some unexpected transcriptional associations. Pinealocytes (Cluster_1854; sole member of the Pin Group) (**Supp Fig S9B**) produce and secrete melatonin and are the principal cell type of the pineal gland. While pinealocytes have been catalogued with neurons in other atlases¹, they are not considered true neurons⁴⁸. In the MSCA, pinealocytes were transcriptionally most similar to tanycytes (Cluster_840; sole member of the Tanycyte Group) (**Supp Fig S9C**), specialized ependymal cells that are found in the 3rd ventricle (3V). Pinealocytes are thought to be evolutionarily related to photoreceptors and retain direct photosensitivity in non-mammalian vertebrates, and potentially in mammals⁴⁹. Tanycytes, which participate in neuroendocrine regulation by sensing and transmitting hormones and other signals from the blood and CSF to the hypothalamus, are also sensitive to melatonin and light levels⁵⁰. Tanycytes derive from diencephalic radial glia along the floor and walls of the third ventricle (3V), while pinealocytes derive from *Pax6*⁺ neuroepithelial cells along the roof of the 3V⁵¹. The shared diencephalic neuroepithelial origin and participation in neuroendocrine signalling within these two cell types may contribute to their transcriptional similarity. The relative similarity between retinal ganglion cells and pinealocytes in the context of these results should be established.

We resolved expected populations of oligodendrocyte lineage cells, including oligodendrocyte progenitor cells (OPCs), committed oligodendrocyte precursor (COPs), immature oligodendrocytes (ImOligo) and mature oligodendrocytes (Oligo). Consistent with lower heterogeneity of oligodendrocyte subtypes compared to other neuroglial types⁵² we resolved two main Groups and 12 clusters. Overall, oligodendrocyte lineage Groups had less spatial diversity relative to astrocytes and microglia (next sections). Both the Oligo OPALIN and Oligo PLEKHG1 Groups were enriched in the white matter, with the Oligo OPALIN Group slightly more prevalent in gray matter structures (**Fig 3B**). OPCs formed a single cluster (Cluster_706) and were fairly evenly distributed across the subcortex, while ImOligos tended to avoid dorsal striatum and other telencephalic structures (**Fig 3B**).

Ependymal diversity maps to discrete ventricular domains

The mouse whole-brain atlas¹ resolved multiple subtypes of ependymal cells, including several with previously uncharacterized spatial specificity. We resolved 11 clusters of ependyma in the MSCA. As we lacked multiome samples and spatial sections around the 4th ventricle and spinal cord, we likely underestimate the regional and transcriptomic diversity of tanycyte and ependymal cells. Based on mapping to the mouse atlas and expression of *TTR*, Cluster_780, Cluster_787, and Cluster_788 are likely choroid plexus ependymal types. Choroid plexus clusters were found in the posterior lateral ventricles (LV) but did not map to the anterior LV (Cluster_780 and Cluster_788) (**Fig 3C**). Cluster_782 corresponded to mouse hypendymal cells and was localized exclusively to the posterior 3V (**Fig 3C**). Other marmoset ependymal clusters exhibited many-to-many relationships to mouse ependymal clusters (**Table S2**). Spatial mapping indicated strong regional biases of several of the marmoset ependymal populations, including two ependymal clusters that were selective to the 3V (Cluster_784, Cluster_785) (**Fig 3C**).

Astrocyte subtypes reflect developmental boundaries and anatomical axes

Astrocytes are regionally heterogeneous in the mammalian brain. Within the neocortex, different molecular and morphological subtypes reside in gray and white matter⁵³. Studies that sampled multiple brain structures indicate further molecular distinctions that reflect classical developmental boundaries, such as differences between the telencephalic and diencephalic astrocytes^{52,54}. In our atlas, astrocytes formed three major Groups: Astro-TE NN, Astro-NT NN, and Astro-WM NN. Group membership predominantly reflected the expected distinction between telencephalic and non-telencephalic astrocytes, as well as between gray and white matter, with some exceptions (**Fig 3D**, top; **Supp Fig S9D,E**). For example, two abundant astrocyte clusters (Cluster_681 and Cluster_684) generally avoided the telencephalon in favor of diencephalic structures, but also were strongly enriched in the septum (**Fig 3D**, middle). While spatially overlapping, these two clusters also had different anterior-posterior and dorsal-ventral biases: Cluster_681 was preferentially found in the thalamus and posterior non-telencephalic structures, while Cluster_684 was proportionally more abundant in hypothalamus and anterior structures. Dorsal-ventral gradients were evident within telencephalic-preferential astrocytes as well: Cluster_808 was relatively enriched in dorsal striatum and amygdala, while Cluster_807 favored ventral striatum and hippocampus (**Fig 3D**, bottom). We also observed smaller populations of astrocytes that were spatially biased to specific brain structures, including hippocampal formation and/or amygdala (Cluster_810 and Cluster_812), dentate gyrus (Cluster_813), thalamus (Cluster_677), hypothalamus (Cluster_683 and Cluster_690), and septum/ventral pallidum (Cluster_682) (**Supp Fig S9F**).

Microglia are highly enriched in marmoset BNST

Microglia and other brain-resident immune populations such as macrophages derive from hematopoietic stem cells that migrate into the CNS early in development prior to blood-brain-barrier closure. At early postnatal stages in mice, microglia cluster around ventricles and in the developing white matter before continuing to migrate throughout the parenchyma⁵⁵. By postnatal day 12-14 in mouse, microglia are “tiled” across major compartments of the brain⁵⁶ which enables them to efficiently survey local tissue environments. Rodent microglia exhibit regional differences in morphology and transcriptional identity, as well as in overall densities⁵⁷. However, absolute and relative density measurements across regions have varied across studies⁵⁸. As regional heterogeneity of microglial populations has mostly been studied in mouse or in developing human/primate (when microglia migration has not yet concluded)⁵⁹⁻⁶¹, we sought to characterize the extent of microglial regional specialization in the adult marmoset subcortex.

We sampled 18,369 microglia in the multiome dataset and resolved one Group and 14 Clusters. Some Clusters appeared to have disproportionate contribution from particular tissue dissections, suggesting molecular subtypes may exhibit regional biases. Spatial mapping indicated that the most prominent increase in microglial density was found in the BNST and hypothalamus (**Fig 3E**). This abundance change was driven by particular Clusters (Cluster_854, Cluster_855, Cluster_857 and Cluster_858; **Fig 3E**). While microglia are known to concentrate around the ventricles and deep white matter during development⁵⁵, this does not account for the observed regional enrichment in BNST, as our spatial and multiomic samples were acquired from adult animals. We noted that the BNST and anterior hypothalamus exhibit increased overall cell density

in Nissl preparations in some marmoset atlases^{62–64}. Moreover, ISH expression of *TREM2* – a marker for microglia – mirrors the increased cell density in these areas^{63,64} (**Supp Fig S9G**), corroborating the observation that marmoset microglia appear to be absolutely and proportionately more abundant in these structures (**Fig 3E**).

Spatial conservation and divergence between gene expression and chromatin accessibility

Gene regulatory networks across primate subcortex.

To uncover gene regulatory programs underlying primate subcortical diversity, we inferred gene regulatory networks (GRNs) using SCENIC+⁶⁵. This approach integrates chromatin accessibility, gene expression, and transcription factor binding motif analysis to identify enhancer-driven regulons (eRegulons) linking transcription factors with their predicted binding sites (in ATAC peaks) and the associated target genes. In total we identified 550 eRegulons (321 direct, 229 extended) in neuronal cells and 299 eRegulons (173 direct, 126 extended) in non-neuronal cells. These eRegulon catalogs establish a foundation for exploring the modular regulatory logic and spatial organization that shapes cell-type diversity across the primate subcortex.

Many eRegulons are active in specific classes of neurons or glia (**Fig 4A, Supp Fig S10A**). One example of this cell-type specificity is the eRegulon associated with the transcription factor *OTX2*, which is highly active in the thalamic GABAergic neuron Group (**Fig 4A**). In mouse, *Otx2* (the homolog of *Drosophila orthodenticle*) is expressed in the telencephalon, diencephalon, and mesencephalon during development⁶⁶ and is regulated by Wnt and Fgf signalling to specify the midbrain/hindbrain boundary^{67,68}. Within neurons, we found 142 putative direct targets of *OTX2*, including *GRIK1*, which encodes the GluK1 subunit of kainate receptors (**Fig 4B**). SCENIC+ predicts *OTX2* activity in multiple peaks associated with *GRIK1* expression, including the intronic peak shown in **Figure 4C**, which contains a consensus *OTX2* binding site and whose accessibility is highly correlated with *OTX2* and *GRIK1* expression. Associations like these form the basis of SCENIC+ gene regulatory network (GRN) inference, and the associated peaks are hereafter referred to as candidate cis-regulatory elements (cCREs).

Spatial projections verify that *GRIK1* and other neuronal targets of *OTX2* are largely but not exclusively restricted to the thalamus (**Fig 4D**). Other Groups with high *OTX2* expression include midbrain GABAergic types (particularly those associated with the superior colliculus) as well as ependymal glia cells and tanycytes (**Fig 4E**), similar to cell type patterns in the adult mouse brain¹. Consequently, we also discovered an *OTX2* eRegulon in non-neurons, adding an additional 69 putative *OTX2* targets. Interestingly, while some *OTX2* targets are strongly expressed in both *OTX2*⁺ neurons and glia, a large fraction show substantially differential expression patterns, including the TFs *GATA3* and *DMBX1* (in neurons) and *LMX1B* and *PDLIM5* (in glia) (**Fig 4F**).

We hypothesized that other cell-type specific TFs might be acting with *OTX2*, which could explain its apparent cell type specific activities. We found that cCREs for cell-type specific target genes also show cell-type specific accessibility (**Supp Fig S10B-C**). This is consistent with the idea that

OTX2 binding sites are only accessible in the presence of some other co-acting TFs. Looking for TFs that are co-enriched in cell-type specific OTX2 cCREs, we found strong neuronal co-enrichment with *DMBX1* and *GATA3* (**Supp Fig S10D**), which have n=36 and n=20 shared gene targets, respectively. While *DMBX1* is confounded by motif similarity with OTX2, co-immunoprecipitation of OTX2 and *DMBX1* also suggests they physically act together⁶⁹. In the glial cells, *NFIA* is enriched for acting in the same cCREs as OTX2, though the absolute number is small (n=4 OTX2+*NFIA* targets).

Regulatory signatures of regionally confined and dispersed neurons

Regional identity exerts persistent influence on brain cell type transcriptomes^{1,20,36}. However, extensive cell migration during development can result in similar cell types residing in different structures. For example, telencephalic interneurons derive from ganglionic eminence (GE) progenitors and migrate widely across the mammalian forebrain^{70,71}. Individual progenitors can give rise to daughter cells that populate multiple distinct brain structure^{72,73}. Consistent with their shared origin, hierarchical clustering based on RNA expression shows that telencephalic interneurons cluster preferentially by initial progenitor class (e.g. medial, caudal, or lateral – MGE, CGE, or LGE; **Fig 1B**).

To better understand the influence of region and developmental origin on gene regulatory programs, we regrouped our regionally localized neuron clusters according to their anatomical region (based on spatial mapping) and major neurotransmitter usage (based on snRNA-seq). We further subdivided GE-derived neurons (all GABAergic or cholinergic) according to their structure of origin (MGE, CGE, or LGE), and we also separated Medium Spiny Neurons (MSNs) from other GABAergic neurons. We then assessed how these groups of neurons compared to one another in terms of the eRegulons they use (**Fig 5, Supp Fig S11**).

Among GABAergic neurons, MGE and CGE-derived interneurons show similar eRegulon usage regardless of their final anatomical residence (**Fig 5A**). This is especially true of MGE-derived interneurons, which primarily consists of two lineage-related subtypes (*PVALB*⁺ or *SST*⁺) that are highly similar regardless of their final telencephalic localization (cortex, hippocampus, or striatum) (**Fig 5B**). However, even these similar cell types have diverged, either as a prerequisite for, or as a consequence of their distinct migrations. When we compared transcription-factor expression between cortical and striatal MGE interneurons, *NKX2-1* emerged as a consistent striatum-enriched regulator in both the *PVALB* and *SST* subtypes (**Fig 5C**).

NKX2-1 is a known marker of MGE progenitors during development and has previously been shown to be down-regulated in cortex-bound MGE interneurons^{74,75}. Focusing on the TF-TF regulatory interactions of *NKX2-1* (**Fig 5D**), we find an interconnected network in which *NKX2-1* activates downstream effectors such as *LHX6*, *LHX8*, *ISL1*, *ARX*, *DLX1/6*, and *PPARA*, creating a self-reinforcing transcriptional circuit that maintains striatal identity. Many of these partners (*LHX6/8*, *ISL1*) are known striatal regulators previously shown to be downstream of *NKX2-1*^{76,77}. Conversely, previously identified cortical TFs (including *SATB1*, *SATB2*, ETS-family, *MEF2C*,

TCF/NR3C1) are specifically enriched in cortex-localized cells and lie outside of the NKX2-1 activator network^{78–81}.

Aside from the well-known role of *NKX2-1* in MGE interneuron migration, we detected striatum-enriched *NFIB* in both PVALB and SST interneurons (**Fig 5C**). Our eRegulon analysis indicates that *NFIB* activates the chemorepellents *SLIT3* and *EFNB2*, both of which are also co-enriched in striatal interneuron subtypes. Slit-Robo and Ephrin signaling have previously been implicated in MGE interneuron migration^{70,82}, and NFI transcription factors regulate forebrain GABAergic maturation⁸³. These findings are consistent with our regulatory network inference and support the hypothesis that *NFIB* regulates interneuron migration through Slit-Robo and Ephrin signaling pathways.

In contrast to GABAergic neurons, glutamatergic neurons show a stark regional divide in their eRegulon usage that mostly separates telencephalic from non-telencephalic structures (**Fig 5D**). Interestingly, in a striking case of unexpected regulatory convergence, the main outgroup among GABAergic neurons – the LGE-derived neurons, primarily MSN projection neurons – also clusters with the telencephalic Glutamatergic projection neurons (**Supp Fig S11A**). In hierarchical clustering on samples from many brain regions, these two classes of cells tend to be closely proximal in both mice and primates, suggesting conserved transcriptional similarity^{20,28,29}. Although LGE-derived neurons (including MSNs) and telencephalic glutamatergic neurons arise from distinct progenitor domains^{84,85}, they converge on a similar set of eRegulons (**Supp Fig S11B-D**). These shared eRegulons include TFs previously associated with both excitatory telencephalic neurons and the LGE (*EGR3*, *E2F3*, *RCOR1*, *FOXG1*)^{86–90}, as well as TFs with known roles in excitatory telencephalic neurons but not previously described in LGE-derived cells (*RFX3*, *MBNL2*)^{91–94}. These overlapping regulatory programs likely reflect convergent functional requirements rather than shared developmental lineages.

Previously, we found the gene *CRACDL* to be among the most highly enriched in both striatal MSNs and telencephalic glutamatergic neurons²⁰. *CRACDL* encodes a brain-enriched, vertebrate-specific protein that localizes to the nucleus and to microtubules and the centrosomes (Human Protein Atlas)⁹⁵. We again found specific *CRACDL* enrichment in striatal MSNs and telencephalic glutamatergic neurons (**Fig 5E, top**). *CRACDL* is a target gene in eight of our eRegulons, three of which are “+/+” direct activators (where TF/peak-with-motif/target gene activity are all positively correlated) (**Fig 5E, bottom**). Of these, only the eRegulon associated with *SMAD3* (of the TGF- β superfamily) has high activity in the striatal MSNs; the other two eRegulons (*TCF4* and *ZBTB18*) are specific to glutamatergic neurons (and both are predicted to be direct regulators of the vesicular glutamate receptor *VGLUT1/SLC17A7*). To get a broader view of this convergent regulatory landscape, we constructed a *CRACDL*-centered eRegulon network and compared its activity across the telencephalon Glutamatergic and striatal MSN types (**Supp Fig S11E**). This extended network contains a number of self-reinforcing activating interactions with *TCF4* in the glutamatergic types, but this interconnected network is mostly inactive in the striatal MSNs. This shows that *CRACDL* expression is ensured by a robust network in the excitatory projection neurons of the telencephalon, while the striatal MSNs activate only a part of this network to satisfy a convergent functional requirement.

Epigenetic regulation of gene expression gradients in primate hippocampal formation

Spatial mapping of primate hippocampal formation.

The hippocampus is considered ‘archicortex’ – a phylogenetically ancient component of the cerebral cortex. In most mammals and birds, the hippocampal formation is located dorsomedially just below the corpus callosum, consistent with its early patterning by the cortical hem^{71,96}. With the expansion of the neocortex and temporal lobe in particular, the primate hippocampal formation shifted away from its ancestral location and occupies a more ventral position. Leveraging our dense spatial sampling, we examined the complete cellular architecture of the marmoset hippocampus along its longitudinal (anterior-posterior) extent. HPF structures were detected in 25/47 MSCA spatial sections. 11 Groups and 104 clusters in our taxonomy spatially mapped to the hippocampal formation (HiF). Six of the Groups largely reflected distinctions between hippocampal subfields (**Figure 6A**).

Generally, cross-species mapping of HPF Groups to the whole mouse brain atlas¹ corroborated the Group level spatial mappings, suggesting broad evolutionary conservation with some modifications. As an example, we focused on the indusium griseum (IG) – a tiny limbic structure adjacent to the corpus callosum. The IG is linked to the dentate gyrus via the fasciola cinereum (FC), a delicate layer that wraps around the splenium of the corpus callosum. The IG remains proximate to the dorsal hippocampus in rodents. In primates, it retains a dorsal location vestigial to the ancestral position of the hippocampus, even though the hippocampus has shifted ventrally. Recently, a spatial transcriptomic study in mice suggested similarity between IG and CA⁹⁷, while other reports suggest greater similarity of IG neurons to DG neurons^{98,99}. Cluster_955 within Group CA2-CA3-IG spatially mapped to the marmoset IG as well as to a highly restricted location in the hippocampus corresponding to the posterior CA3 field, at or near the border with CA4 (**Supp Fig S12A**). Cross species comparison indicated a correspondence to mouse clusters 0400-0401 in Yao et al¹, which were grouped into the mouse “025 CA2-FC-IG” supertype (**Supp Fig S12A**). While the marmoset IG-CA3 cluster spanned the A-P extent of the callosum and was highly restricted to the posterior-most sections of CA3, mouse IG cells were present in anterior parts of the callosum and were restricted to dorsal hippocampus. The mapping to dorsal mouse hippocampus and posterior marmoset hippocampus is consistent with the hippocampal longitudinal axis rotation between rodents and primates, while the relative positional mapping and transcriptional correspondence in each species suggest that IG may be more closely related to CA subfields than to DG.

To facilitate analysis of A-P variation within and across Groups, we first conducted pairwise alignment between all spatial sections across the z-axis using PASTE2¹⁰⁰. The alignment algorithm leverages spatial proximity and gene expression similarity of cells on adjacent slides to adjust the x and y coordinates of each cell. Importantly, the alignment is independent of the cluster identity yet the post-alignment results show that cells from the same Clusters occupy comparable spatial locations across sections, indicating continuous transitions along the anterior-posterior axis (**Fig 6B**).

Cell types in hippocampal subfields vary along the anterior-posterior axis

Dorsal-ventral (D-V) gradients of function, connectivity, and gene expression shape the topography of mouse hippocampus^{1,101,102}. The mouse D-V axis corresponds to the longitudinal (Posterior to Anterior) hippocampal axis in primates. Accordingly, gradients of gene expression have also been described along the A-P axis of the human¹⁰³ hippocampus. Focusing on glutamatergic neurons, we observed marked spatial localization of individual clusters along the marmoset hippocampal A-P axis (**Fig 6A-B, Supp Fig S12**). Subfield- and A-P-restricted clusters were found within major hippocampal structures including DG, CA1, CA2, CA3, and were also present in Pro/Subiculum (ProS/SUB). For example, Clusters_942-945 were found only in the CA3 head, Clusters_924-926 were in CA3 head and body, Clusters_930-931 were in CA3 body/tail, and Cluster_949 was only in CA2/3 tail (**Fig 6A**). Similarly A-P restricted Clusters were found in DG, CA1 and ProS (**Fig 6A**). Based on expression of known markers (*CALB2+PROX1-*), we observed two Clusters of Mossy cells (Cluster_953 and Cluster_957) that spanned the entire DG A-P axis, while Cluster_956 Mossy cells were found only in the DG head (**Supp Fig S12B**).

Consistent with the rotational difference (from D-V to P-A) along the hippocampal longitudinal axis in primates, several Clusters that localized to the anterior hippocampus in the marmoset mapped to cells in the mouse ventral hippocampus. For example, supertype 0076 CA3 Glut_2 of the whole mouse brain atlas is found in ventral CA3 subfields and mapped to CA3 head and anterior body MSCA Clusters within the CA3 Glut Group. Conversely, marmoset Clusters_950-952 were biased to the posterior body and tail of CA3 and mapped to mouse supertype 0078 CA3 Glut_4, located in dorsal CA3 (**Supp Fig S12C**).

Gene expression gradients along the primate hippocampal A-P axis.

To examine the logic of A-P distributions of HPF Clusters in more detail, we embedded snRNAseq Clusters in each Group on a 1-dimensional (1D) UMAP space and compared each Cluster's median 1D coordinate to its modal spatial z-section in Xenium (Methods). In Groups CA3 Glut, CA1-ProS Glut, and SUB-ProS Glut, median 1D UMAP positions showed strong correlation with longitudinal spatial z coordinates, indicating that the 1D UMAP axis recapitulates the A-P positions of each cluster (**Fig 6C, Supp Fig S12D**).

Focusing on Groups that showed strong correlation between Cluster 1D UMAP and the corresponding modal z position spatially, we sought to identify genes with significant AP gradients by computing the cell-level correlation between gene expression and their 1D UMAP coordinates (Methods). Combining results from all three Groups tested, 168 genes were enriched anteriorly, showing decreasing expression toward the posterior pole, and 136 genes were posterior-biased with increasing expression towards posterior sections ($p\text{-adj} < 0.05$, $|\text{spearman } r| > 0.5$). We then examined whether genes that correlated with A-P position were private to or shared across groups. Surprisingly, most genes with longitudinal graded expression (286/304 genes, 88%) were private to one of the three Groups (**Fig 6D**).

To assess whether these Group-specific longitudinal programs are conserved across primates, we compared our results to Ayhan et al¹⁰⁴, who used single-nucleus sequencing to compare anterior and posterior dissections of the human hippocampus. We found significant overlaps

between differentially expressed genes reported in the human CA1 and gradient genes in the marmoset CA1-ProS Group ($p\text{-adj} < 0.05$; **Fig 6E**), suggesting conservation in genes that drive expression gradients in the primate hippocampus.

Gene regulatory gradients along the primate hippocampal A-P axis.

In classical morphogen gradients (such as Bicoid in *Drosophila*), target enhancers can be activated or repressed in a concentration dependent manner. In such cases, the enhancer is competent everywhere but is only active when the TF concentration is sufficiently high. In other instances – such as spinal cord patterning – enhancers themselves show graded accessibility along the patterned axis¹⁰⁵. Distinct enhancers are active at the rostradorsal and caudoventral poles of the mouse hippocampal primordium in development¹⁰⁶. However, how gene expression gradients along the adult primate hippocampal longitudinal axis arise still remains elusive. Motivated by these models, we hypothesized there are four gene regulatory configurations that could contribute to the observed gene gradients above (**Fig 6F**). Under scenario A, both the TF and its associated enhancer show graded abundance or accessibility, predicting a reinforced gradient in target gene expression. Under scenario B, the concentration gradient of TF alone encodes position while in scenario C, the TF is relatively uniform and spatially restricted CRE accessibilities gates the global TF input. Under scenario D, neither TF expression nor enhancer accessibility is detectably graded, suggesting that there are other epigenetic mechanisms underlying the observed gene gradients.

To test these scenarios, we extended our gradient analysis to compute cell-level correlation between binned accessibility of putative enhancers identified in all eRegulons and the 1D UMAP coordinates for each Group (Methods). Using eRegulon triplets linking TFs, enhancers and target genes, we calculated the percentage of gradient target genes under each possibility outlined in Fig 6F. In CA3 Glut neurons, the largest fraction of gradient genes were associated with triplets in which both TF and enhancer were graded (**Fig 6G**). Triplets in which neither component was graded only accounted for a small proportion of gradient genes.

To illustrate this with a concrete example, we next examined a representative eRegulon triplet in the *TCF4*^{+/+} activator eRegulon in CA3 Glut. To investigate whether all components of a graded eRegulon (i.e. the TF, the enhancer, and the target gene) show the same spatial trend, we first examined *TCF4* itself and found that its expression is significantly graded and increases posteriorly. Within the *TCF4*^{+/+} eRegulon, the Ephrin receptor *EPHA5* (a downstream target gene of *TCF4*) likewise exhibits a significant gradient (**Fig 6H,I**). We further validated this target gene gradient by directly visualizing log counts per thousand (CPT) of *EPHA5* expression of CA3 Glut cells measured in the independent Xenium spatial transcriptomics data in the 3D-aligned space (**Fig 6I**), and by examining ISH of *EPHA5* in the Marmoset Gene Atlas (**Supp Fig S12E**). Accessibility of the putative enhancer that has *TCF4* motifs and correlated accessibility with target gene *EPHA5* also shows similar correlation with A-P position. Together, our results suggest that longitudinal gene expression gradients in excitatory hippocampal cell types may be primarily governed by the coordinated activity of spatially graded TFs and enhancers, although additional regulatory mechanisms are likely involved.

Evolutionary modifications to neurons in primate thalamus

Spatial organization of glutamatergic populations.

The mammalian thalamus is a collection of largely glutamatergic nuclei. Each nucleus forms reciprocal loops with different areas of the neocortex and basal ganglia to mediate sensory, motor, or higher-order function^{107–109}. Like neocortex, primate thalamus has disproportionately enlarged territories associated with higher-order function, such as the medial pulvinar¹² and mediodorsal nucleus¹⁰⁷. We found that the spatial locations of the 19 thalamic glutamatergic Groups distinguished known anatomical subdivisions, including the anterodorsal nucleus (AD), the intralaminar nuclei (ILN), the lateral dorsal nucleus (LD), the ventrolateral nuclei (VLN), and the lateral geniculate nucleus (dLG) (**Fig 7A**).

Laminar organization of the primate Lateral Geniculate Nucleus (LGN) and shared cell type with inferior pulvinar (IPul)

The dorsal lateral geniculate nucleus (dLG) receives direct projections from retinal ganglion cells and has undergone evolutionary reorganization and relocation, detaching from the main thalamic body. In primates, the dLG is a laminated structure; in rodents, it has a core-shell organization^{110,111}. The inferior pulvinar nucleus (IPul), adjacent to the dLG, has also expanded significantly in primates to accommodate reciprocal connections with visual cortices¹¹². We found three groups concordant with anatomical divisions of these neighboring thalamic nuclei (**Fig 7B**). Two distinct excitatory groups (LGmc-LGpc Glut and LGko Glut) recapitulate the four-layer laminar architecture of the dLG in marmosets (**Fig 7B**). We identified a third excitatory group (LGko-IPul Glut) that spans the dLGN-IPul boundary (**Fig 7B**). This shared population is consistent with prior microdissection-based transcriptomic studies and immunohistochemistry^{21,113}, suggesting potential convergent connectivity or shared functions across these nuclei.

Glutamatergic populations in primate-expanded medial pulvinar

Unlike the inferior and lateral pulvinar, the medial pulvinar – a specialized nucleus of the primate thalamus that connects to prefrontal cortex and may be involved in distinct aspects of visual cognition – does not have a clear anatomical homologue in rodents^{12,114}. Previous work indicates that by the perinatal period, nuclei of the marmoset pulvinar exhibit gene expression differences not found in mouse pulvinar¹⁴. Consistent with molecular distinctions across marmoset pulvinar nuclei, Cluster 728 is biased to lateral pulvinar, while Cluster 729 was biased to medial pulvinar (**Fig 7C**). Cluster 730 was localized to the cap of the medial pulvinar (**Fig 7C**). Direct comparison of Cluster 728 and Cluster 729 revealed differential expression of 486 genes with more than two fold enrichment in either cluster ($p\text{-adj} < 0.05$). Notably, among them are genes reported previously to delineate the primate medial and lateral pulvinar, such as *CALB1* and *NNAT*^{14,115,116}.

Primate-expanded thalamic inhibitory neurons are transcriptionally related to midbrain neurons

Aside from the RT, the rodent thalamus is almost entirely glutamatergic. In mouse, non-RT thalamic inhibitory neurons comprise only ~6% of all thalamic neurons, most of which are found in the lateral geniculate nucleus (LGN)^{9,117}. Unlike in rodents, primates have abundant GABAergic populations in both sensory and higher order thalamic nuclei (Hunt 1991, Dixon 2001), with

reports estimating as many as 30-50% of neurons in some thalamic nuclei are GABAergic. From our spatial data, we estimate that 20% of marmoset thalamic neurons (excluding those in RT, HN, and dLG) are GABAergic (**Fig 7D**).

Prior studies suggest the expanded population of primate thalamic GABAergic neurons may arise from multiple origins. Reports in nonhuman primates have found evidence for diencephalic and telencephalic origins^{118,119}, or midbrain and diencephalic GABAergic populations⁹, while a study in human midgestation found evidence for midbrain, diencephalic and telencephalic thalamic GABAergic populations¹²⁰. Though many fate-specifying genes are developmentally regulated, adult transcriptomes tend to strongly reflect cells' developmental lineage^{1,20,36}. We reasoned that transcriptional similarities between thalamic GABAergic populations and other cell types might provide additional clues as to their origins.

Non-RT thalamic GABAergic populations comprise three Groups in the MSCA: THM GABA, THM-MB GABA, and MB-THM GABA (**Fig 7D**). Several Clusters within these Groups were restricted to particular thalamic nuclei, such as Cluster_38, which form an LGN-specific GABAergic population (**Supp Fig S13A**). However, many GABAergic Clusters were distributed across both thalamic and midbrain nuclei (**Supp Fig S13B**). Based on spatial mapping, 26% of cells belonging to THM GABA, THM-MB GABA, and MB-THM GABA groups were found outside of the thalamus. The observation that the marmoset midbrain and thalamus share many GABAergic Clusters led us to wonder whether some or all of the primate-expanded thalamic GABAergic population emerged from a conserved midbrain population present in mouse. Supporting this possibility, THM GABA, MB-THM GABA and THM-MB GABA groups in marmoset were assigned to a single mouse subclass: 213 SCsg Gabrr2 Gaba in cross-species mapping (**Supp Table S3**). This subclass predominates in the superior colliculus of the mouse midbrain, though also extends a sparse population to the mouse LGN (**Supp Fig S13C**).

In addition to marmoset to mouse mapping, we also inspected the direct pairwise reciprocal mapping results between the human brain atlas³⁶ and marmoset, as well as an additional snRNA-seq mouse whole brain atlas generated by Slide-seq² to comprehensively search for transcriptionally similar cell types to the THM GABAergic cells across mammalian species. MSCA cells from all three THM GABA Groups were assigned to the Midbrain-derived inhibitory Supercluster from Siletti et al³⁶, which contains samples from microdissections of thalamus, midbrain and pons (**Fig 2A, Supp Fig S13B**). Between primate datasets (human and marmoset) and mouse datasets, this population was consistently dominated by predictions to mouse midbrain and tegmental neurons in both mouse datasets^{1,2}, reinforcing a conserved midbrain-like transcriptional identity for non-RT thalamic GABAergic cells¹.

Primate thalamic inhibitory neurons share CREs with midbrain neurons

Using our SCENIC+ eRegulons, we found that midbrain GABAergic neurons were the most similar to thalamic GABAergic neurons (**Supp Fig S14A, S11A**). Some of the most strongly co-enriched eRegulons are associated with the TFs *GATA3*, *OTX2*, *BACH1*, and *ESRRG* (**Supp FigS14B**).

Aside from any active eRegulon usage, we hypothesized that the chromatin landscape might retain additional evidence of cell lineage, for instance, if some accessibility patterns are carried over from an earlier developmental state. Earlier, we found that midbrain glutamatergic and GABAergic neurons conspicuously co-enrich for numerous midbrain-specific genes and peaks (**Supp Fig S4E**). If primate-expanded GABAergic populations in the thalamus originate from that same conserved midbrain GABAergic population, we might expect retention of some of these highly midbrain-specific marker peaks in their chromatin accessibility profiles. To find these markers, we ranked ATAC peaks according to their specificities for midbrain glutamatergic and midbrain GABAergic cell types. We then defined midbrain markers as the set of peaks that maximize shared specificity for these two types (see Methods). While this same procedure produced highly distinct, regionally-specific peak sets for most brain regions (**Supp Fig S14C**), we found that a number of otherwise highly midbrain-specific peaks are also accessible in thalamic GABAergic neurons (**Fig 7E**). Projecting the eigenvalue for this peak set onto our spatial data, we can clearly visualize how these peaks are shared among glutamatergic and GABAergic midbrain neurons, but are otherwise only present in thalamic GABAergic neurons (**Fig 7F**).

Developmental data in mice suggests that excitatory and inhibitory neurons of the superior colliculus originate from a common pool of resident, multipotent progenitors¹²¹. Taken together, the cross-species mapping and transcriptomic and regulatory concordance between thalamic GABAergic and midbrain neurons supports the conclusion that primate GABAergic neurons in the non-RT nuclei of the thalamus primarily originate from midbrain sources⁹ and whose expansion constitute a primate adaptation compared with rodents.

Discussion

Understanding how the brain supports diverse computations requires detailed knowledge of cell populations within and across brain structures, as well as the extent to which they are shared between species. In this study, we spatially resolved a taxonomy of molecularly-defined cell assemblies across marmoset subcortical structures. The dense spatial sampling, coupled with cross-species mapping of brain cell atlases in mouse and human, enable a comprehensive picture of spatial and molecular variation in the mammalian brain. Across the cell taxonomy, relative changes in chromatin accessibility and gene expression are largely aligned, but not perfectly concordant, suggesting that transcriptional identity is mostly (but not always) mirrored by cis-regulatory architecture. As an illustration, we show that spatial gradients in the hippocampus reflect spatially graded transcription factor expression coupled with permissive enhancers. We identify lineage and regional context as major influences that shape transcriptional and gene regulatory identity of cell types. These influences are prominent in the case of primate-expanded thalamic GABAergic neurons, which we find share the greatest similarity to species-conserved midbrain populations.

A surprising outcome is the extent to which the resolved Groups and even Clusters of the snRNA-seq defined taxonomy mapped to anatomically precise locations in the spatial atlas. The spatial gene panel, which contained 300 genes (only 23 of which were transcription factors), was constructed using an existing snRNA-seq dataset that sampled a limited number of marmoset subcortical structures²⁰, and was moreover optimized for cross-species integration of basal ganglia structures²⁵. Despite this, most Groups and a substantial proportion of Clusters had anatomically plausible spatial mapping across the subcortex. This suggests that a relatively modest panel of genes is sufficient to accurately capture differences between a large number of neuronal populations. Given that small cell populations had lower Cluster prediction accuracy, and that average distances between ground-truth Clusters and erroneously predicted Clusters was low (**Supp Fig S4A**), additional snRNA-seq sampling and larger gene panels will further refine spatial mapping. However, our results suggest that ‘non-optimal’ gene panels with modest numbers of genes can still capture substantial diversity between large sets of cell populations. More broadly, focusing cell type definitions on quantitative variation in gene expression, perhaps more so than on the presence of markers or an arbitrary set of differentially expressed genes, could support the concept of cell types as “peaks” in the joint probability distribution of expression levels of many genes¹²². Such a framework can be especially powerful in the context of cross-species comparisons of evolutionarily divergent cell populations^{4,5}.

Enhancers can be exploited to develop genetic tools targeted to a particular brain cell type^{20,123–127}—in many instances, enhancers have cell type specific activities in a distinct anatomical location and are not accessible in transcriptionally similar cell types located in different brain structures. Such regionally restricted activity underscores how enhancers can maintain both cell type and regionally-restricted fingerprints (**Figs 5-7**). Conversely, the same enhancers can also be shared across distinct spatial locations and unrelated cell types. The spatially-resolved nature of our cell taxonomy facilitates screening and prediction of enhancer activities across regions and cell types in the primate subcortex to better anticipate such “off-target” cases (**Supp Fig S5**).

The functional organization of the hippocampus is associated with molecular and connectivity gradients along its major axes in rodents and primates^{102,103}. We established that cell types within distinct hippocampus subfields participate in subfield-specific gene regulatory networks. While transcription factors and target genes show spatially graded expression, the enhancers in these networks are also spatially graded across the longitudinal axis (**Fig 6**). This suggests that hippocampal molecular gradients, unlike classical morphogen gradients, require the concerted orchestration of gradients across each actor of the GRN. Whether and how the establishment of gradients in different subfields are orchestrated across development will be important to understand.

Because postmitotic brain cells usually remain within their developmental compartment of origin, regional identity and cell type identity are typically intertwined^{128,129}. Telencephalic interneurons, which originate from a common origin and migrate to distinct tissues^{80,130,131}, constitute an exception. Such cases enable lineage and regional identity to be decoupled. Our eRegulon-based comparisons show that MGE- and CGE-derived interneurons retain highly similar gene-regulatory programs regardless of their ultimate destination, yet exhibit specific regulatory adjustments (**Fig 5**). Sustained *NKX2-1* and *NFIB* activity in MGE-derived striatal interneurons may be remnants of migration-associated regulatory programs that continue to shape their molecular identities. In contrast, glutamatergic projection neurons display a sharper regional divide, separating telencephalic from non-telencephalic lineages. Intriguingly, LGE-derived medium spiny neurons, though GABAergic, converge transcriptionally with telencephalic glutamatergic projection neurons, with both expressing TFs like *RFX3*, *EGR3*, *MBNL2*, and *RCOR1*, as well as the genes they regulate (**Supp Fig S11**). This convergence appears to be an example of evolutionarily and developmentally distant neuron classes adopting similar gene regulatory programs to meet common functional demands.

Migration events that cross developmental boundaries are a strong test for disentangling lineage and regional context. We found that the expanded population of GABAergic neurons in the primate thalamus, which are largely absent from the mouse thalamus, likely arose from a conserved population of midbrain GABAergic neurons, particularly those in the superficial layers of the superior colliculus. Once in the primate thalamus, they retain a signature of chromatin accessibility that is shared with GABAergic and glutamatergic populations of the midbrain (**Fig 7**). Differential migration of conserved cell types offers a potent way for evolution to reorganize brain structures and circuits to shape species-specific functional properties^{3,7}. For example, using comparative analysis of neural precursors across development, a recent study showed that a population of LGE-derived olfactory bulb neural progenitors in mouse follows a distinct dorsal migratory stream in macaques and becomes redirected to the prefrontal cortex⁷. Future work that identifies shared lineages between primate-expanded thalamic GABAergic populations and their conserved midbrain counterparts will be important for understanding how and when this migration occurs, as well as the functional consequence of such adaptations to primate subcortical circuits.

In addition to diverse regionally localized neuronal populations with developmental imprints, we also uncover spatially biased nonneuronal populations, particularly astrocytes and microglia (**Fig 3**). Astrocyte Groups segregate along developmental (telencephalon vs. non-telencephalon) and grey vs. white matter boundaries, while individual Clusters show more granular anatomical

selectivity. Unlike astrocytes that tile the brain relatively uniformly, marmoset microglia are disproportionately concentrated in BST and anterior hypothalamus, indicating that neuroimmune signaling may not be uniformly distributed across the primate subcortex. Together, these findings demonstrate that nonneuronal architectures comprise regionally tuned cell types that likely contribute to circuit specialization and structure-specific disease vulnerability. Future research dissecting the functional properties of each of these glial cell types is needed to better understand how they shape and modulate local circuit dynamics.

Leveraging the tri-modality characterization in the marmoset subcortical cell atlas, we are able to link high-resolution molecular taxonomies to three-dimensional anatomical localization, offering a spatially grounded template of primate subcortical cell type diversity (**Fig 1**). Beyond serving as a reference for the marmoset brain, this atlas also provides a foundation for refining existing taxonomies and anchoring future ones. Through joint cross-species embedding, we disambiguate closely related heterogeneous human cell types (**Fig 2**, e.g. Splatter) and refine spatial annotations within complex anatomical subdivisions (**Fig 6-7**, e.g. amygdala, hypothalamus, thalamus and hippocampal formation), increasing the interpretability of existing cell atlases. Based on the successful creation of a cross-species basal ganglia cell type taxonomy conserved across marmoset, macaque, and human³¹, we anticipate our subcortical atlas will be valuable for analysis in other primate species, including humans. Another powerful application will be to situate human genetic risk variants within spatially-resolved primate cell types and anatomical structures¹³², which could increase biological interpretability of such efforts.

Ongoing work in BICAN will eventually integrate these data into forthcoming cross-species whole brain atlases. We anticipate that the comprehensive sampling enabled by the compact size of the marmoset brain will serve as a bridge for elucidating the structural and molecular organization of primate brains across evolution. By making these data available through our Cytosplore Viewer and other platforms, we provide a multimodal resource applicable to a broad range of neuroscience applications.

Experimental model and subject details

Animals

Marmoset experiments were approved, overseen, and conducted in compliance with the Massachusetts Institute of Technology IACUC under protocol number 2303000479.

Method details

Tissue processing

Adult marmosets (6 individuals, 4M and 2F, 2-7.5 years old with a median age of 5.6 years; **Supp Table S5**) 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 $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, and 10 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 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 $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 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.

Nuclei isolation

Single nuclei isolation of marmoset samples for 10x sequencing was conducted as described in ([dx.doi.org/10.17504/protocols.io.5qpvk1exl4o/v2](https://doi.org/10.17504/protocols.io.5qpvk1exl4o/v2)). Briefly, frozen marmoset neural tissue microdissections ranging from 14 to 60 mg were placed into a 3 mL dounce homogenizer containing 2 mL homogenization buffer (2M sucrose, 2M KCl, 1M Tris, and 1M MgCl_2 , 10% Triton X-100, 100X Protease Inhibitor Cocktail, Protector RNase Inhibitor, and 100 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 μ m 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 blocking buffer and DAPI (4',6-diamidino-2-phenylindole) at a final concentration of 0.1 μ g/mL.

Fluorescence-activated nuclei sorting (FANS)

Nuclei suspensions were sorted with the Sony MA900 cell sorter using a 130 μ m 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 μ L 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 μ L above the invisible pellet of nuclei. An equal amount of 2X nuclei buffer (20X nuclei buffer (PN2000153), Protector RNase inhibitor, 100 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 μ L 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 μ L was ideal for the 10X Chromium for a recovery of 10,000 cells/ μ L.

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

Full experimental details for this dataset are described in Hewitt, Turner, et al.³¹. Briefly, a fresh hemisphere from a five-year-old male marmoset was slabbed into six, 5-mm thick, coronal slabs from anterior-to-posterior, which were then flash frozen. Slabs were sectioned on a Leica cryostat at 10 µm onto 10X Xenium slides and collected at 200 µm intervals. A panel of 300 genes was constructed from a combination of computationally-derived and manually selected marker genes, using an early version of the Basal Ganglia Consensus Taxonomy²⁵ and computational search via mFISHtools¹³³ and GeneBasis¹³⁴. The gene panel was designed to target basal ganglia cell types and select overlapping genes with parallel macaque and human gene panels. 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” (<https://zenodo.org/records/16915694>), a version of NCBI GenBank/RefSeq GCF_011100555.1/GCA_011100555.2 modified to contain mitochondrial gene annotations generated using the mitos2 webserver (<https://doi.org/10.1093/nar/gkz833>) and mitochondrial chromosome CM021961.1 (<https://doi.org/10.1186/s13059-021-02336-9>).

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 (v1.30) (<https://doi.org/10.1093/bioinformatics/btw354>). 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 cellbender¹³⁵ (v0.3.0) to remove ambient RNA. scDbIFinder¹³⁶ (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 Seurat¹³⁷ (v5.0.3) before exporting to h5ad.

Scanpy¹³⁸ (v1.10.4) was used for the calculation of top 4000 highly variable genes (HVG) with “seurat_v3” as flavor and donor name as batch key. Using the HVG subset, we corrected for

donor batch effects by training a scVI²⁶ (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 (v1.0.0) with the same clustering parameters as described in the companion cross-species BG taxonomy paper²⁵. 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::HCPC¹³⁹ (v2.11). The resulting dendrogram was split into neurons and nonneurons, and visualized using ggtree¹⁴⁰ (v3.6.2) and ggnewscale¹⁴¹ (v0.4.10). The Cluster-level metacells are used throughout other analyses.

Marker gene identification and Differential Gene Expression Analysis

Marker genes for every Cluster and Group were computed using NSForest¹⁴² (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 DESeq2¹⁴³ (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 apegglm¹⁴⁴ (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, as defined by the species-to-species ortholog tables from Ensembl BioMarkt (<https://useast.ensembl.org/biomart/martview/72badb26965c442fe7469e8e96b8b3bd>). For large reference datasets with over 1 million cells, we used MapMyCells (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 are listed in Supplemental Table S4. Mapping results from each reference dataset are summarized per Cluster and Group in Supplemental Table S2 and S3.

Additional pairwise reciprocal mapping among whole brain human³⁶, mouse scRNA-seq¹ and mouse snRNA-seq² 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 and Turner et al³¹. 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 (**Supp Fig S1G**).

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 (RRID:SCR_024672) flat mapping algorithm with bootstrapping (100 iterations, bootstrap factor 0.95) and using all 300 genes from the gene panel (**Supp Table S6**) 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.

Generation of MSCA Taxonomy

Groups in the initial draft taxonomy were generated by aggregating clusters with similar transcriptomic signature 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)¹⁴⁵. Groups were then assigned to the corresponding Class derived from the mouse taxonomy¹. 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, Fu, Schmitz et al.²⁵.

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 mouse¹, marmoset MSCA, and human³⁶ 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 expressolog⁴⁶.

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 SnapATAC³⁴ (v2.8.0) (<https://doi.org/10.1038/s41592-023-02139-9>). 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 also generated a reduced peak set by filtering peaks with non-zero signal in >50% of the cluster-level pseudobulks in at least one Group, yielding 462,698 peaks for downstream analysis. bigWig files were generated for our taxonomic Groups using “export_coverage” (normalization="CPM", counting_strategy="insertion", bin_size=1).

Gene Regulatory Network (GRN/eRegulon) calling and scoring

Due to performance constraints, SCENIC+⁶⁵ (v1.0a2) was run separately on 50000 neuronal and 50000 non-neuronal cells sampled using geometric sketching¹⁴⁶ based on our RNA scVI latent space. We performed Topic modelling with pycisTopic³⁵ (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 pycistarget¹⁴⁷ (v1.1). Finally, we ran the SCENIC+ workflow with default settings producing 321 direct and 229 extended neuronal “eRegulons” (regulons), and 173 direct and 126 extended non-neuronal regulons. eRegulon activity was scored at the cluster-level using RNA or ATAC pseudobulks with the R package AUCell⁶⁵ (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 mean of the cluster z-scores is used.

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

Highly variable genes were selected as described above. A separate subset of the snRNAseq data was prepared using the 300 spatial gene panel from the Xenium spatial transcriptomics data. A 5-fold cross validation was performed on each of the 4000 highly variable genes snRNAseq data subset, 300 spatial gene panel snRNAseq data subset, and the 300 spatial gene panel Xenium data. Only the genes were subset; all cells were included for this comparison.

For classification accuracy, 80% of the data was used for training (4 folds) and cell type labels were predicted on the remaining 20% of data not used in training (1 fold); this process was repeated for the 5 folds for both the Group and Cluster labels. Distances between each cluster were computed for each data subset by first taking the mean across all cells in each cluster to use as the center for each cluster. Then, pairwise cosine distances were computed between the centers. The predicted distances were computed using a weighted average: for a given cell type A, the predicted distance for cell type A was calculated by weighting the distances to each cell type B (i.e. distance between A and B, for each cell type B in the set of all cell types) by the

proportion of the cells belonging to cell type A that were predicted to be cell type B (i.e. the predictions used to calculate the 5-fold cross validation accuracy). K-nearest-neighbors (KNN) classifiers from cuML were used for the 5-fold cross validation using cosine as the distance metric and 15 neighbors.

Cytosplore Viewer

Cytosplore Viewer leverages the fact that ATAC-seq and RNA-seq are annotated with the same, fine-grained clustering, which 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.

Alignment of Xenium Spatial Transcriptomics data in three-dimensions

To align consecutive Xenium slices, we used PASTE2¹⁰⁰, 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.

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 S2E**), 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 pseudoreplicate 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.

Hippocampal Gradient Analysis

Cells from hippocampal Groups CA3 Glut, CA1-ProS Glut, SUB-ProS Glut, and DG Glut in the multiomics and spatial transcriptomics datasets were included for this analysis. In the Xenium spatial transcriptomics dataset, because the hippocampus was only present from sections $z = 21$ onwards, any clusters mapped to anterior sections were excluded as low-confidence. In the multiomic dataset, additional clusters with fewer than 25 cells mapped across the remaining spatial transcriptomics sections were considered low-confidence and were excluded. 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. Clusters in DG Glut showed no significant correlation between 1D UMAP coordinate and spatial z position and were therefore excluded from subsequent analyses. To identify genes with A-P gradients in the hippocampus, we computed correlations between gene expression and the 1D UMAP coordinates in the multi-omics dataset. Similarly, we identified peaks with A-P gradients by correlating the binned chromatin accessibility ($n_bins = 50$) against the 1D UMAP coordinates due to the sparsity of the ATAC data. P values for all correlations were adjusted for multiple testing using Benjamini-Hochberg. To test the frequencies of each hypothesis in **Fig 6F**, we subsetting the eRegulon triplets identified above based on graded TF and/or enhancer peak, and then calculated the percentage of target genes that were significantly graded within each subset. All

Regional ATAC marker analysis

Related to **Figure 7E-F** and **Supp Fig S9E**, 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.

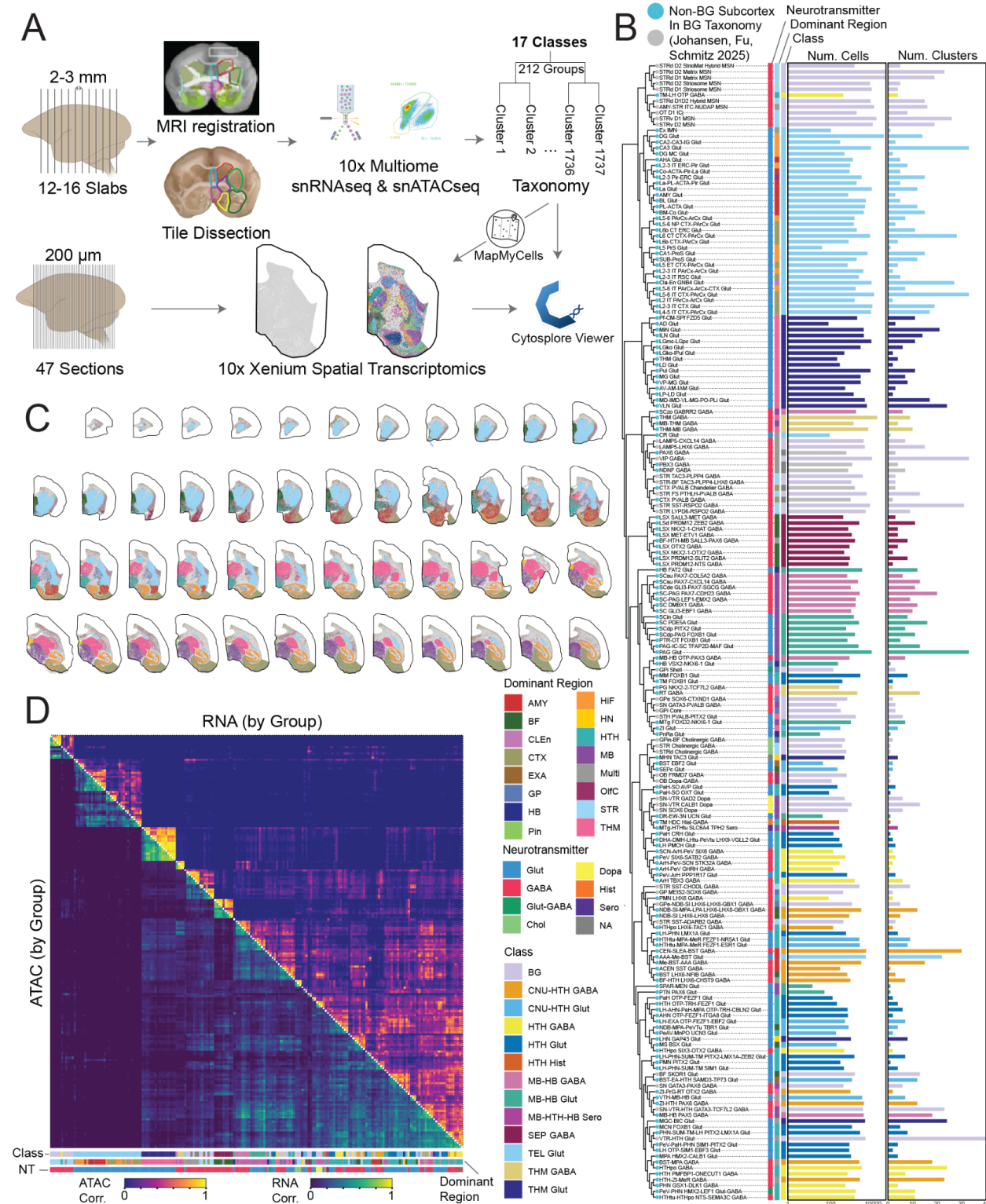
Data and Code Availability

Raw 10X Multiome data can be accessed at NeMO (<https://assets.nemoarchive.org/col-vdk0cmw>) and all processed data will be publicly accessible (<https://brain-map.org/consortia/hmba>). Code used in the analyses in the manuscript will be deposited at <https://github.com/KrienenLab/MSCA>.

Acknowledgements

This publication was supported by and coordinated through the Brain Initiative Cell Atlas Network (BICAN). This research was supported by the National Institutes of Health award UM1MH130981. The work was additionally supported by NIH DP2MH140136 and the Klingenstein–Simons Fellowship (F.M.K.), and the Netherlands Organization for Scientific Research NWO:024.004.012 (B.L.). We thank MIT DCM veterinarian staff and animal care technicians who assisted in anesthesia, post-operative 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.

Figures



and 20% other nonneurons for 10x Multiome sequencing. Downstream computational analysis generates a nested hierarchical cell type taxonomy of Class, Group and Cluster from multiomics data and predicts cluster labels on Xenium spatial transcriptomics samples using MapMyCells to obtain spatially-resolved cluster identities. A Cytosplore Viewer linking multiomics and spatial transcriptomics data is openly available for interactive data exploration. (B) 194 neuronal transcriptomic Groups hierarchically clustered based on the expression of top 4000 highly variable genes. Groups included in the companion cross-species consensus Basal Ganglia taxonomy²⁵ are indicated before group names. Each color tile after group names (from left to right) shows the neurotransmitter, spatially localized dominant region and Class identity of the Group. Bar plots indicate the number of cells and number of Clusters within each Group. (C) Overview of 47 Xenium spatial transcriptomics sections from anterior to posterior. Each cell is assigned a transcriptomic Cluster ID and colored by its corresponding dominant region label. (D) Heatmap showing pairwise spearman correlation of transcriptomic and chromatin accessibility similarity between all Groups. Columns and rows of the heatmap are in the order of the dendrogram in (B).

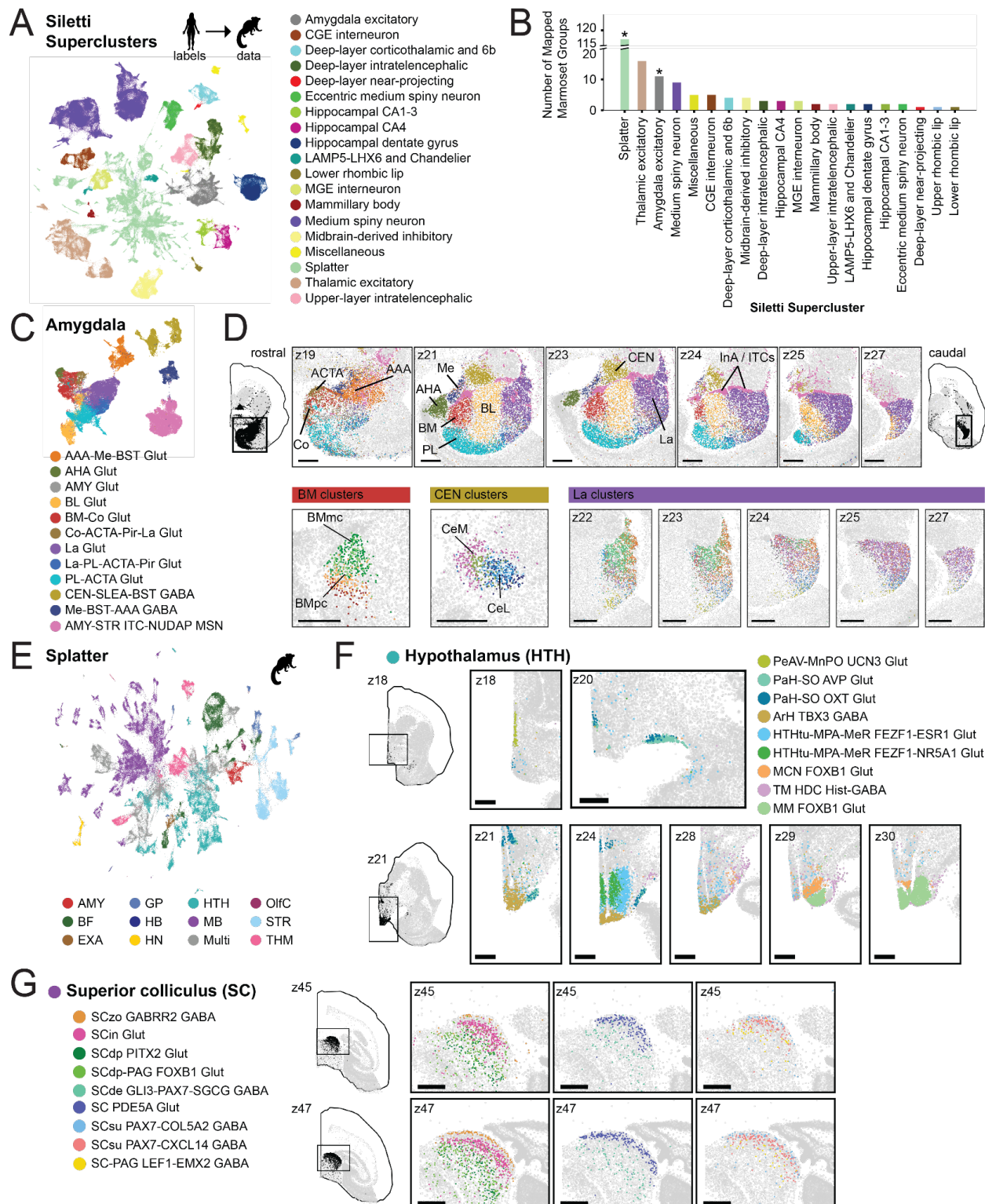


Figure 2. Cross-species spatial mapping of heterogeneous neuron populations. (A) UMAP embedding of marmoset subcortical neurons colored by predicted human Supercluster identity from Siletti et al³⁶. (B) Number of Groups in the marmoset subcortex atlas per human Supercluster with asterisks indicating spatially localized Superclusters presented in C-G. (C)

UMAP embedding of cells spatially localized in the amygdala colored by Group identity. (D) Top: Representative spatial sections of amygdala Groups (top) and Clusters within specific amygdala subdivisions (bottom). (E) UMAP embedding of marmoset subcortical cells in the human Splatter Supercluster colored by their dominant spatial location. (F) Representative spatial sections of heterogeneous HTH Splatter Groups. (G) Representative spatial sections of Splatter Groups in the laminated SC.

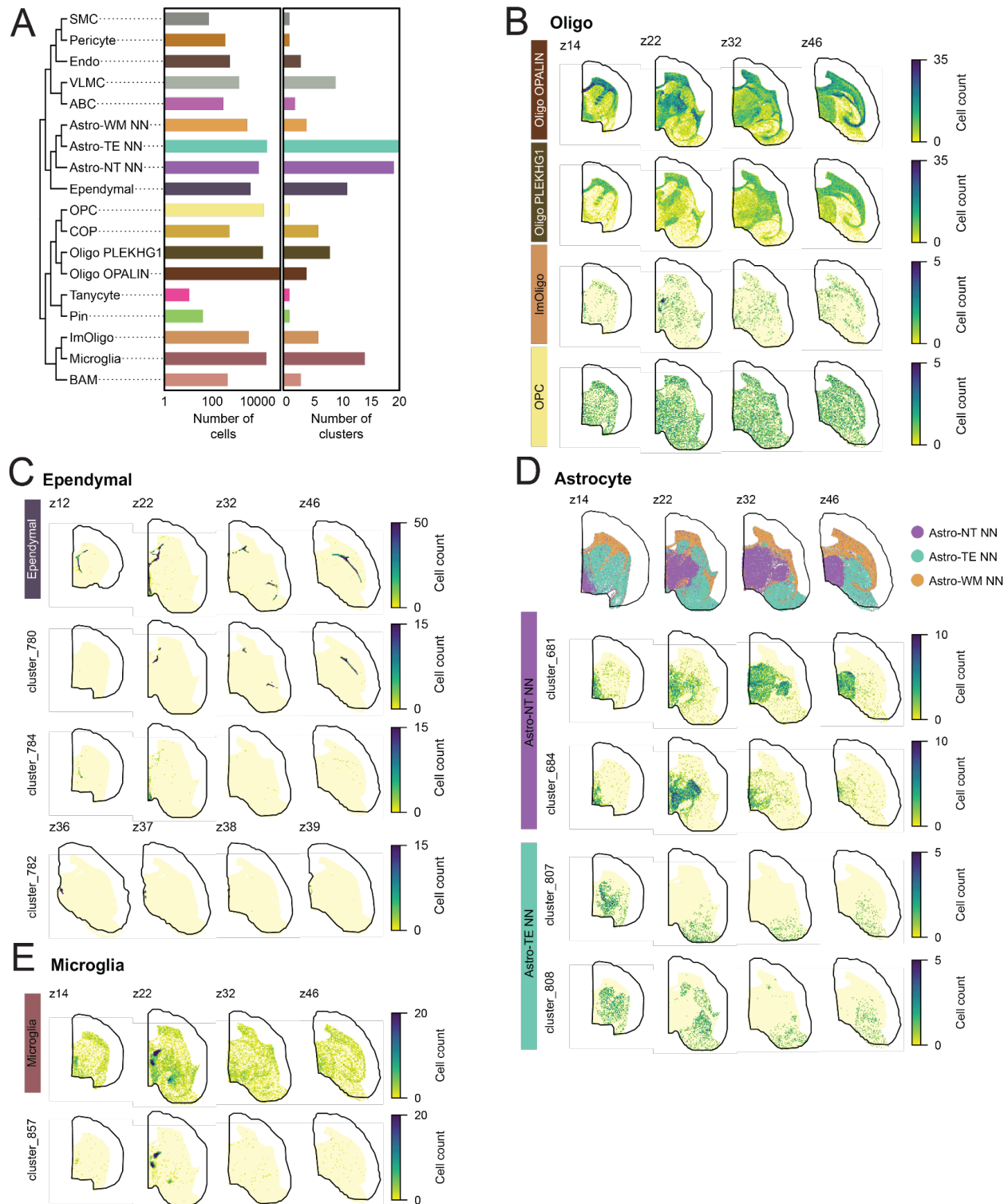


Figure 3. Regional zonation of non-neuronal cell types. (A) Hierarchically clustered tree of 18 nonneuronal Groups based on the expression of top 4000 highly variable genes. Bar plots indicate the number of cells and number of Clusters within each Group. **(B)** 2D histograms of cell counts for the four oligodendrocyte Groups (top to bottom): Oligo OPALIN, Oligo PLEKHG1, ImOligo, OPC. **(C)** 2D histograms of cell counts for the ependymal Group (top) and three

clusters with localization near specific ventricles: Cluster_780, Cluster_784, Cluster_782. **(D)** Regional zonation of the three astrocyte Groups: Astro-NT NN, Astro-TE NN, Astro-WM NN (top). 2D histogram of cell counts for two Astro-NT NN clusters with regional specialization: Cluster_681 and Cluster_68 (middle). 2D histogram of cell counts for two Astro-TE NN clusters with regional specialization: Cluster_807 and Cluster_808 (bottom). **(E)** 2D histogram of cell counts for the Microglia Group (top) and the predominant cluster driving the high density of microglia in the bed nucleus of the stria terminalis (BST), Cluster_857 (bottom). Bins for all 2D histograms are 150 um x 150 um.

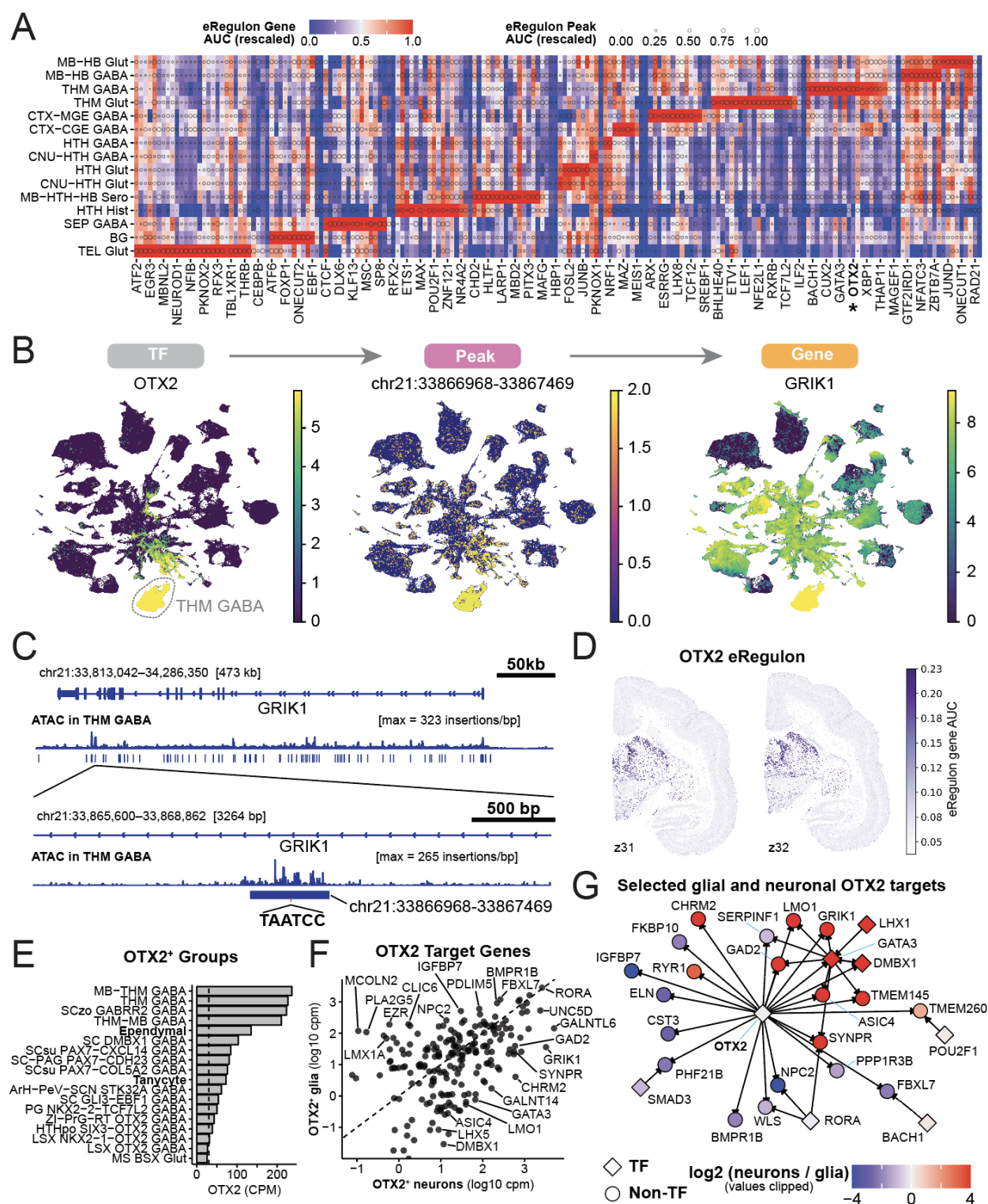


Figure 4. Gene regulatory networks across primate subcortex. (A) Direct activator (“+/+”) eRegulons are shown for each neuronal Class. Every third eRegulon is labelled. OTX2 eRegulon denoted by asterisk. (B) UMAP plots for an example eRegulon triplet. Peak accessibility and target gene expression are both correlated with OTX2 expression ($r=0.77$ and 0.57 , respectively). (C) Within Class THM GABA, ATAC-seq insertion events near the peak and

target gene from panel B. Peak is an apparent intronic enhancer and contains a perfect canonical OTX2 binding motif. (D) Representative spatial sections show high OTX2 eRegulon activity in the thalamus. (E) Top Groups by OTX2 expression. Glial groups are in bold. Groups with >30 CPM were called OTX2⁺ for subsequent analysis. (F) Expression of OTX2 target genes in neuronal vs. glial groups. (G) A small network plot for selected neuron- or glia-specific OTX2 targets. TF-Gene interactions only shown if SCENIC+ importance >1.5.

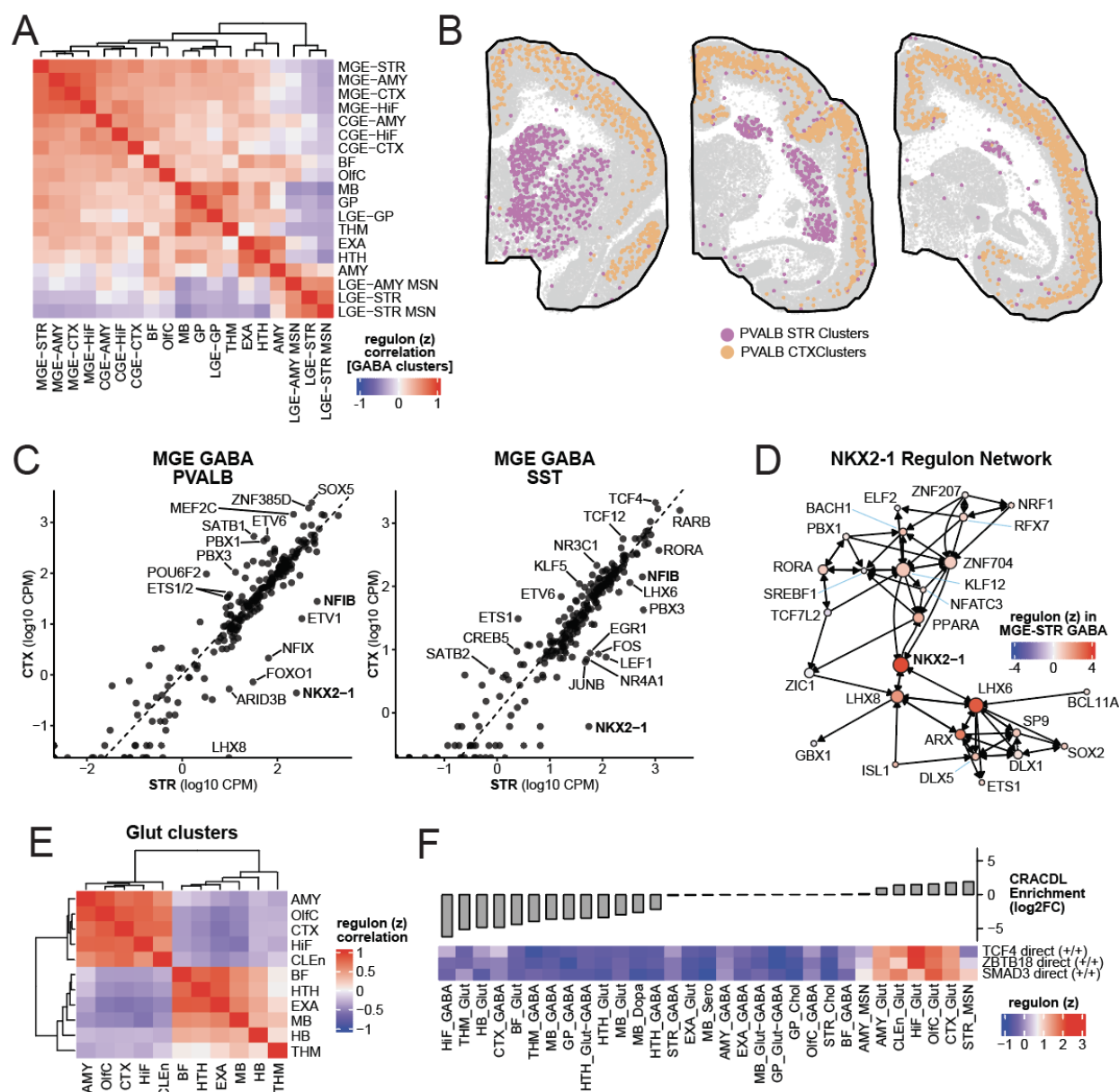


Figure 5. Regional dynamics of gene regulatory network activity. (A) GABAergic Cluster-level eRegulon scores (z-score of gene AUC) are averaged by region, with MSNs and GE-derived clusters separated. Pearson correlations are shown. (B) Spatial mapping of STR- and CTX-localized PVALB interneurons. (C) TF expression comparison in STR- vs CTX-localized clusters. (D) TF-TF eRegulon network centered on NKX2-1, colored by eRegulon activity (gene AUC z-score) averaged across MGE-STR GABA clusters. Network limited to two hops upstream or downstream of NKX2-1. (E) Glutamatergic clusters are compared, as in A. (F) CRACDL enrichment is apegm-shrunken log2FoldChange vs. all neuronal clusters localized to other regions. For eRegulon scores, cluster-level gene AUC is z-scored and averaged for each grouping.

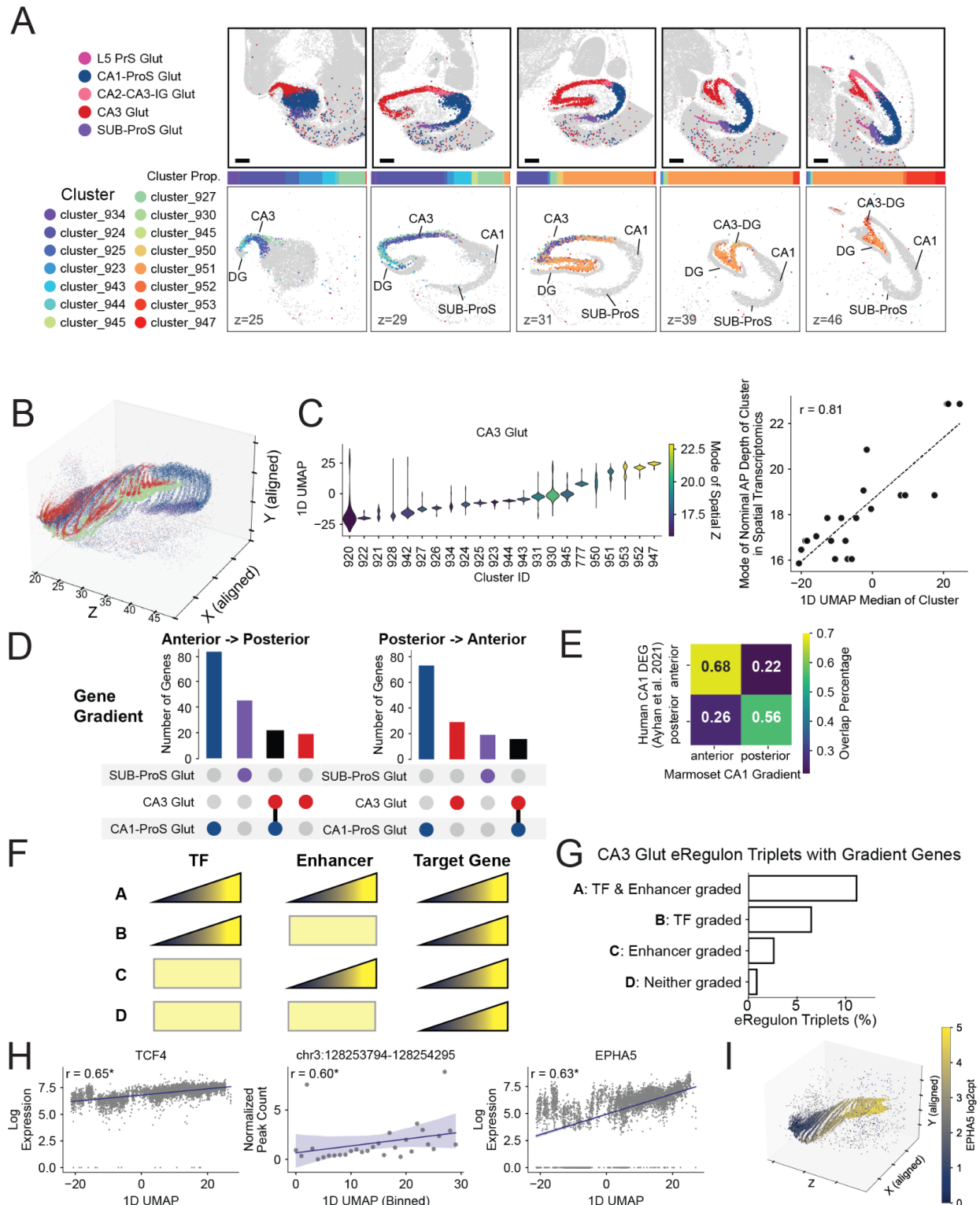


Figure 6. Epigenetic regulation of spatial gradients in primate hippocampal formation. (A) Representative spatial sections of Groups in hippocampal subfields (top) and Clusters in CA3 (bottom). (B) Three-dimensional representation of hippocampal subfields in spatially aligned sections. (C) Left: distribution of 1D UMAP coordinates for cells in each CA3 glutamatergic

cluster, with color indicating the modal spatial z-section (anterior-posterior) of the cluster. Right: scatter plot showing correlation of median 1D UMAP coordinates and modal spatial z-section for each cluster ($r = 0.81$; $p < 0.05$). (D) UpSet plots showing overlaps of genes whose expression is significantly correlated with 1D UMAP coordinates across hippocampal glutamatergic Groups. Genes were partitioned by the sign of their correlation with modal spatial z, yielding anterior->posterior (left) and posterior->anterior (right) gene sets. (E) Heatmap of percentage of marmoset CA1 gradient genes that overlap with differentially expressed genes in anterior and posterior human CA1¹⁰⁴. (F) Schematic of possible scenarios of gradients from TFs and enhancers that drive target gene gradients. (G) Percentage of eRegulon triplets with target gene gradients under each possibility outlined in (F) ($|\text{spearman } r| > 0.5$; $p\text{-adj} < 0.05$). (H) Correlation of example triplet (TF -> enhancer peak -> target gene) in TCF4 +/+ eRegulon. Left to right: Log gene expression of transcription factor TCF4 ($r = 0.65$; $p\text{-adj} < 0.05$), binned accessibility of enhancer peak chr3:128253794-128254295 ($r = 0.60$; $p\text{-adj} < 0.05$); log expression of target gene EPFA5 ($r = 0.63$; $p\text{-adj} < 0.05$) for each cell against their 1D UMAP coordinates. (I) Log cpt of *EPFA5* expression in cells of CA3 Glut in aligned spatial transcriptomics sections.

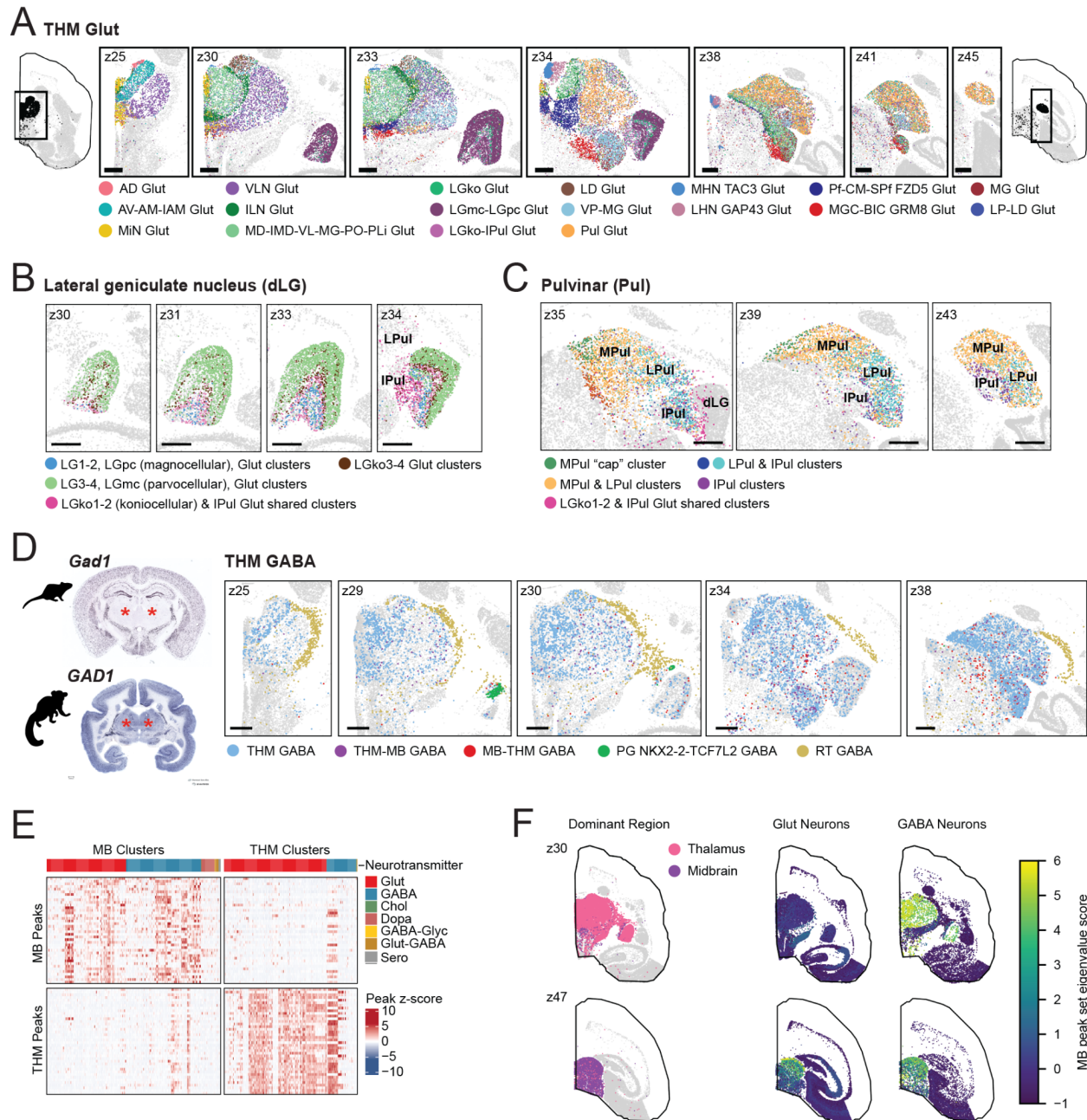


Figure 7. Evolutionary modifications to primate thalamus. (A) Representative spatial sections of Groups in THM Glut Class. (B) Representative spatial sections showing the laminar organization of the dLG. (C) Representative spatial sections of the Pulvinar nucleus highlighting clusters biased to the medial (Cluster_729) and lateral (Cluster_728) divisions. (D) (left) ISH with probes against *Gad1* in mouse (RRID:SCR_002978) and *GAD1* in marmoset^{14,64} showing absence and presence of GABAergic cells in the thalamus for mouse and marmoset, respectively. Location of thalamus indicated with red asterisks. (right) Representative marmoset spatial sections of Groups in THM GABA Class. (E) Heatmap showing regional marker peaks (shared in localized Glut & GABA neurons) in all clusters located within MB and THM. (F)

Representative spatial sections showing eigenvalue score of MB peak set in excitatory and inhibitory neurons in the THM (top) and MB (bottom).

Supplemental Table S1. MSCA taxonomy membership

Supplemental Table S2. Cluster-level summary of quality control statistics, marker genes, and cross-species mapping results.

Supplemental Table S3. Group-level summary of quality control statistics, marker genes, and cross-species mapping results.

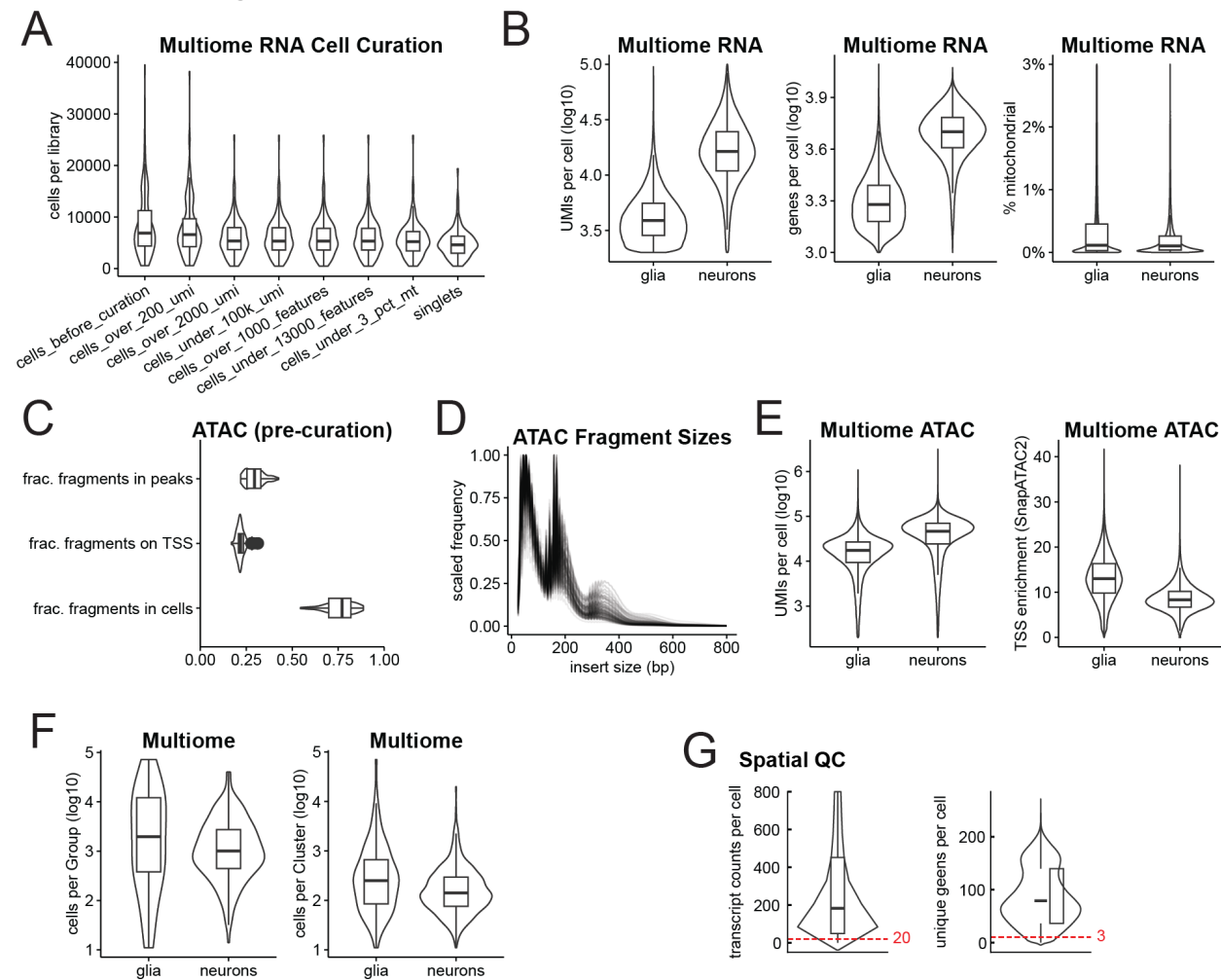
Supplemental Table S4. All public datasets referenced in MSCA and methods used for mapping.

Supplemental Table S5. Animal information.

Supplemental Table S6. Xenium Spatial Transcriptomics Gene Panel

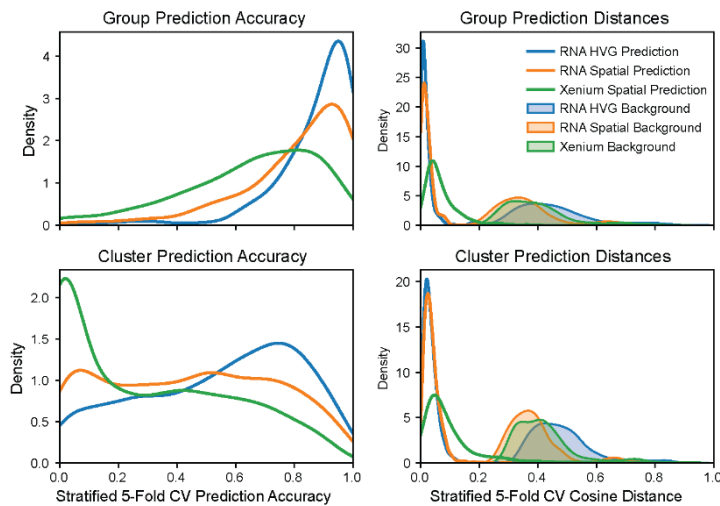
Supplemental Table S7. eRegulon triplets in MSCA

Supplemental Figures

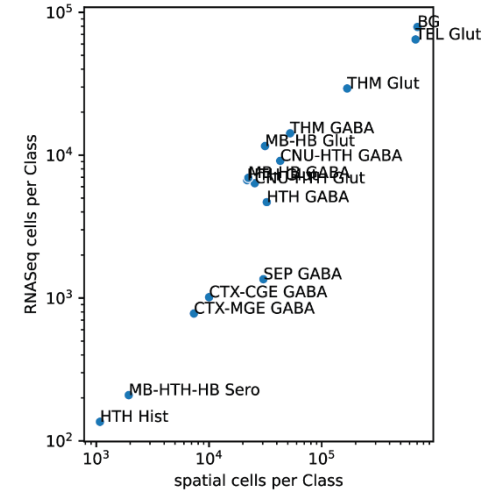


S1. Quality metrics of multiomics and spatial transcriptomics data, related to Fig 1. (A) Cell curation in multiome libraries, based on RNA metrics. Only libraries passing library-level QC are shown. (B) Multiome RNA metrics for curated cells. (C) ATAC metrics from cellranger-arc alignment summary for libraries passing library-level QC. (D) ATAC insert size distributions for all libraries passing QC, showing robust retention of nucleosomal fragments. (E) ATAC metrics for curated cells. (F) Multiome cells in each Group and Cluster. (G) Xenium spatial transcriptomics metric cutoffs for curated cells.

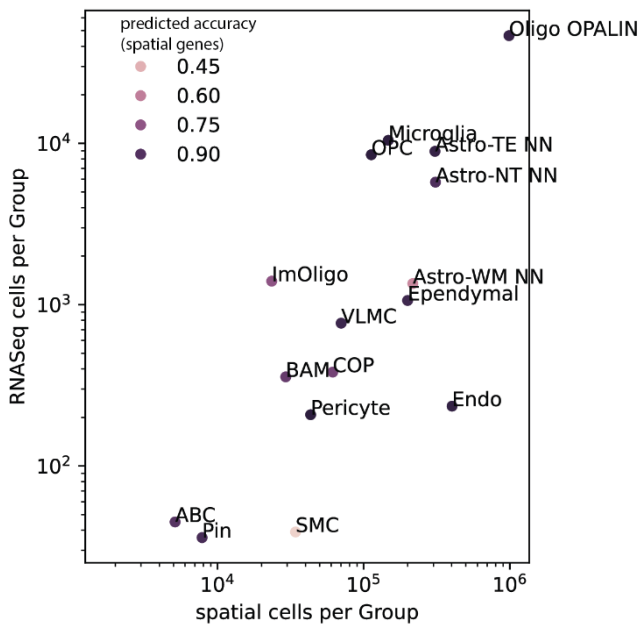
A Spatial Gene Panel (n=300) vs HVGs (n=4000) Cross Validation



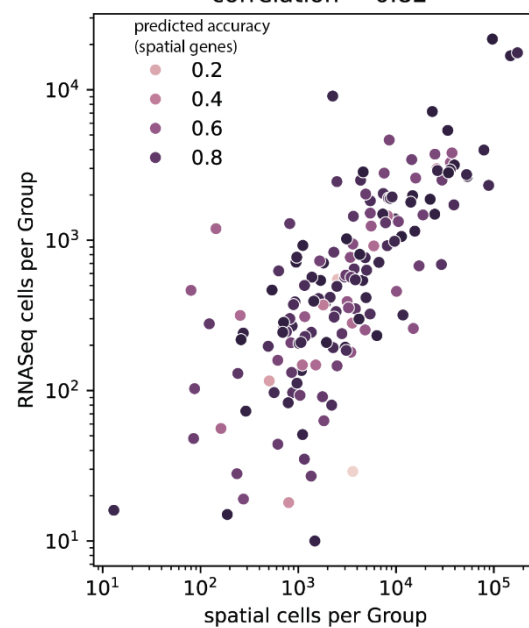
B Neuron Class populations correlation = 0.98



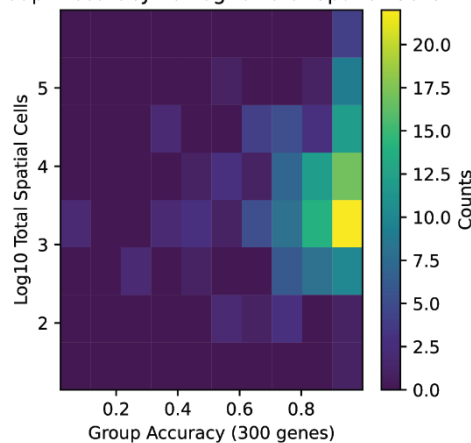
C Nonneuron Group populations correlation = 0.88



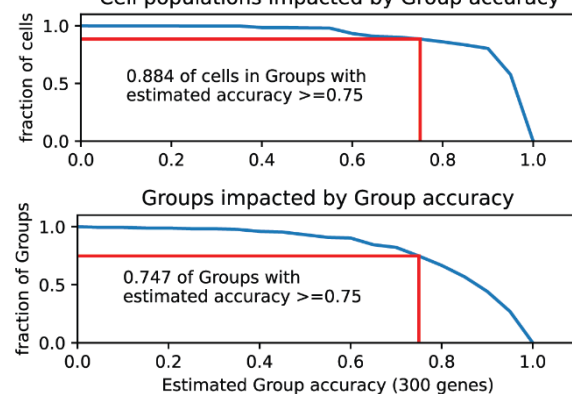
D Neuron Group populations correlation = 0.82



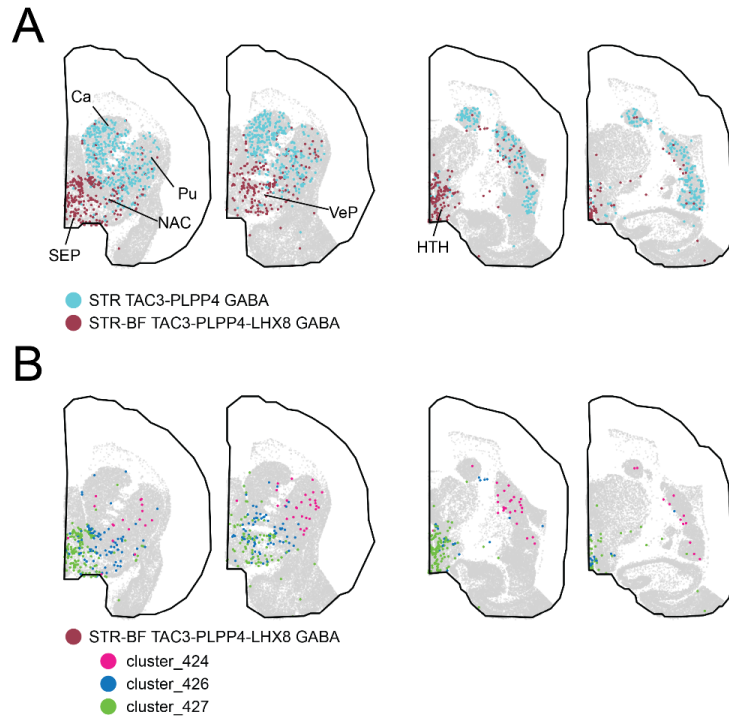
E Group Accuracy vs Log10 Total Spatial Cells



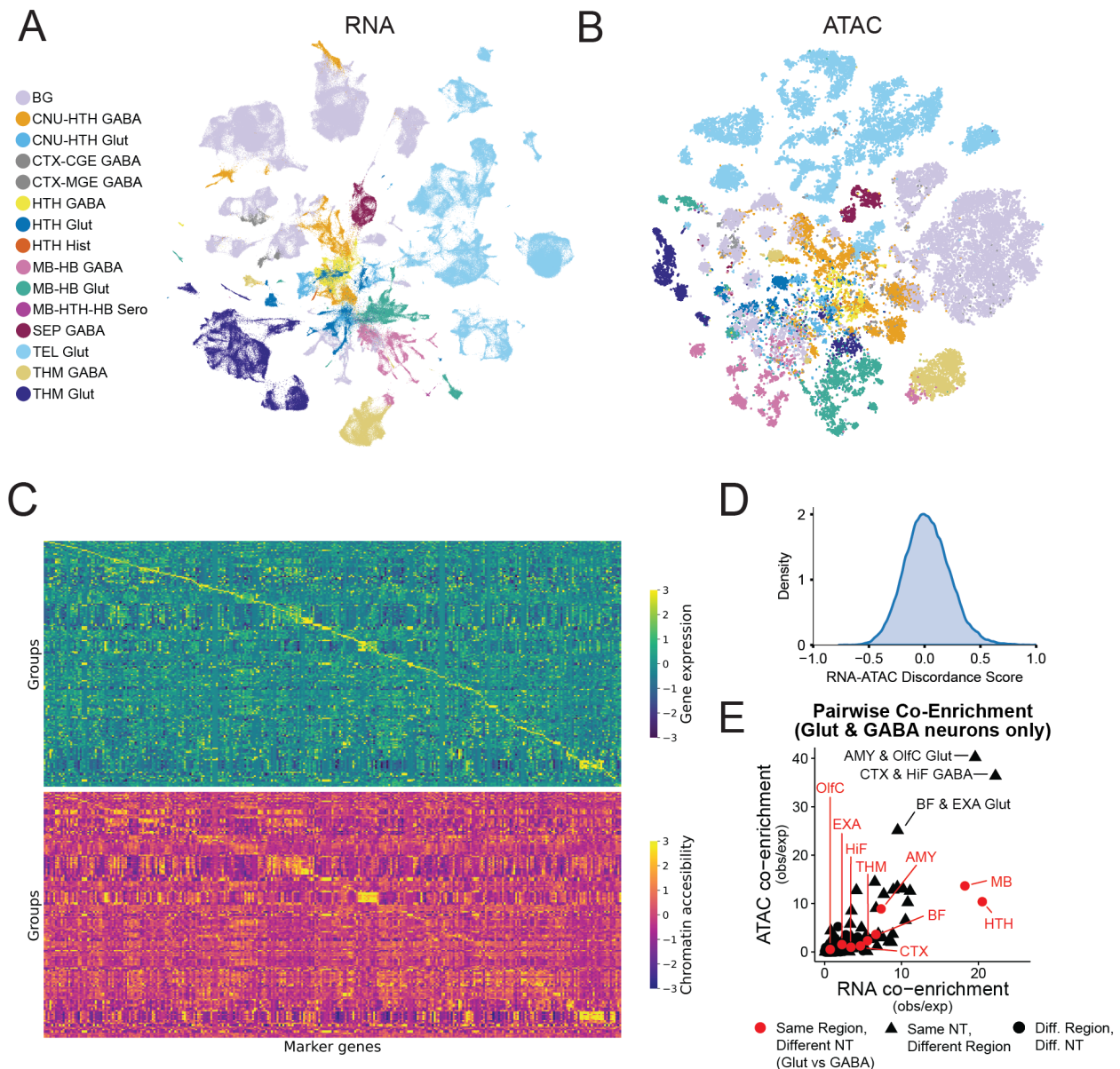
F Cell populations impacted by Group accuracy



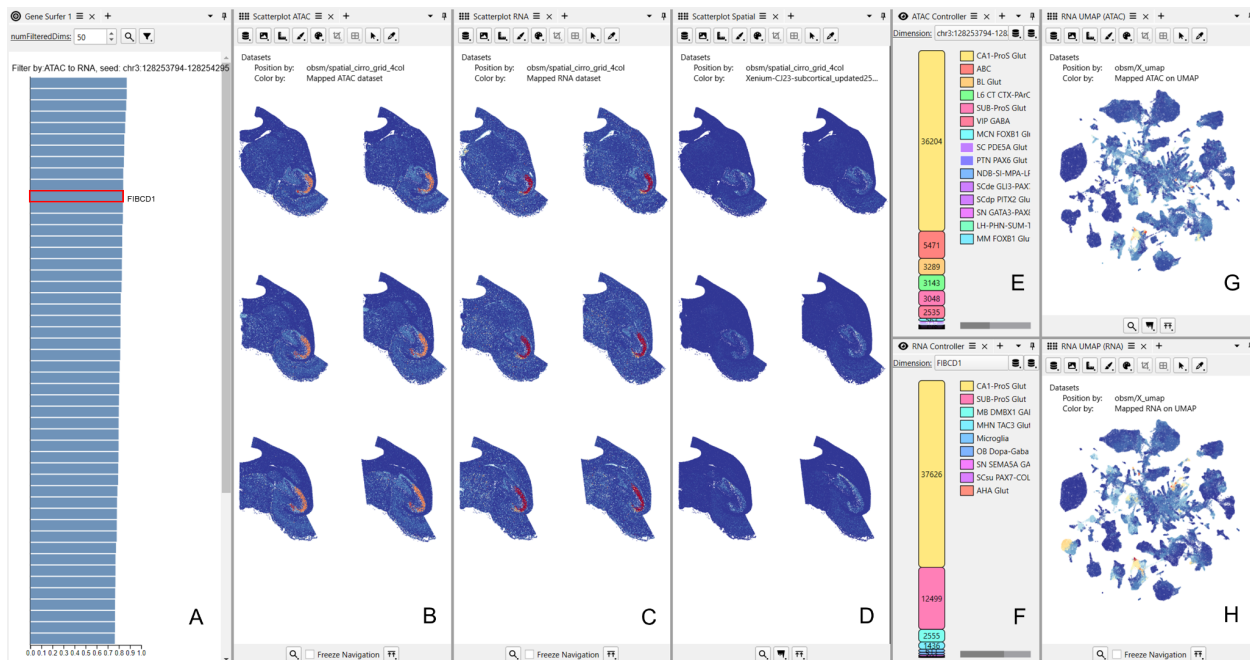
S2. Gene panel and label accuracy, related to Fig 1. (A) Predicted 5-fold cross-validation accuracies and distances of Group (top) and Cluster (bottom) labels in the RNASeq and spatial data. Using the subset of 300 genes from the spatial data reduced accuracy of Cluster labels but maintained high Group prediction accuracy, with mean predicted accuracies 0.84 and 0.82 for non-neuron and neuron Groups, respectively. Mispredicted cell types are closely related to the ground truth, as seen in comparison to the background, computed using a uniform distribution across cell types, which have larger cosine distances. RNASeq and spatial populations (each from one animal, “Nutmeg” and “Tank”) are highly correlated at the Class level (B) and at the Group level (C,D). The impact of Group-level accuracy on the whole atlas is shown in a 2D heatmap of predicted Group accuracy and number of spatial cells (E). For an example cutoff of estimated Group accuracy of 0.75, less than 12% of cells and 26% of Groups are impacted.



S3. TAC3 interneuron Groups are shared between striatum and hypothalamus, related to Fig 1. (A) Spatial mapping of TAC3+ GABAergic interneurons previously characterized in the primate striatum (Krienen 2020; Johansen et al 2025) indicates that one of the Groups (STR-BF TAC3-PLPP4-LHX8 GABA) is widely distributed across the striatum (Ca, Pu, NAC), septum (SEP), ventral pallidum (VeP) and hypothalamus (HTH). (B) Clusters within Group STR-BF TAC3-PLPP4-LHX8 GABA show further regional compartmentalization within these structures.

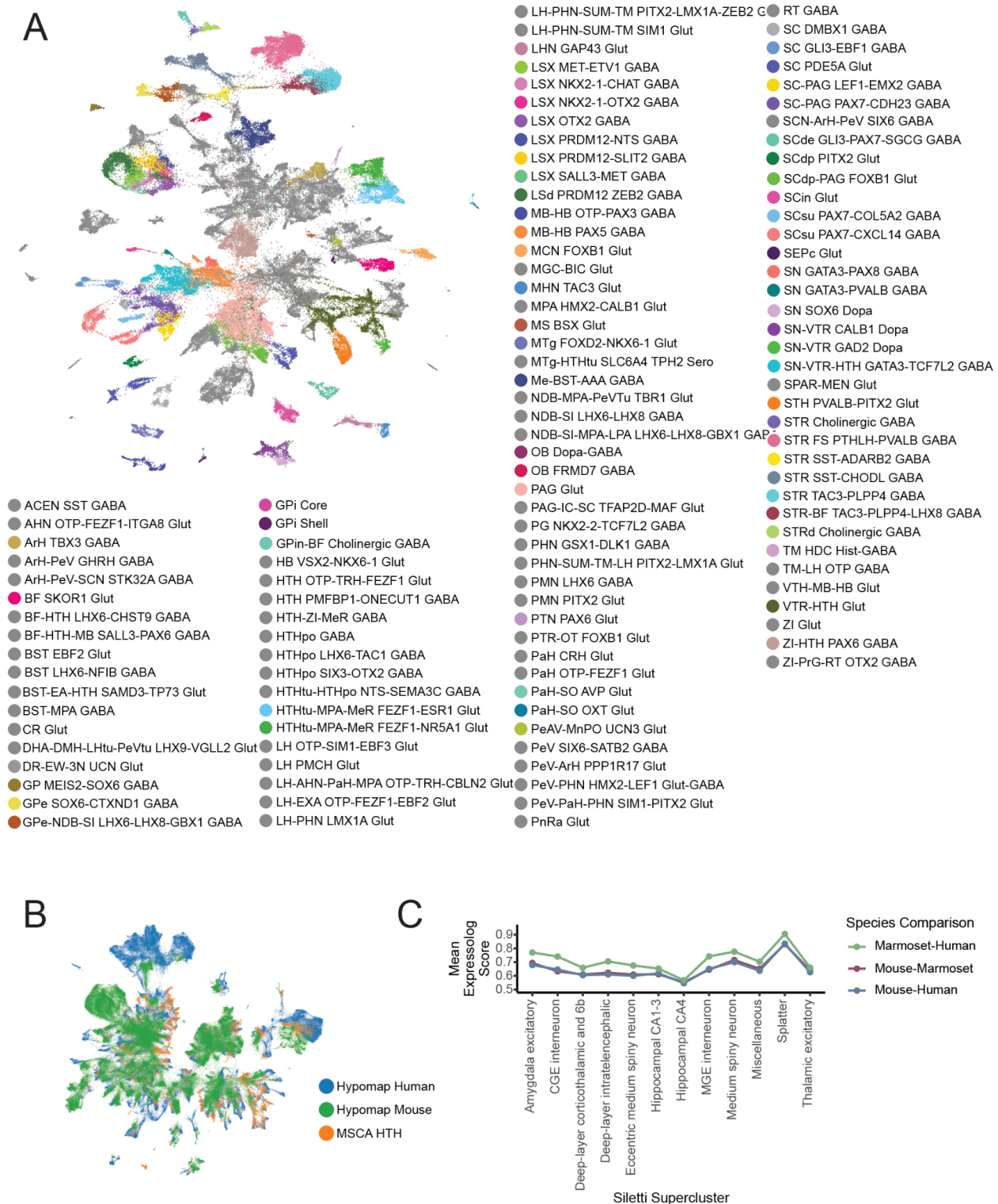


S4. Concordance between RNA-seq and ATAC-seq data, related to Fig 1. (A) UMAP embedding of all neurons based on gene expression colored by Class. (B) tSNE embedding of sub-sampled neurons based on chromatin accessibility-derived cisTopics colored by Class. (C) Heatmap of marker gene expression (top) and matched chromatin accessibility of peaks in the corresponding marker gene (bottom) of each group. (D) Density plot of RNA-ATAC discordance score calculated from pairwise cosine distances between z-scored Group centroids in RNA (scVI embedding) and ATAC (Topic embedding) latent spaces. (E) Neurons are grouped by anatomical region and neurotransmitter. For every possible pair, co-enrichment is based on the number overlaps in the top 1% of genes (RNA) or top 2% of ATAC peaks according to apegIshrunken log2FoldChange estimates (see Methods section on co-enrichment analysis).



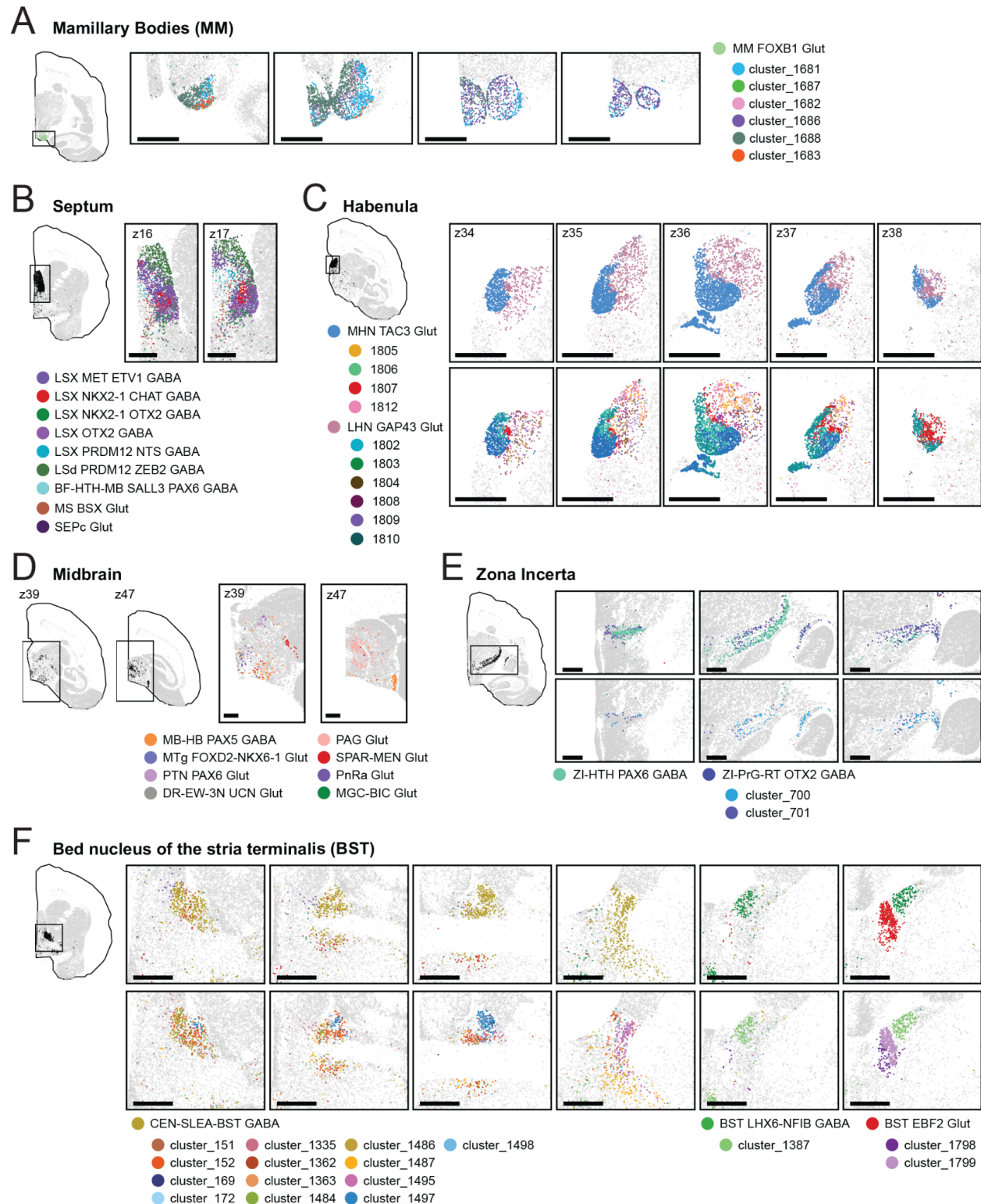
S5. Cytosplore Viewer Interface, related to Fig 1. (A) Bar chart showing ATAC peaks or RNA genes ranked by correlation to the selected seed feature. (B) Spatial map colored by the selected ATAC peak (chr3:128253794–128254295). (C) Spatial map colored by the selected RNA gene (*FIBCD1*). (D) Spatial map colored by the spatially measured gene (*EPHA5*). (E) Stacked bar chart showing cell-type proportions among the top 1% of cells for the selected ATAC peak. (F) Stacked bar chart showing cell-type proportions among the top 1% of cells for the selected RNA gene. (G) RNA-seq UMAP colored by the selected ATAC peak. (H) ATAC-seq UMAP colored by the selected RNA gene.

shows the spearman correlation of gene expression between the corresponding marmoset Groups and human Superclusters.

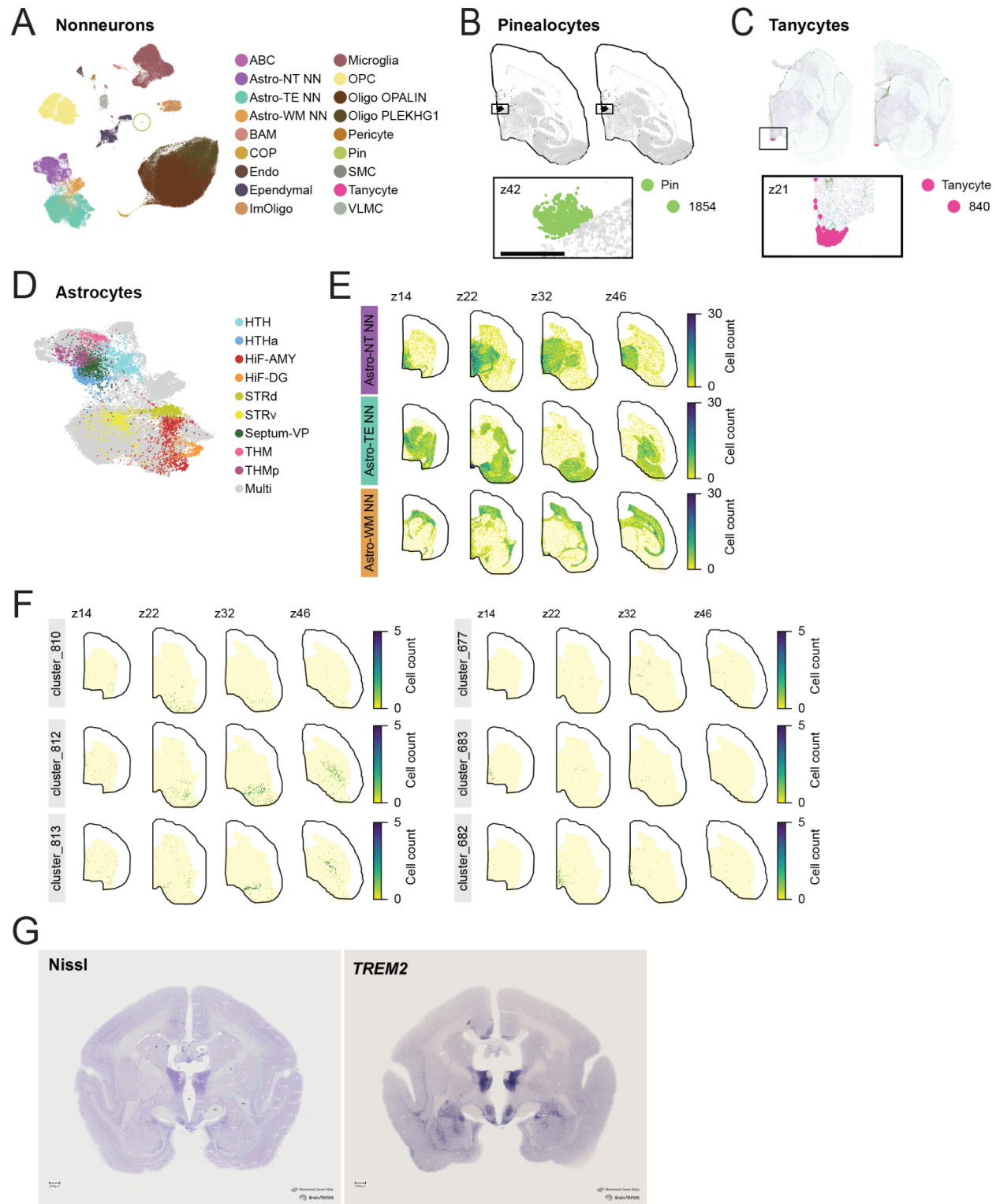


S7. Heterogenous and complex “Splatter” cells, related to Fig 2. (A) UMAP embedding of cells mapped to human Splatter Supercluster colored by MSCA Groups discussed in the manuscript. (B) Integration of MSCA hypothalamus Groups and mouse and human Hypomap datasets (C) Mean expressolog scores of homologous clusters across human, marmoset and

mouse within each Supercluster. Higher expressolog scores indicate stronger transcriptomic similarities between comparisons.

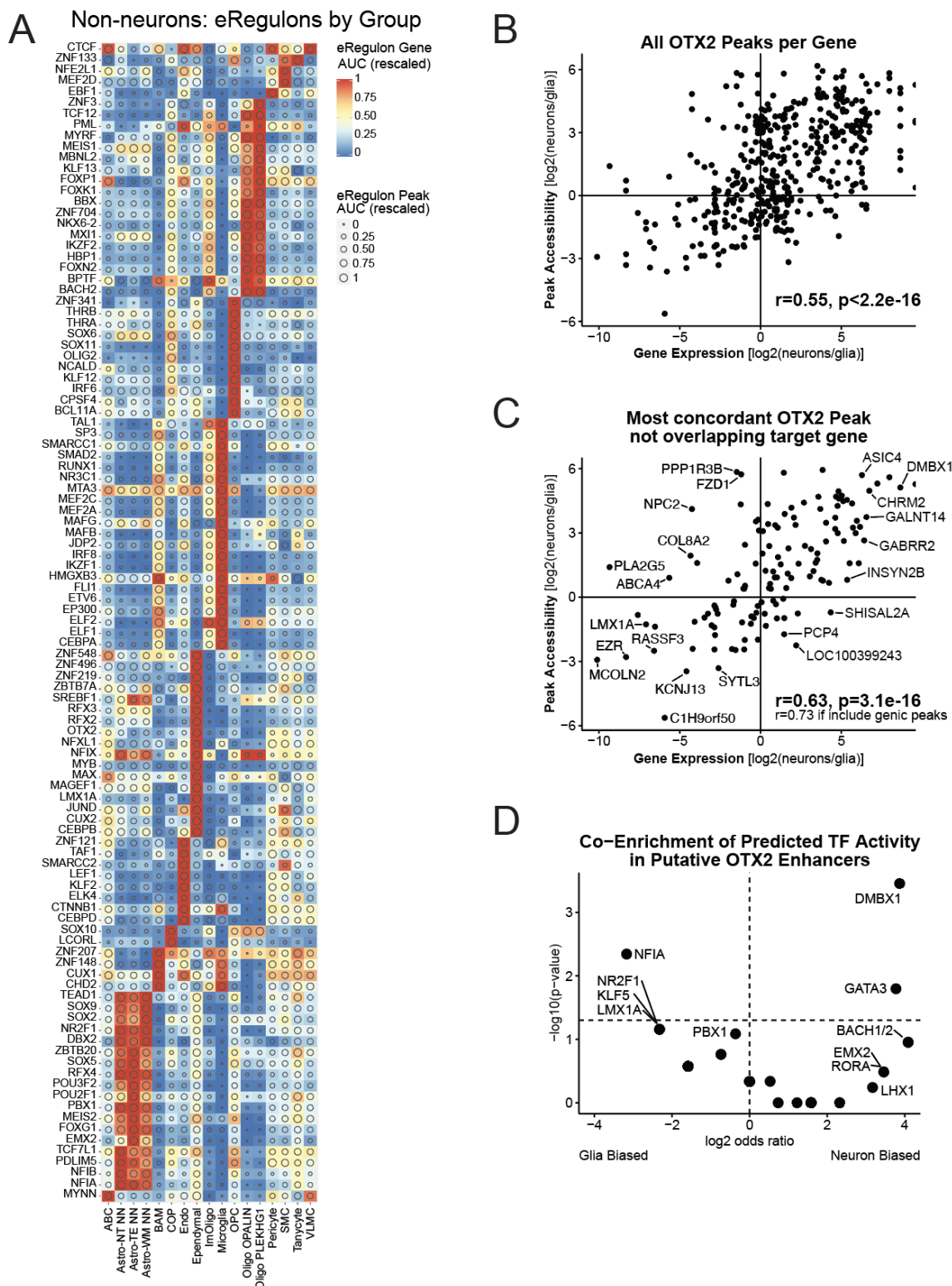


S8. Representative spatial sections of “Splatter” Groups in different brain regions, related to Fig 2. (A) Mamillary bodies, (B) Septum, (C) Habenula, (D) Midbrain, (E) Zona Incerta, (F) Bed nucleus of the stria terminalis



S9. Nonneurons, related to Fig 3. (A) UMAP embedding of nonneurons colored by Groups. (B) Spatial localization of the pinealocyte Group (Cluster_1854) to the pineal gland. (C) Spatial localization of tanycytes (Cluster_840) (D) UMAP embedding of astrocytes colored by spatially

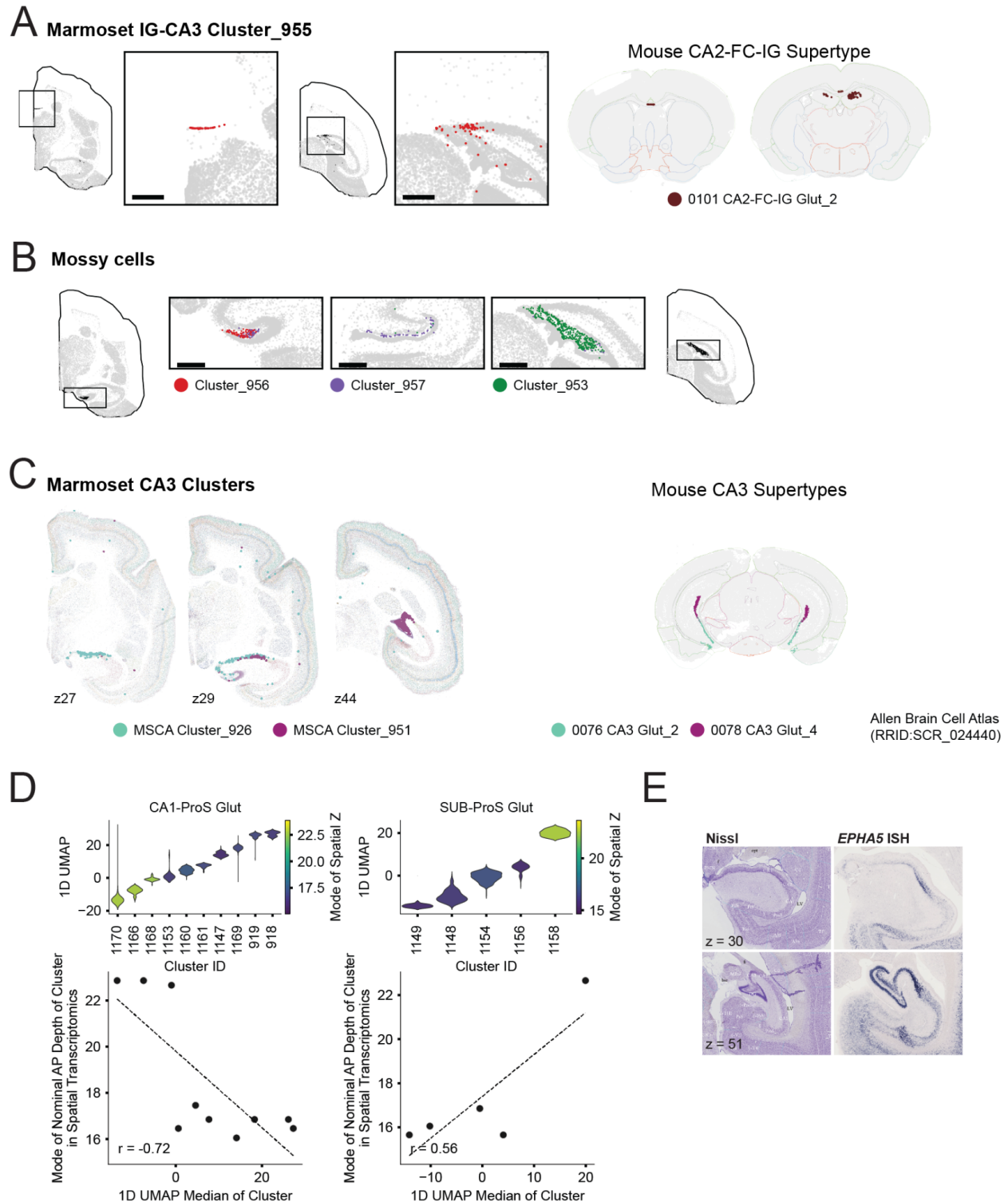
biased clusters and their regions. (E) 2D histogram of cell counts showing the regional zonation of the three astrocyte Groups. (F) 2D histogram of cell counts showing the specific regional zonation of six clusters. (G) Nissl and TREM2 ISH staining from the Marmoset Gene Atlas¹⁴ (1Y, Male)



S10. Non-neuronal Activator eRegulons and extended OTX2 analysis, related to Fig 4. (A) Heatmap showing all direct activator (+) eRegulons detected in non-neurons. Non-neurons grouped at the Group level. (B) Comparison of OTX2 target gene expression and the accessibility of peak with inferred OTX2 binding/activity. Of 436 peaks, 202 (46%) are within the

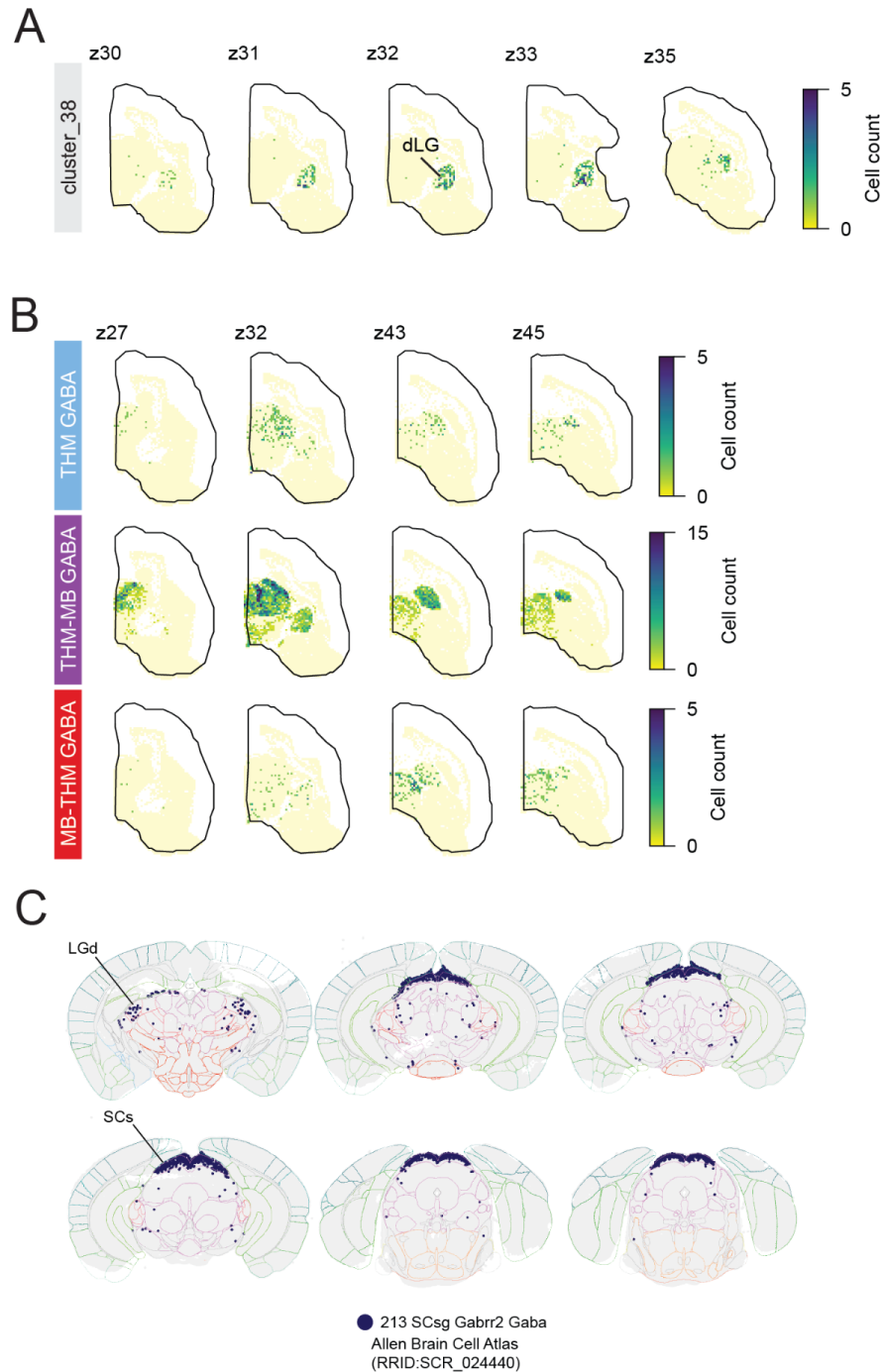
target gene itself (90 within 500bp of any TSS, and 57 peaks contain the target gene TSS). (C) Same as B, but peaks were subset to only those not found within the target gene. Of those, the peak with most similar enrichment to the target gene's expression was selected. (D) Peaks with predicted OTX2 activity sometimes have other TF motifs and predicted activity. Each point represents a candidate co-binding TF detected in the same OTX2-bound peaks linked to differentially regulated target genes. Odds ratio is for co-binding in neuron versus glia-biased OTX2 loci (log2FoldChange magnitude >0.75 for gene and peak), and p-values are unadjusted from Fisher's exact test (line for p=0.05).

averages all other neuronal clusters. Z-score differences are shown for individual eRegulons. (C) Cluster-level eRegulon z-scores are averaged within each grouping (columns). eRegulons are shown if $z > 1$ for at least one LGE and one TEL Glut grouping. Bold: eRegulons that are > 1 in panel B for both LGE and TEL Glut. (D) For selected eRegulons from panel C, the TF's expression and eRegulon (direct +/+) activity (gene AUC z-score) are shown (both are the mean for all clusters in each grouping). (E) All direct eRegulons within two hops of CRACDL are shown, with nodes colored by gene AUC z-scores and sized by gene expression z-scores.

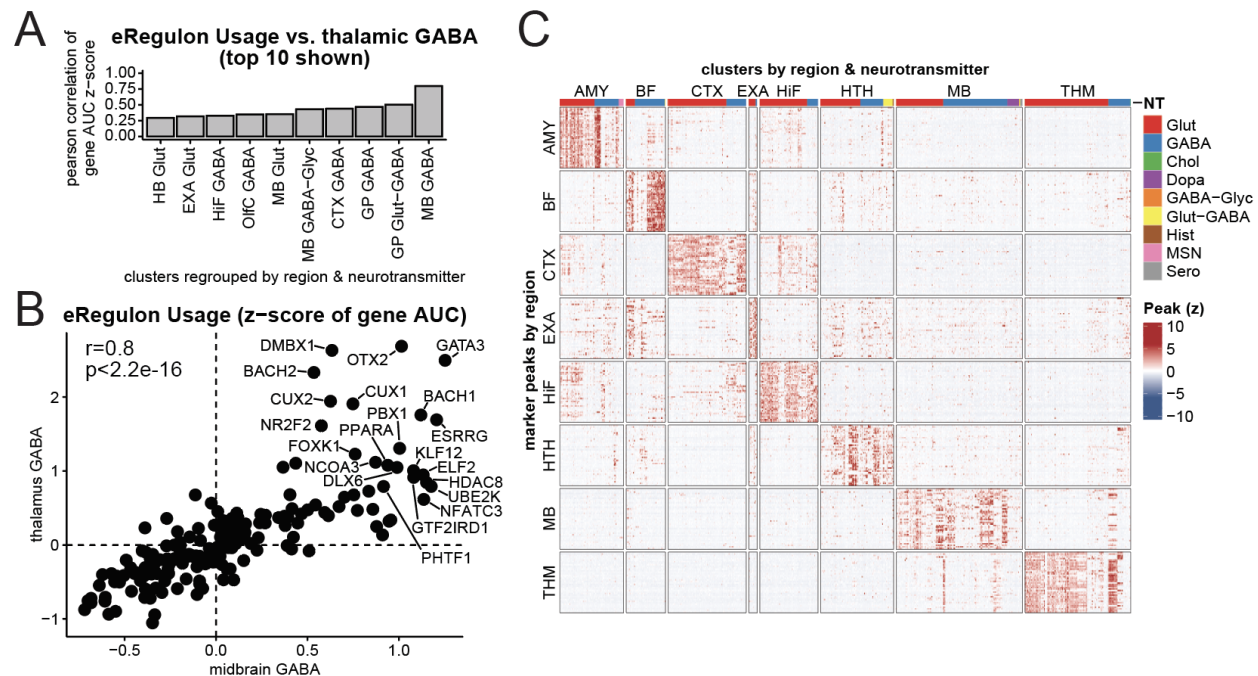


S12. Distribution of transcriptomic clusters and regulatory programs in the hippocampus, related to Fig 6. (A) Marmoset cluster 995 localizes to IG and posterior CA3 (left), and maps to mouse supertype 0101 CA2-FC-IG Glut_2 (right). (B) Spatial localization of mossy cells in DG. (C) Clusters with anterior or posterior bias in marmoset CA3 (left) map to mouse ventral and dorsal CA3 supertypes (right), respectively. (D) Top: Distribution of 1D

UMAP coordinates for cells in each cluster of CA1-ProS (left) and SUB-ProS (right) Group, with color indicating the modal spatial z-order of the cluster. Bottom: Scatter plots showing correlation of median 1D UMAP coordinates and modal spatial z-section for each cluster in CA1-ProS Glut (left; $r = -0.72$, $p < 0.05$) and SUB-ProS Glut (right; $r = 0.56$, $p < 0.05$). (F) EPHA5 ISH staining from the Marmoset Gene Atlas showing higher expression of EPHA5 in posterior CA3 in comparison to anterior CA3 (3M, Female).



S13. GABAergic Groups in the primate Thalamus, related to Fig 7. (A) 2D histograms of cell counts showing the regional localization of Cluster_38 to the dLG. (B) 2D histograms of cell counts for the three non-RT thalamic GABAergic Groups. (C) Mouse subclass 213 SCsg Gabrr2 Gaba from the Allen Brain Cell Atlas (RRID:SCR_024440).



S14. eRegulons and peaks common to thalamus and midbrain GABAergic neurons, related to Fig 7. (A) eRegulon usage (cluster gene AUC z-score) is averaged for each dominant region and neurotransmitter (so not Group or Class), and then the top 10 correlations with thalamic GABA neurons (from Groups THM GABA, THM-MB GABA, RT GABA, ZI-HTH PAX6 GABA (1 Cluster excluded), ZI-PrG-RT OTX2 GABA, PG NKX2-2-TCFL2 GABA) are shown. (B) Specific eRegulon usage (gene AUC z-scores) are shown for all midbrain- or thalamus-localized GABAergic neurons. (C) Top regional marker peaks identified in each anatomical region (row) and the peak accessibility z-score for each cluster (column).

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. 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>.
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. 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>.
6. Niepoth, N., Merritt, J.R., Uminski, M., Lei, E., Esquibies, V.S., Bando, I.B., Hernandez, K., Gebhardt, C., Wacker, S.A., Lutz, S., et al. (2024). Evolution of a novel adrenal cell type that promotes parental care. *Nature* 629, 1082–1090. <https://doi.org/10.1038/s41586-024-07423-y>.
7. 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>.
8. 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>.
9. 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>.
10. 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>.
11. 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>.
12. 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>.
13. 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. B Biol. Sci.* 360, 665–691. <https://doi.org/10.1098/rstb.2005.1626>.
 14. 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.* 118, e2020125118. <https://doi.org/10.1073/pnas.2020125118>.
 15. 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>.
 16. 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>.
 17. 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>.
 18. 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>.
 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 N. Y. N* 1991 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. 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>.
 23. 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>.
24. 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>.
25. Johansen, N.J., Fu, Y., Schmitz, M., Dubuc, A., Wirthlin, M., Garcia, A.D., Hewitt, M., Turner, M., Seeman, S., Long, B., et al. (2025). Cross-species consensus atlas of the primate basal ganglia.
26. 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>.
27. transcriptomic_clustering: improved clustering pipeline for transcriptomic data https://github.com/AllenInstitute/transcriptomic_clustering/tree/dev.
28. 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>.
29. 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>.
30. Reilly, M.B., Tekieli, T., Cros, C., Aguilar, G.R., Lao, J., Toker, I.A., Vidal, B., Leyva-Díaz, E., Bhattacharya, A., Cook, S.J., et al. (2022). Widespread employment of conserved *C. elegans* homeobox genes in neuronal identity specification. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1010372>.
31. 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.
32. 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>.
33. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>.
34. 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>.
35. 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> <https://doi.org/10.5281/zenodo.7857024>.

36. 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>.
37. 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>.
38. 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>.
39. 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>.
40. Franjic, D., Skarica, M., Ma, S., Arellano, J.I., Tebbenkamp, A.T.N., Choi, J., Xu, C., Li, Q., Morozov, Y.M., Andrijevic, D., et al. (2022). Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* 110, 452-469.e14. <https://doi.org/10.1016/j.neuron.2021.10.036>.
41. 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>.
42. Hashikawa, Y., Hashikawa, K., Rossi, M.A., Basiri, M.L., Liu, Y., Johnston, N.L., Ahmad, O.R., and Stuber, G.D. (2020). Transcriptional and Spatial Resolution of Cell Types in the Mammalian Habenula. *Neuron* 106, 743-758.e5. <https://doi.org/10.1016/j.neuron.2020.03.011>.
43. Wallace, M.L., Huang, K.W., Hochbaum, D., Hyun, M., Radeljic, G., and Sabatini, B.L. Anatomical and single-cell transcriptional profiling of the murine habenular complex. *eLife* 9, e51271. <https://doi.org/10.7554/eLife.51271>.
44. 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>.
45. 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>.
46. 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>.
47. Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* 513, 532–541. <https://doi.org/10.1002/cne.21974>.
48. Ueck, M., and Wake, K. (1977). The pinealocyte--a paraneuron? A review. *Arch. Histol. Jpn. Nihon Soshikigaku Kiroku* 40 Suppl, 261–278. https://doi.org/10.1679/aohc1950.40.supplement_261.
49. Blackshaw, S., and Snyder, S.H. (1997). Developmental Expression Pattern of Phototransduction Components in Mammalian Pineal Implies a Light-Sensing Function. *J. Neurosci.* 17, 8074–8082. <https://doi.org/10.1523/JNEUROSCI.17-21-08074.1997>.
50. Kannangara, H., Cullen, L., Miyashita, S., Korkmaz, F., Macdonald, A., Gumerova, A., Witzum, R., Moldavski, O., Sims, S., Burgess, J., et al. (2023). Emerging roles of brain tanycytes in regulating blood-hypothalamus barrier plasticity and energy homeostasis. *Ann. N. Y. Acad. Sci.* 1525, 61–69. <https://doi.org/10.1111/nyas.15009>.
51. Ibañez Rodriguez, M.P., Noctor, S.C., and Muñoz, E.M. (2016). Cellular Basis of Pineal Gland Development: Emerging Role of Microglia as Phenotype Regulator. *PLoS ONE* 11, e0167063. <https://doi.org/10.1371/journal.pone.0167063>.
52. 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*. <https://doi.org/10.1016/j.neuron.2025.09.011>.
53. 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>.
54. Khakh, B.S. (2025). On astrocyte-neuron interactions: Broad insights from the striatum. *Neuron* 0. <https://doi.org/10.1016/j.neuron.2025.08.009>.
55. Liwang, J.K., Kronman, F.N., Pi, H.-J., Wu, Y.-T., Vanselow, D.J., Manjila, S.B., Parmaksiz, D., Shin, D., Ben-Simon, Y., Taormina, M., et al. (2025). epDevAtlas: mapping GABAergic cells and microglia in the early postnatal mouse brain. *Nat. Commun.* 16, 9538. <https://doi.org/10.1038/s41467-025-64549-x>.
56. Barry-Carroll, L., Greulich, P., Marshall, A.R., Riecken, K., Fehse, B., Askew, K.E., Li, K., Garaschuk, O., Menassa, D.A., and Gomez-Nicola, D. (2023). Microglia colonize the developing brain by clonal expansion of highly proliferative progenitors, following allometric scaling. *Cell Rep.* 42. <https://doi.org/10.1016/j.celrep.2023.112425>.
57. Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, Scheiwe, C., Nessler, S., Kunz, P., van Loo, G., et al. (2019). Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* 566, 388–392.

<https://doi.org/10.1038/s41586-019-0924-x>.

58. Tan, Y.-L., Yuan, Y., and Tian, L. (2020). Microglial regional heterogeneity and its role in the brain. *Mol. Psychiatry* 25, 351–367. <https://doi.org/10.1038/s41380-019-0609-8>.
59. Martins-Ferreira, R., Calafell-Segura, J., Leal, B., Rodríguez-Ubreva, J., Martínez-Saez, E., Mereu, E., Pinho E Costa, P., Laguna, A., and Ballestar, E. (2025). The Human Microglia Atlas (HuMicA) unravels changes in disease-associated microglia subsets across neurodegenerative conditions. *Nat. Commun.* 16, 739. <https://doi.org/10.1038/s41467-025-56124-1>.
60. Penna, E., Cunningham, C.L., Saylor, S., Kreutz, A., Tarantal, A.F., Martínez-Cerdeño, V., and Noctor, S.C. (2021). Greater Number of Microglia in Telencephalic Proliferative Zones of Human and Nonhuman Primate Compared with Other Vertebrate Species. *Cereb. Cortex Commun.* 2, tgab053. <https://doi.org/10.1093/texcom/tgab053>.
61. Li, Y., Li, Z., Yang, M., Wang, F., Zhang, Y., Li, R., Li, Q., Gong, Y., Wang, B., Fan, B., et al. (2022). Decoding the temporal and regional specification of microglia in the developing human brain. *Cell Stem Cell* 29, 620–634.e6. <https://doi.org/10.1016/j.stem.2022.02.004>.
62. Paxinos, G. (2012). *The marmoset brain in stereotaxic coordinates* 1st ed. (Academic Press).
63. 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.* 118, e2020125118. <https://doi.org/10.1073/pnas.2020125118>.
64. 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>.
65. 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>.
66. 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>.
67. 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>.
68. 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>.
69. 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>.
70. 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>.
71. 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>.
72. 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>.
73. Bandler, R.C., Vitali, I., Delgado, R.N., Ho, M.C., Dvoretzskova, 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>.
74. Nóbrega-Pereira, S., Kessar, 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>.
75. 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>.
76. Zhao, Y., Marín, O., Hermes, 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. U. S. A.* 100, 9005–9010. <https://doi.org/10.1073/pnas.1537759100>.
77. 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. Off. J. Soc. Neurosci.* 27, 3078–3089. <https://doi.org/10.1523/JNEUROSCI.3055-06.2007>.
78. 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. Off. J. Soc. Neurosci.* 32, 17690–17705. <https://doi.org/10.1523/JNEUROSCI.3583-12.2012>.
79. 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>.
80. 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>.
81. 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>.
82. Andrews, W., Barber, M., Hernandez-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>.
83. Bright, A.R., Kotlyarenko, Y., Neuhaus, F., Rodrigues, D., Feng, C., Peters, C., Vitali, I., Dönmez, E., Myoga, M.H., Dvoretzskova, 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>.
84. Puellas, L., Kuwana, E., Puellas, 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%253C409::aid-cne3%253E3.0.co;2-7](https://doi.org/10.1002/1096-9861(20000828)424:3%253C409::aid-cne3%253E3.0.co;2-7).
85. 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>.
86. 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>.
87. 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. Cold Spring Harb. N* 1, 140–152.
88. 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>.
89. 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>.
90. 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>.

91. 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>.
92. 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>.
93. 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. (2025). Single-cell multiomic human brain atlas reveals regulatory drivers of cortical regionality. Preprint at Neuroscience, <https://doi.org/10.1101/2025.08.07.667992> <https://doi.org/10.1101/2025.08.07.667992>.
94. 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/10.1016/j.neuron.2012.05.029>.
95. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science* 347, 1260419. <https://doi.org/10.1126/science.1260419>.
96. 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>.
97. Cohen, L., Halpern, A., Blosser, T.R., Che, P., Pan, X., and Zhuang, X. (2025). Whole-transcriptome-scale and isoform-resolved spatial imaging of single cells in complex tissues. Preprint at bioRxiv, <https://doi.org/10.1101/2025.08.27.672533> <https://doi.org/10.1101/2025.08.27.672533>.
98. 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>.
99. 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>.
100. Liu, X., Zeira, R., and Raphael, B.J. (2023). Partial alignment of multislice spatially resolved transcriptomics data. *Genome Res.*, genome;gr.277670.123v1. <https://doi.org/10.1101/gr.277670.123>.
101. 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>.
102. 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>.
103. 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>.
104. 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>.
105. 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>.
106. 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>.
107. Goldman-Rakic, P.S., and Porrino, L.J. (1985). The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *J. Comp. Neurol.* 242, 535–560. <https://doi.org/10.1002/cne.902420406>.
108. Alexander, G.E., Crutcher, M.D., and DeLong, M.R. (1991). Chapter 6 Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, “prefrontal” and “limbic” functions. In *Progress in Brain Research The Prefrontal Its Structure, Function and Cortex Pathology.*, H. B. M. Uylings, C. G. Van Eden, J. P. C. De Bruin, M. A. Corner, and M. G. P. Feenstra, eds. (Elsevier), pp. 119–146. [https://doi.org/10.1016/S0079-6123\(08\)62678-3](https://doi.org/10.1016/S0079-6123(08)62678-3).
109. Haber, S.N., and Calzavara, R. (2009). The cortico-basal ganglia integrative network: The role of the thalamus. *Brain Res. Bull.* 78, 69–74. <https://doi.org/10.1016/j.brainresbull.2008.09.013>.
110. Murray, K.D., Rubin, C.M., Jones, E.G., and Chalupa, L.M. (2008). Molecular Correlates of Laminar Differences in the Macaque Dorsal Lateral Geniculate Nucleus. *J. Neurosci.* 28, 12010–12022. <https://doi.org/10.1523/JNEUROSCI.3800-08.2008>.
111. Cruz-Martín, A., El-Danaf, R.N., Osakada, F., Sriram, B., Dhande, O.S., Nguyen, P.L., Callaway, E.M., Ghosh, A., and Huberman, A.D. (2014). A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature* 507, 358–361. <https://doi.org/10.1038/nature12989>.
112. Cortes, N., Ladret, H.J., Abbas-Farishta, R., and Casanova, C. (2024). The pulvinar as a hub of visual processing and cortical integration. *Trends Neurosci.* 47, 120–134. <https://doi.org/10.1016/j.tins.2023.11.008>.
113. Huo, B., Zeater, N., Lin, M.K., Takahashi, Y.S., Hanada, M., Nagashima, J., Lee, B.C., Hata, J., Zaheer, A., Grünert, U., et al. (2019). Relation of koniocellular layers of dorsal lateral geniculate to inferior pulvinar nuclei in common marmosets. *Eur. J. Neurosci.* 50, 4004–4017.

<https://doi.org/10.1111/ejn.14529>.

114. Grieve, K.L., Acuña, C., and Cudeiro, J. (2000). The primate pulvinar nuclei: vision and action. *Trends Neurosci.* 23, 35–39. [https://doi.org/10.1016/S0166-2236\(99\)01482-4](https://doi.org/10.1016/S0166-2236(99)01482-4).
115. Gutierrez, C., Cola, M.G., Seltzer, B., and Cusick, C. (2000). Neurochemical and connectional organization of the dorsal pulvinar complex in monkeys. *J. Comp. Neurol.* 419, 61–86. [https://doi.org/10.1002/\(SICI\)1096-9861\(20000327\)419:1%253C61::AID-CNE4%253E3.0.CO;2-I](https://doi.org/10.1002/(SICI)1096-9861(20000327)419:1%253C61::AID-CNE4%253E3.0.CO;2-I).
116. Homman-Ludiye, J., and Bourne, J.A. (2019). The medial pulvinar: function, origin and association with neurodevelopmental disorders. *J. Anat.* 235, 507–520. <https://doi.org/10.1111/joa.12932>.
117. Jager, P., Ye, Z., Yu, X., Zagoraïou, 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>.
118. 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](https://doi.org/10.1016/S1472-9288(02)00002-X).
119. 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>.
120. 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>.
121. 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>.
122. 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>.
123. Blankvoort, S., Witter, M.P., Noonan, J., Cotney, J., and Kentros, C. (2018). Marked Diversity of Unique Cortical Enhancers Enables Neuron-Specific Tools by Enhancer-Driven Gene Expression. *Curr. Biol.* 28, 2103-2114.e5. <https://doi.org/10.1016/j.cub.2018.05.015>.
124. 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>.
125. 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>.
126. Caplan, K.A., Chen, K., Wienisch, M., Wang, Y., Chai, R., Kang, I.-H., Chen, C.S., Rosario, R. del, Krienen, F., 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> <https://doi.org/10.1101/2025.11.06.687071>.
127. 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>.
128. 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](https://doi.org/10.1016/S0166-2236(03)00234-0).
129. 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>.
130. 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>.
131. 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>.
132. 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> <https://doi.org/10.1101/2025.11.16.688724>.
133. Miller, J. (2025). mfishtools: Building Gene Sets and Mapping mFISH Data.
134. 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>.
135. 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>.
136. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDbtFinder. *F1000Research* 10, 979. <https://doi.org/10.12688/f1000research.73600.2>.
137. 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>.
138. 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>.
139. 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>.
140. Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinforma.* 69, e96. <https://doi.org/10.1002/cpbi.96>.
141. Campitelli, E. (2019). ggnewscale: Multiple Fill and Color Scales in ggplot2. (Zenodo). <https://doi.org/10.5281/zenodo.2543772> <https://doi.org/10.5281/zenodo.2543772>.
142. 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>.
143. 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>.
144. 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>.
145. HOMBA ontology — Common Coordinate Framework and Multi-Species Atlas Primer (CCF-MAP) https://alleninstitute.github.io/CCF-MAP/docs/HOMBA_ontology_v1.html.
146. 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>.
147. Winter, S.D., and cbravo (2025). aertslab/pycistarget: pycistarget v1.1. (Zenodo). <https://doi.org/10.5281/zenodo.14627914> <https://doi.org/10.5281/zenodo.14627914>.