



## NEUROSCIENCE

# A marmoset brain cell census reveals regional specialization of cellular identities

Fenna M. Krienen<sup>1,2,†\*</sup>, Kirsten M. Levandowski<sup>2,3,4</sup>, Heather Zaniewski<sup>2,3,4</sup>, Ricardo C.H. del Rosario<sup>1,2</sup>, Margaret E. Schroeder<sup>2,3,4</sup>, Melissa Goldman<sup>1,2</sup>, Martin Wienisch<sup>2,3,4</sup>, Alyssa Lutservitz<sup>1</sup>, Victoria F. Beja-Glasser<sup>2,3,4</sup>, Cindy Chen<sup>2,3,4</sup>, Qiangge Zhang<sup>2,3,4</sup>, Ken Y. Chan<sup>2</sup>, Katelyn X. Li<sup>3,4</sup>, Jitendra Sharma<sup>3,4</sup>, Dana McCormack<sup>2,3,4</sup>, Tay Won Shin<sup>3,4,5</sup>, Andrew Harrahill<sup>3,4</sup>, Eric Nyase<sup>3,4</sup>, Gagandeep Mudhar<sup>6</sup>, Abigail Mauermann<sup>3,4,5</sup>, Alec Wysoker<sup>2</sup>, James Nemesh<sup>2</sup>, Seva Kashin<sup>2</sup>, Josselyn Vergara<sup>2</sup>, Gabriele Chelini<sup>7</sup>, Jordane Dimidschstein<sup>2</sup>, Sabina Berretta<sup>8,9</sup>, Benjamin E. Deverman<sup>2</sup>, Ed Boyden<sup>3,4,5</sup>, Steven A. McCarroll<sup>1,2,†\*</sup>, Guoping Feng<sup>2,3,†\*</sup>

Copyright © 2023  
The Authors, some rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
License 4.0 (CC BY).

The mammalian brain is composed of many brain structures, each with its own ontogenetic and developmental history. We used single-nucleus RNA sequencing to sample over 2.4 million brain cells across 18 locations in the common marmoset, a New World monkey primed for genetic engineering, and examined gene expression patterns of cell types within and across brain structures. The adult transcriptomic identity of most neuronal types is shaped more by developmental origin than by neurotransmitter signaling repertoire. Quantitative mapping of GABAergic types with single-molecule FISH (smFISH) reveals that interneurons in the striatum and neocortex follow distinct spatial principles, and that lateral prefrontal and other higher-order cortical association areas are distinguished by high proportions of *VIP*<sup>+</sup> neurons. We use cell type-specific enhancers to drive AAV-GFP and reconstruct the morphologies of molecularly resolved interneuron types in neocortex and striatum. Our analyses highlight how lineage, local context, and functional class contribute to the transcriptional identity and bio-distribution of primate brain cell types.

## INTRODUCTION

The vertebrate brain is a complex tissue composed of highly heterogeneous and spatially regionalized cell types (1–8). The molecular composition of brain cells is influenced by many factors, including the cells' function, developmental and phylogenetic history, and regional context; each factor contributes in some part to overall cellular identity. The marmoset, a small, New World monkey, diverged from humans ~45 million years ago and has a near-lissencephalic brain that is the size of a squirrel. Despite its small size, the marmoset brain retains many of the structural and functional innovations that have accumulated along the primate lineage (9–14). Because of its small size, the marmoset brain also offers practical advantages for studying primate cell type diversity: Sequencing 1 million neurons would achieve coverage of 0.001% of the total neurons in the marmoset brain, yet would represent far lower coverage (0.00001%) of the neuronal composition of the human brain (15). Previous single-cell sequencing studies of the marmoset brain focused on single brain regions (16, 17) or on specific cell classes

across regions (10, 18). Inclusion of both closely and distantly related brain structures and cell types can yield insights into the sometimes surprising developmental and ontological relationships between them (19).

To better appreciate what shapes a primate brain cell's identity, we conducted a census of brain cell types of the adult marmoset using single-nucleus RNA sequencing (snRNA-seq). We generated snRNA-seq (10x Genomics 3' v3.1) data from 2.4 million unsorted brain cell nuclei across 8 neocortical and 10 subcortical locations from 10 young adult marmosets (4 male, 6 female) and resolved clusters from all major neuronal and nonneuronal cell classes. snRNA-seq data were generated as part of the Brain Initiative Cell Census Network (BICCN; RRID:SCR\_015820) and are available on the BICCN Data Center (RRID:SCR\_022815; <https://biccn.org/data>) as well as the NeMO archive (RRID:SCR\_016152; [https://scicrunch.org/resolver/RRID:SCR\\_016152](https://scicrunch.org/resolver/RRID:SCR_016152); <https://doi.org/10.1101/2022.10.18.512442>). Processed data can be accessed and explored through the cellxgene data repository (RRID:SCR\_021059).

All neuron-containing brain structures in the central nervous system (CNS) have both excitatory and inhibitory neuronal populations, although the proportions and degree of developmental relatedness between these two populations vary by structure. In the neocortex and other telencephalic structures of vertebrates, distinct populations of adult neurons are often categorized by their primary neurotransmitter status, typically either inhibitory [ $\gamma$ -Aminobutyric acid (GABA)-ergic] or excitatory (glutamatergic). In other brain structures, primary neurotransmitter usage appears less essential to a cell type's identity. We found that the transcriptional identities of excitatory and inhibitory neurons within telencephalic brain structures segregate strongly, consistent with previous studies in other mammalian species (4, 5, 17, 20). In contrast, there is much

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. <sup>3</sup>McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. <sup>4</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>5</sup>Howard Hughes Medical Institute, Cambridge, MA 02139, USA. <sup>6</sup>Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA. <sup>7</sup>Center for Mind/Brain Sciences, University of Trento, Piazza della Manifattura n.1, Rovereto (TN) 38068, Italy. <sup>8</sup>Basic Neuroscience Division, McLean Hospital, Belmont, MA 02478, USA. <sup>9</sup>Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA.

\*Corresponding author. Email: [fkrienen@princeton.edu](mailto:fkrienen@princeton.edu) (F.M.K.); [mccarroll@genetics.med.harvard.edu](mailto:mccarroll@genetics.med.harvard.edu) (S.A.M.); [fengg@mit.edu](mailto:fengg@mit.edu) (G.F.)

†These authors contributed equally to this work.

‡Present address: Princeton Neuroscience Institute, Princeton, NJ, USA.

greater transcriptional similarity between GABAergic and glutamatergic neuronal types in nontelencephalic compartments. Moreover, few gene expression distinctions present in telencephalic glutamatergic neurons are shared with glutamatergic neurons in nontelencephalic brain regions.

While primary neurotransmitter status did not drive transcriptional similarity between neurons, their brain structure of origin did: The adult transcriptomic identity of most neuronal types appears to be shaped more by their developmental or regional identity than by their neurotransmitter signaling repertoire. In line with previous findings in diverse mammalian species (2, 5, 18), we also found regional heterogeneity in nonneuronal cells, particularly astrocytes.

The neocortex is parcellated into a mosaic of functional areas. While our snRNA-seq results and others (2, 10, 21) indicate that cell type proportions vary considerably between discrete cortical areas, spatially resolved approaches are better suited for appreciating large-scale trends and finer discontinuities in the biodistribution of cell types (22). Although the marmoset brain is small, potentially owing to secondary dwarfism of Callitrichids (23), its neocortex is still highly functionally parcellated and retains primate-specific specializations (9–14). Moreover, the small size of the marmoset neocortex enables quantitative, spatial mapping of cell types across the entire anterior-posterior extent within single sections. We therefore used single-molecule fluorescence in situ hybridization (FISH) (smFISH) to spatially profile major interneuron types across marmoset striatum and neocortex. We found that in the striatum, major interneuron types are distributed as medial-lateral gradients. In contrast, the marmoset neocortex has a complex topography of interneuron concentrations that do not follow a single spatial axis. Lateral prefrontal areas, in particular, which have undergone major expansion in primate evolution (24), are typified by higher proportions of *VIP*<sup>+</sup> and *LAMP5/ID2*<sup>+</sup> interneurons and lower proportions of *SST*<sup>+</sup> interneurons relative to all other neocortical areas.

As cellular morphology is critical to neuron function and type definition (25–28), we sought to provide a compendium of morphologically reconstructed interneurons in striatum and neocortex. We reconstructed a set of molecularly characterized interneurons in marmosets that had received systemic delivery of an adeno-associated virus (AAV) carrying a reporter (AAV9-hDlx5/6-GFP-fGFP) under an interneuron-selective regulatory element (29). Reconstructions are available for download from the Brain Image Library (BIL; <https://doi.org/10.35077/g.609>). As AAV9-hDlx5/6-GFP-fGFP did not label a primate-specific, *TAC3*<sup>+</sup> striatal interneuron type that we previously discovered using snRNA-seq (10), we acquired single-nucleus ATAC-seq (snATAC-seq) data from marmoset striatum and used *TAC3* interneuron-specific enhancers to successfully express green fluorescent protein (GFP) in these cells, enabling their morphological reconstruction.

Comprehensive cell type atlases of mouse (1, 4, 5, 30) and regional sampling across humans and nonhuman primates brains (2, 18, 21, 31) have begun to facilitate comparative analysis of brain cell type features (10, 17, 32, 33). Together, our census of major transcriptomically defined brain cell types and quantitative mapping of interneuron biodistribution and morphology in the marmoset serve as a resource for the primate neuroscience community and for studies of cell type development and evolution.

## RESULTS

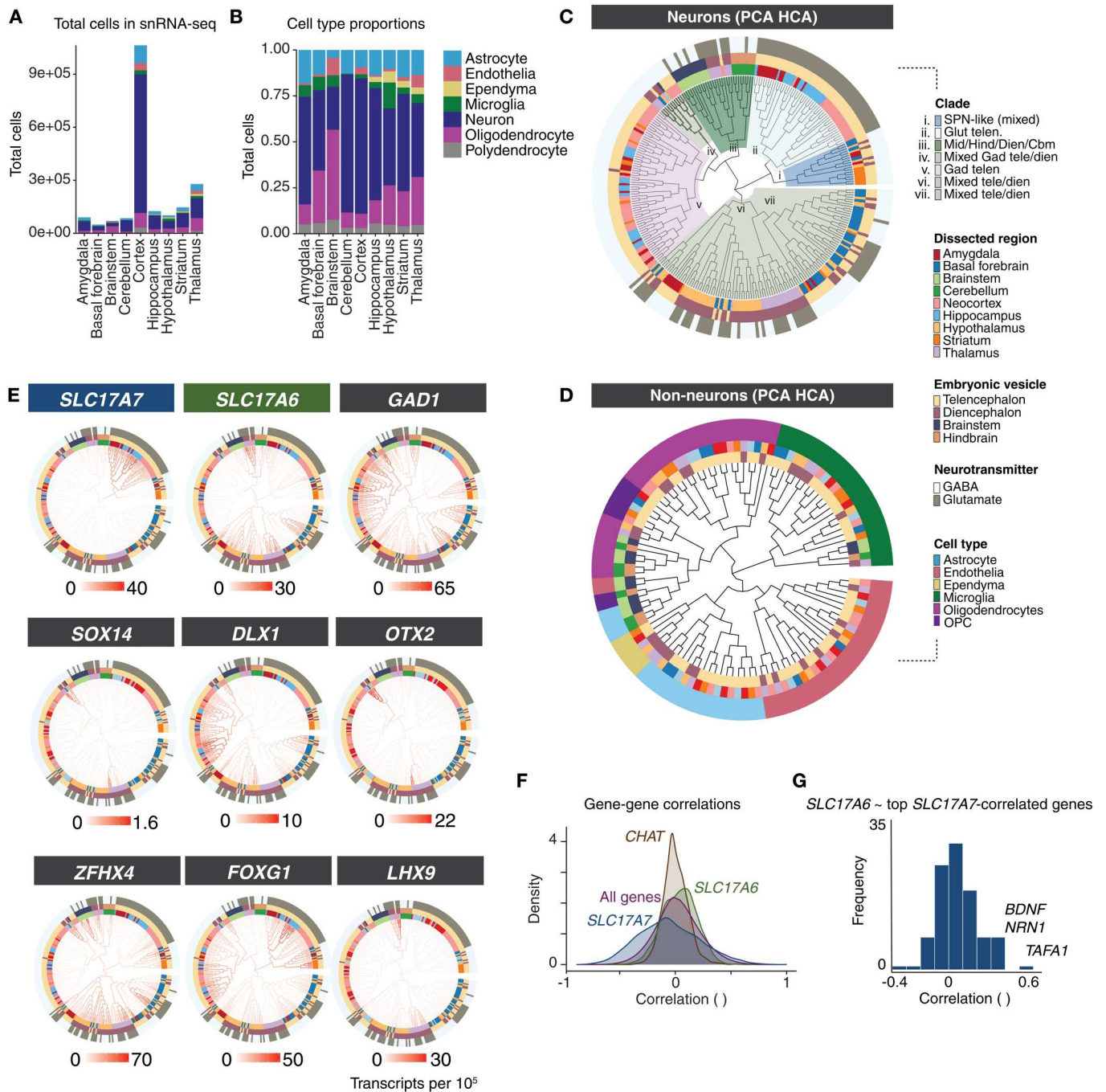
### A transcriptomic census identifies major marmoset brain cell types

We acquired snRNA-seq data from 18 brain regions collected using 10x Genomics 3' 3.1 chemistry across 10 young adult marmosets (ages 2 to 5 years) as well as a small dataset from the lateral prefrontal cortex (PFC) of two aged animals (Fig. 1, fig. S1, and table S1). The number of donors per brain structure varied (min = 2, max = 10; fig. S1), as did the cell sampling rate per brain structure (Fig. 1A, fig. S1A, and table S2). Neocortex was the most comprehensively sampled in terms of total numbers, donors, and regional dissections (Fig. 1A, fig. S1A, and table S2). We acquired data from cerebellum, brainstem, hypothalamus, thalamus, amygdala, striatum (separate dissections for caudate, putamen, and nucleus accumbens), hippocampus, basal forebrain, and neocortex. We did not attempt to dissect each brain structure in its entirety; this census should be viewed as containing representation from each of these structures, but not exhaustive profiling thereof. For example, the hypothalamus and brainstem contain tremendous cellular heterogeneity that varies across relatively small spatial scales due to the presence of specialized brain nuclei and rare cell types (1, 3, 34); our census only sampled small portions of these structures and is not comprehensive. Within the neocortex, we separately sampled eight neocortical locations (prefrontal, temporal pole, S1, M1, A1, V2, V1, and lateral parietal association), and within PFC, from four distinct subdivisions (dorsolateral, medial, orbital, and ventral) (fig. S1, B and C).

Nuclei from each brain structure were pooled across donors and analyzed to identify major cell types and their proportions by brain structure (Fig. 1, B to D). Using linear discriminant analysis [scPred; (35)] trained on a supervised set of cell class labels, we identified and discarded low-quality cells and doublets, and assigned each nucleus to its probable major type—astrocyte, endothelia, ependyma, microglia/macrophage, neuron, oligodendrocyte, or oligodendrocyte precursor cell (OPC). After major cell type assignment, nuclei were clustered by brain structure to reveal subtype diversity within each major class. We used a previously described clustering pipeline (4) based on independent components analysis (fastICA); each clustering analysis involved additional curation of doublets and outlier cells, followed by a second round of subclustering of major clusters. At each of these curation stages, independent components (ICs) that loaded on individual donors or batches were excluded and reclustering was performed to attenuate donor and batch effects on clustering results. For each cluster, a “metacell” (conceptually similar to “pseudo-bulking”) was generated by summing transcript counts of individual cells of each cluster together, followed by scaling and normalizing to 100,000 transcripts per nucleus. These regional and cell type-resolved metacells were the starting point for most cross-region analyses. The final snRNA-seq dataset size after curation contained 2.01 million cells (table S2).

### Hierarchical clustering reveals regional specialization of marmoset brain cells

Cell types are identified based on many factors including their function, morphology, developmental origin, and regional context (36–38). We studied how transcriptional profiles related to each other across 288 neuronal clusters (metacells) that were sampled across all brain structures. Individual replicates and other variables (age, sex) were generally proportionately represented across resolved



Downloaded from <https://www.science.org> at Massachusetts Institute of Technology on November 28, 2023

**Fig. 1. Single-nucleus RNA sequencing of marmoset brain.** (A) Number of nuclei per brain structure. Colors indicate cell classes. (B) Proportions of cell classes across brain structures. (C) Neurons in each dissected region are clustered separately, and then pseudo-bulked “metacells” of all 288 clusters are arranged in the dendrogram using hierarchical clustering of top 100 PCA scores of expressed genes. Outer rings colored by: dissected subregion (inner ring), cephalic vesicle (second ring), and major neurotransmitter usage (outer ring). Seven major “clades” are colored and labeled according to cell type and regional composition of clade. (D) Same hierarchical clustering procedure as (C) but for nonneuronal cell types. Outer ring colors indicate major nonneuronal cell class. Inner ring colors indicate region dissection and vesicle; colors as in (C). (E) Expression of markers for glutamatergic neurons (*SLC17A7*, *SLC17A6*), GABAergic neurons (*GAD1*), and TFs (*FOXG1*, *ZFH4*, *SOX14*, *DLX1*, *OTX2*, *LHX9*) are plotted as heatmaps on dendrogram shown in (C). These TFs are largely restricted to specific clades or are associated with particular cephalic vesicles. (F) Gene-gene correlation (Spearman tau) distributions across all neuron clusters in (C) for *SLC17A7*, *SLC17A6*, and *CHAT* (marker for cholinergic neurons) as well as background (all genes). The distribution of pairwise correlations to *CHAT* had a lower SD (mean  $r = 0.002$ , SD = 0.116) relative to baseline gene-gene correlations (mean  $r = 0.02$ , SD = 0.199;  $F_{5078,17315809} = 0.33$ ,  $P_{\text{adj}} < 1 \times 10^{-15}$ ). In contrast, pairwise correlations to *SLC17A7* were much broader relative to the background distribution of all gene-gene correlations. (G) Distribution of cross-cell type correlation to *SLC17A6* of genes most correlated with *SLC17A7* (top 116 genes with Spearman tau > 0.5 to *SLC17A7*).

clusters, with several exceptions that were likely mainly to differences in dissection across donors (figs. S1 and S2). We used hierarchical clustering to situate the neuronal types on a dendrogram computed using distances calculated from the top 100 principal component (PC) scores across neuron types (HCA-PCA; Fig. 1C and table S2). Most telencephalic types (neocortex, amygdala, hippocampus, striatum) clustered distinctly from diencephalic and hindbrain types, indicating that developmental lineage continues to shape the transcriptional identity of adult primate brain cell types. However, of the seven major clades, four contained mixtures of cell types at the terminal (leaf level) originating from distinct cephalic compartments. For example, basal forebrain neuron types intermingled with hypothalamic types, suggesting closer transcriptional similarity of two structures that are typically considered to occupy distinct cephalic compartments.

We next conducted HCA-PCA clustering on 111 nonneuronal metacells curated from all sampled brain structures (Fig. 1D). Major clades of this dendrogram reflected nonneuronal class; hematopoietic lineages (microglia and brain-resident macrophages) were the most transcriptionally divergent from the other nonneuronal types. Ependymal cells, detected only in structures that border the ventricles (particularly numerous in thalamus, striatum, and hypothalamus dissections), were transcriptionally most similar to astrocytes (Fig. 1D). As prior transcriptomic studies in mouse (1, 5), marmoset (18), and human brains (2) have also shown, astrocytes were particularly heterogeneous across major subdivisions (fig. S2, A to H). We validated the differential expression of two genes that distinguished astrocytes in the cortex (telencephalic origin, *FOXG1*) and thalamus (diencephalic origin, *SPARC*) in one marmoset using smFISH (fig. S2, F to H).

### Neurotransmitter usage is not strongly associated with transcriptomic identity

The neurotransmitters used by neurons are central to their function, and neurons are often classified by their primary neurotransmitter. However, synthesis of the major neurotransmitters—glutamate (in mammals, excitatory) and GABA or glycine (inhibitory)—requires only a modest number of genes, and GABA itself is synthesized from glutamate (39). As neuronal transcriptomes tend to show strong influences of their regional or developmental origin (Fig. 1C), the extent to which neurotransmitter utilization is strongly associated to the transcriptomic identity of neuron types may differ by brain structure. Transcriptomically defined neuron types in the neocortex and other telencephalic structures hierarchically partition into GABAergic and glutamatergic types (5, 8, 17), which reflects both their distinct developmental origins and their distinct neurotransmitter repertoire. In other brain structures such as the hypothalamus, the distinction between neurons using GABA or glutamate is much less clear at the transcriptional level (1, 3), likely as a consequence of shared developmental lineage.

To determine how neurotransmitter identity was associated with a general transcriptional identity across all neurons expressing the same neurotransmitter, we examined expression of genes encoding the most prevalent vesicular glutamate transporters (*SLC17A6*, *SLC17A7*) and GABA synthesis enzymes (*GAD1*, *GAD2*) (Fig. 1, C to F). If primary neurotransmitter usage was strongly associated with the expression of many other genes, we would expect that neurons expressing the same neurotransmitter-associated genes (GABAergic or glutamatergic) would preferentially group together

regardless of other factors (such as developmental origin). However, we did not find evidence for strong global transcriptomic identities of GABAergic and glutamatergic neurons. Neuronal types from each set were distributed across the tree, suggesting widely divergent transcriptomic identities of cell types that share a common neurotransmitter (Fig. 1C).

In telencephalic structures such as neocortex, hippocampus, and amygdala, glutamatergic neurons express *SLC17A7* (VGLUT1) and segregate from GABAergic telencephalic neurons (Fig. 1C). Yet, even within the telencephalon, neurotransmitter identity did not drive global transcriptional similarity between major clades. GABAergic projection neurons, such as spiny projection neurons (SPNs) of the striatum, were transcriptionally closer to telencephalic glutamatergic neurons than SPNs were to telencephalic GABAergic interneurons (Fig. 1C).

Although glutamatergic neurons from distinct cephalic origins do not cluster together, maintaining glutamatergic neurotransmission or associated function could require a common core set of genes. To assess how neurotransmitter utilization relates to genome-wide RNA expression patterns, we examined distributions of gene-gene correlations across cell types (Fig. 1F). Surprisingly few genes are strongly positively correlated with both *SLC17A6* (VGLUT2) and *SLC17A7* (VGLUT1) expression, even those associated with glutamate synthesis and packaging. One hundred sixteen genes had correlated expression to *SLC17A7* (Spearman's tau > 0.5). The median correlation of *SLC17A6* to those 116 genes was centered at tau = 0.05 (Fig. 1G). Only a few genes correlated above tau = 0.5 to both *SLC17A6* and *SLC17A7*, including *BDNF*, *NRN1*, and *TAF1*. Moreover, only 10 genes (*ARPP21*, *BDNF*, *CACNA2D1*, *CHN1*, *CHST8*, *CPNE4*, *LDB2*, *NRN1*, *PTPRK*, and *TAF1*) are differentially expressed (>2.5-fold change) in both telencephalic (*SLC17A7*<sup>+</sup>) glutamatergic neurons and nontelencephalic (*SLC17A6*<sup>+</sup>, *SLC17A7*<sup>-</sup>) glutamatergic neurons relative to GABAergic (*GAD1*<sup>+</sup>) neurons. The bulk of gene expression that distinguishes neuronal types from each other appears incidental to their neurotransmitter identities.

### Adult neuronal types are imprinted by their developmental origin

We inspected RNA expression patterns along with details about each cell type's dissection region of origin to assess which brain structures tended to contain highly similar cell types and which had more dispersed transcriptional profiles. Some tissues, such as the neocortex, gave rise to cell types that exclusively clustered within their cephalic domain (Fig. 1C), a result unlikely to be driven by ambient RNA contamination since regionally variable genes were not over-enriched in neocortical ambient RNA estimates (fig. S3A). However, within a given cephalic domain, cell types from distinct brain structures were often more similar to types sampled from other brain structures. For example, while hippocampal cell types were all found in telencephalic clades in the dendrogram, many individual hippocampal types were more like types in the amygdala or neocortex than they were like other hippocampal types (Fig. 2).

Out of 62 neocortical neuron types, only two types joined clades outside of the major GABAergic and glutamatergic telencephalic branches: (i) the *MEIS2*<sup>+</sup> prefrontal GABAergic types (fig. S1C and Fig. 2), which formed a clade most similar to diencephalic and midbrain *OTX2*<sup>+</sup> types, and (ii) neocortical Cajal-Retzius



**Fig. 2. Gene expression across neural populations.** Expression of broad class marker genes and other genes of interest across all neurons sampled by snRNA-seq. Heatmap colors are scaled to max normalized expression for each row (gene). Dendrogram ordering and metadata colors as in Fig. 1C. Cell types discussed in the main text are labeled at the bottom.

Downloaded from <https://www.science.org> at Massachusetts Institute of Technology on November 28, 2023

neurons, which were more similar to a clade of *LHX9*<sup>+</sup> thalamic neurons despite indications that they predominantly originate from the cortical hem in primates (Figs. 1E and 2) (40). Telencephalic interneuron subtypes tended to cluster with those that express the same class marker. For example, all *PVALB*<sup>+</sup> types sampled across neocortex, amygdala, and striatum clustered closely together (Fig. 2), consistent with a shared developmental origin.

Consistent with previous work which suggested that mammalian thalamus contains both midbrain-derived and forebrain-derived GABAergic interneurons (41), we observed distinct clades of thalamic neurons that were most similar to diencephalic or midbrain populations. Thalamic GABAergic neurons that were *OTX2*<sup>+</sup> were distinct from other thalamic populations (Figs. 1E and 2) and formed a clade with other *OTX2*<sup>+</sup> neurons sampled from brainstem, hypothalamus, and basal forebrain. These populations were *SOX14*<sup>+</sup>, while thalamic *SOX14*<sup>-</sup> populations joined mixed diencephalic-telencephalic clades (Figs. 1E and 2). The dispersion of thalamic neurons to distinct diencephalic and midbrain-dominated clades supports recent work suggesting multiple developmental origins for primate thalamic GABAergic neurons (41). We did not find thalamic GABAergic populations that were transcriptionally similar to telencephalic types (42), potentially owing to nonexhaustive sampling of thalamus, although we note that the next most proximal clade to the thalamic *OTX2*<sup>+</sup> types contained *MEIS2*<sup>+</sup> GABAergic neocortical neurons (Fig. 2).

The amygdala is composed of loosely associated nuclei with diverse phylogenetic and developmental origins (43, 44), and amygdala neuron types distributed widely across the dendrogram (Fig. 1C). Amygdala excitatory neurons, which share properties with cortical and claustrum neurons, clustered with other excitatory telencephalic neurons. In contrast, *FOXP2*<sup>+</sup> amygdala GABAergic neurons, which reside in the intercalated nuclei (19, 45, 46), cluster with striatal SPNs (Fig. 2), in line with recent lineage tracing studies in mice (19) and with analysis of fetal macaque cell types (46). These results underscore the diversity and developmental complexity of cell types comprising the amygdala nuclei.

### Brain cell taxonomies are conserved and are shaped by broad classes of genes

We assessed whether the transcriptomic similarities observed using our hierarchical clustering approach (HCA-PCA) held if using different methods or datasets (Fig. 3). First, we found that the overall dendrogram configuration was broadly conserved when marmoset neuron hierarchical clustering was recomputed using other distance functions (Fig. 3A). To determine whether a consistent pattern of transcriptomic similarity is found in other species, we next repeated PCA-based HCA clustering on a publicly available single-cell RNA sequencing atlas of the adult mouse brain (4).

We found a similar global tree structure between the marmoset and mouse neurons, suggesting broad conservation of the features that drive transcriptomic similarities in neurons across adult primates and rodents (Fig. 3, B and C). In particular, the relative ordering of the major clades was strikingly similar across species: For example, in both species, telencephalic glutamatergic neurons are most similar to a clade of GABAergic neurons that includes striatal SPNs as well as transcriptionally similar types of GABAergic, *FOXP2*<sup>+</sup> neurons in amygdala and hypothalamus (Fig. 3, B to D). Other examples of concordance between species included that all

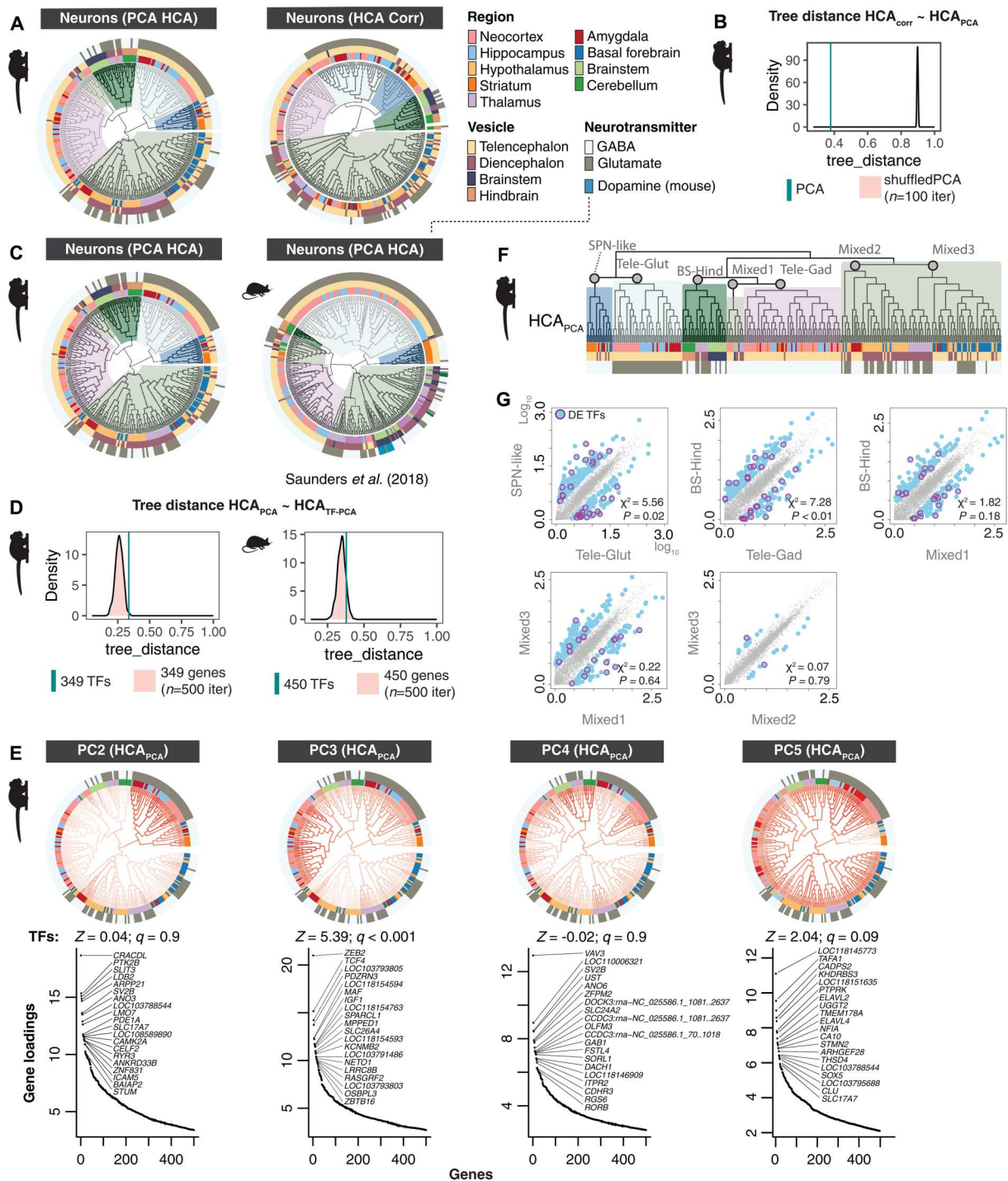
telencephalic GABAergic interneurons formed a large single clade, and the presence of a mixed clade containing thalamic and brainstem types. Cerebellar neuron types were entirely restricted to a single, unmixed clade in both mouse and marmoset (Fig. 3B), although this relative isolation could be due to the lack of other hindbrain structures in the datasets.

Transcription factors (TFs) are master regulators that determine cell type identity in development through temporal patterning, suggesting that they may constitute a key class of genes that maintain transcriptional identity in adults (1). Supporting this view, some TFs are strongly associated with specific brain structures or cephalic compartments, such as *FOXP1* in the telencephalon (Fig. 1E) and *OTX2* in nontelencephalic structures (Fig. 1E). Hierarchical clustering based only on TF genes (349 genes in marmoset, 450 in mouse) was highly similar to the neuronal trees that were computed over all expressed genes (5882 in marmoset, 3528 homologous genes in mouse) (Fig. 3C). However, tree ordering generated by TFs alone did not produce lower tree distances to the original tree compared to similarly sized sets of randomly selected expressed genes (Fig. 3D). This suggests that although TFs undoubtedly play a central role in determining cell identity in development, the suite expressed in adult types are not unique in reflecting the transcriptomic identities of adult neuron types as a whole. Out of the top 20 PCs (which account for 92% variance), 7 PCs were enriched for TFs within the top 200 genes (absolute loadings) compared with randomly sampled genes. Several individual PCs (table S2) loaded on specific clades of the tree, but most showed more complex patterns across clades (Fig. 3E), reinforcing that adult TF expression captures aspects of transcriptional identities but does not drive global transcriptional similarity across neurons.

Borrowing from ancestral state reconstruction methods typically used to estimate evolutionary divergences between genetic sequences or species (47), we applied a maximum likelihood-based approach (fastAnc) to the expression of all genes at each leaf of the marmoset dendrogram (corresponding to observed metacells), along with the branch lengths between adjacent leaves, to reconstruct the transcriptomic state (inferred expression values) at internal nodes of the dendrogram (Fig. 3, F and G). We then compared pairwise “ancestral” expression values of all genes in the parent nodes for each of the seven major clades depicted in Figs. 1C and 3 (F and G). TFs were overrepresented in comparisons of internal nodes that contained cell types stemming from developmentally related brain regions, but were not overrepresented when the leaf clusters stemmed from multiple regions (Fig. 3, F and G). While the expression of some TFs reflects the developmental origins of neurons, some cell types are transcriptionally similar despite having distinct developmental origins. This may reflect aspects of convergence in adult transcriptional profiles (8, 19).

### A hypothalamic-origin neuron type in medial amygdala retains its diencephalic identity

In mammalian CNS, some cell types migrate long distances from proliferative zones to their mature destinations (48, 49). However, neurons generally tend to respect cephalic boundaries and remain within the same subdivision as their progenitors (50). This tight control over migration potential makes it difficult to disentangle the persistent influence of developmental origin from later influences arising from shared tissue context, which could affect all neurons similarly within a brain structure. Cephalic boundary crossings,



**Fig. 3. Conservation of neuronal hierarchy across species and clustering methods.** (A) Comparison between hierarchical clustering (HCA) of marmoset neurons computed on top 100 principal components (PCA) or from a correlation-based distance measure (HCA Corr). (B) Distance between trees computed in (A). Cyan line = tree distance between HCA PCA and HCA Corr. Pink distribution is tree distance scores between HCA<sub>corr</sub> and shuffled PCA scores ( $n = 100$  shuffling iterations). Lower values of tree\_distance ( $x$  axis) indicate higher agreement between dendrogram tree structures. (C) Comparison of marmoset and mouse neurons [mouse atlas data from (4)] using HCA PCA on 5884 genes in marmoset and 3528 1:1 orthologs in mouse. (D) Tree distance comparisons to shuffled distributions using expressed TFs in marmoset and mouse neurons. (E) Top gene loadings for PC2 to PC5 computed over marmoset metacells (table S2). PC scores are plotted on the HCA PCA dendrogram shown in Figs. 1C and 2. Ranked gene loading plots show top 20 genes per PC.  $Z$  scores and  $q$  values show false discovery rate (FDR)-corrected tests of TF enrichments in top loading genes per PC. (F) Marmoset dendrogram in Fig. 1C (HCA<sub>PCA</sub>) indicating major clades compared in (G). (G) “Ancestral” reconstruction (AR) of gene expression profiles of major clades of marmoset neuron types from HCA PCA dendrogram using maximum likelihood estimates. Scatterplots show pairwise comparisons between AR of internal nodes of major clades. Blue dots, genes with  $>3$ -fold change difference between the two ARs; magenta circles, differentially expressed TFs (DE-TFs). Chi-square and  $P$  values describe whether TFs are significantly differentially enriched between the AR pairs.

though rare, do exist (41, 42, 51). Such boundary crossing events can reveal whether cell types that embarked on cross-cephalic migration retain transcriptomic profiles more in common with their tissues of origin, or more in common with their final destinations.

Our data confirm a striking example of cross-cephalic migration in the amygdala. First, we observed an unexpected clustering pattern in our analysis: Despite expressing *SLC17A6* and lacking expression of *GAD1* and *GAD2*, three amygdala neuron clusters showed a closer association with GABAergic rather than glutamatergic types (Fig. 4, A and B). They did not express other genes required for GABAergic transmission such as *SLC32A1* (VGAT) and also lacked the molecular machinery for noncanonical GABA reuptake or release that has been observed in other cell populations (Fig. 2) (52, 53). These findings indicate that these particular amygdala neurons exhibit a "cryptic" transcriptomic identity, in which they are presumably glutamatergic yet have global transcriptional profiles that are more similar to GABAergic types, relative to other telencephalic neurons.

The cryptic amygdalar subtypes display additional atypical gene expression features compared to other amygdala neuronal types, including expression of *OTP* and *SIM1* (Fig. 4B). These TFs are typically expressed in neuronal lineages originating from proliferative zones around the third ventricle. Previous work from other laboratories has shown that in mice, a migratory stream of cells labeled from third ventricle electroporation crosses into the telencephalon (51, 54) and reaches the medial amygdala. This migration is dependent on *Otp* expression; subtypes also express *Sim1* (Fig. 4C) (51). To establish whether these cells are detectable in mouse amygdala, where they were initially described (51, 54), we generated an snRNA-seq dataset consisting of 53,745 nuclei from adult mouse amygdala (P60 to P70). *Otp* protein expression in mouse amygdala becomes largely down-regulated in adulthood (51), although *Otp* mRNA is present in the medial amygdala in P28 mice (Fig. 4D). Consistent with down-regulation of *Otp* protein over development, we did not detect *Otp* gene expression in our snRNA-seq dataset; however, we detected several neuronal clusters that expressed *Sim1* (Fig. 4E).

Mirroring the cryptic transcriptomic profile we observed in the *OTP*<sup>+</sup>/*SIM1*<sup>+</sup> marmoset amygdala neuron types (Fig. 4, A and B), the mouse amygdala *Sim1*<sup>+</sup> neurons clustered with GABAergic neurons, yet expressed *Slc17a6* but not *Gad1* or *Gad2* (Fig. 4E). In mice, these neurons also constitute the majority population in the amygdala that express *Adcyap1* (encoding the protein PACAP) (Fig. 4E), a neuropeptide that is extensively (but not exclusively) expressed in hypothalamic populations (55) and associated with energy homeostasis (56), stress, anxiety (57), and immune responses (58). [In marmosets, *ADCYAP1* is additionally expressed in subsets of *SLC17A7*<sup>+</sup> neurons that do not share the cryptic phenotype (Fig. 4B), indicating that *ADCYAP1* has a distinct pattern of expression in amygdala neurons between mice and marmosets.]

To precisely locate the specific amygdalar nuclei housing the cryptic population, we investigated the expression of *SIM1* in marmosets (59, 60). *SIM1* expression was highly enriched in the neonatal marmoset medial amygdala (MeA) (<https://gene-atlas.brainminds.jp>; Fig. 4F), which mirrors the migration of *Sim1*/*Otp*<sup>+</sup> neurons from the diencephalon in mice into medial amygdala (51). We next verified *SIM1* and *SLC17A6* but not *SIM1* and *GAD1* colocalization in adult marmoset medial amygdala using smFISH (Fig. 4G).

On the basis of these observations, *SIM1*<sup>+</sup>/*OTP*<sup>+</sup> amygdala neurons likely cross from the diencephalon into the telencephalon early in development in both species. This offers a test of the persistence of developmental origin on transcriptomic identities: If amygdalar *SIM1*<sup>+</sup>/*OTP*<sup>+</sup> neurons are globally more similar to hypothalamic types, it suggests that they retain developmental imprint of their origin despite crossing into the telencephalon by P0 (Fig. 4F). We inspected our marmoset neuronal dendrogram to determine whether these amygdala-resident types preferentially associated with telencephalic neurons or diencephalic types. The cryptic amygdala neuron types (as well as several other small amygdala populations that are not *SIM1*<sup>+</sup> or *OTP*<sup>+</sup> but that are transcriptionally similar to the amygdala cryptic types) join a clade with hypothalamic and basal forebrain types (Fig. 4A), separate from most other amygdala types that reside in telencephalic clades. These results suggest that the cryptic amygdala neurons are a conserved population in both mice and primates that likely have diencephalic origins. Consistent with a birthplace imprinting effect, they retain transcriptional identities more similar to diencephalic types than to telencephalic types with which they ultimately reside, despite their early migration into the telencephalon (Fig. 4A).

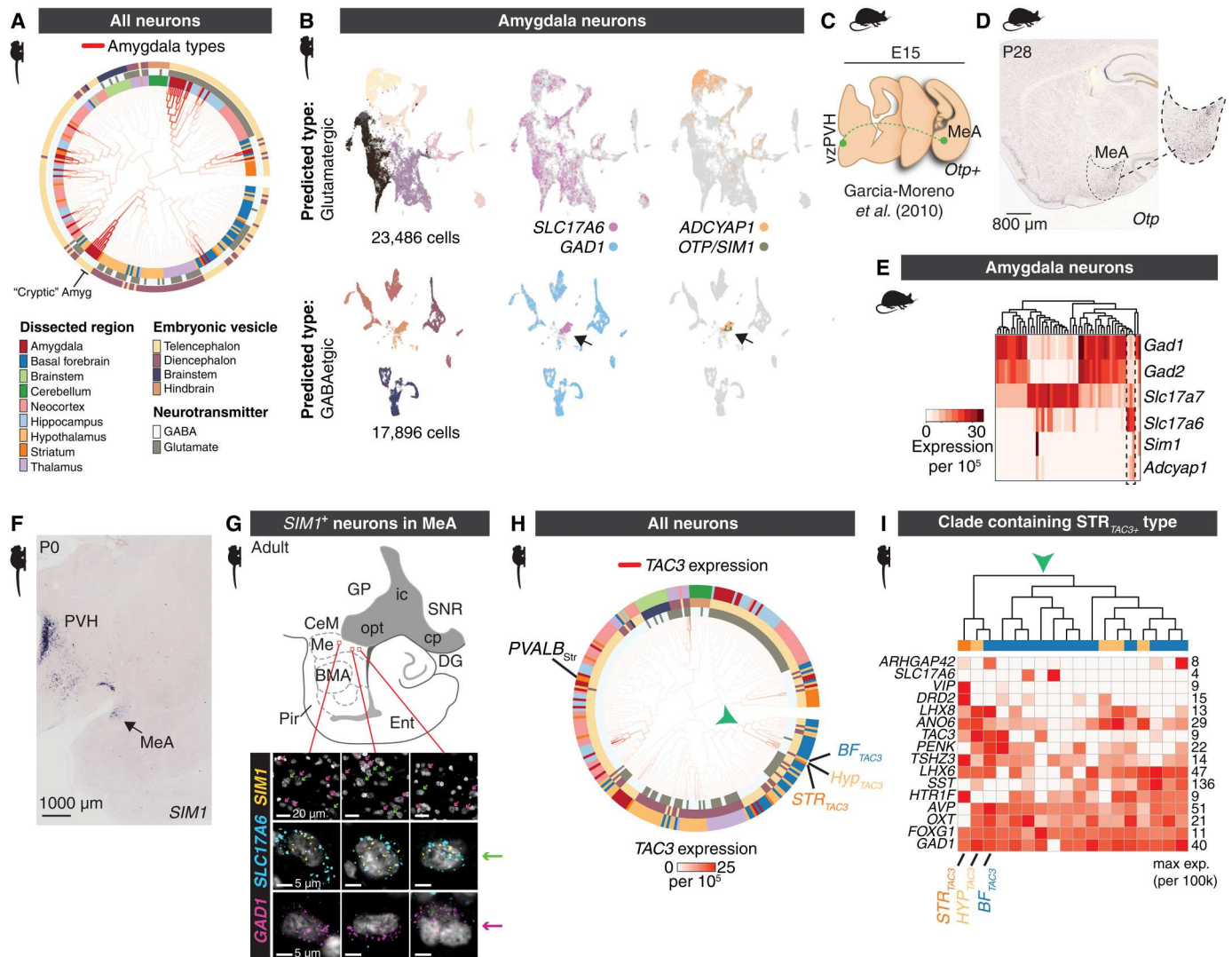
### Primate-specific striatal *TAC3*<sup>+</sup> interneurons are similar to specific *TAC3*<sup>+</sup> diencephalic types

Previously, we discovered a *TAC3*<sup>+</sup>, *LHX6*<sup>+</sup> interneuron subtype in the striatum of humans, macaques, and marmosets that was absent in mice and ferrets (10). Compared with other striatal types, they are transcriptionally most similar to *PVALB*<sup>+</sup> interneurons and, because they expressed the TF *LHX6*, we surmised that they likely also arose from the medial ganglionic eminences (MGEs) (10). Supporting this inference, a recent study of fetal macaque snRNA-seq data from ventral telencephalic progenitor domains found that primate *TAC3*<sup>+</sup> striatal interneurons likely arise from progenitors in the MGE, which diverge from an ancestral progenitor class that also gives rise to conserved *MAF*<sup>+</sup> progenitors that go on to produce *PVALB*<sup>+</sup> and *TH*<sup>+</sup> striatal interneurons (46).

The broader census of brain structures in the current dataset allowed us to examine the transcriptional identity of the striatal *TAC3*<sup>+</sup> type relative to cell types outside of the striatum. *TAC3* is expressed in 20 different neuron types in our dataset, including in cortical GABAergic neurons as well as several amygdala, basal forebrain, thalamic, and hippocampal types (fig. S4, A to C, and Fig. 2). *TAC3*<sup>+</sup> types did not form a single clade in the dendrogram, and their transcriptional resemblance to other neuron types largely reflected their tissue or cephalic origin. For example, thalamic *TAC3*<sup>+</sup> subtypes were found in a thalamic-only clade, while hippocampal *TAC3*<sup>+</sup> neurons were most similar to other hippocampal and amygdala types. The thalamic types were *SLC17A6*<sup>+</sup>, while all other *TAC3*<sup>+</sup> types were GABAergic. These results suggest that *TAC3* expression alone does not mark a transcriptionally uniform set of cell types.

Unexpectedly, the striatal *TAC3*<sup>+</sup> type was most similar not to other striatal interneuron types as we previously concluded (10), but rather to two other *TAC3*<sup>+</sup> GABAergic types found in basal forebrain and hypothalamus (Fig. 4H, *TAC3* smFISH in fig. S4, A and B, and data S1). The broader clade containing these three types (depicted by green arrowhead in Fig. 4I) otherwise consisted entirely of basal forebrain and hypothalamic types. Each of the three *TAC3*<sup>+</sup> populations had distinct gene expression (such as high





**Fig. 4. Examples suggesting cross-cephalic boundary cell type migration.** (A) Locations of amygdala clusters in the neuronal HCA-PCA dendrogram from Fig. 1C. (B) Clustering of marmoset amygdala cells ( $n = 44,165$  neurons) predicted to be glutamatergic (top row) or GABAergic (bottom row) based on a linear classifier (scPred) trained on supervised cell type models. t-distributed Stochastic Neighbor Embedding (t-SNE) for each class is colored by cluster or positive expression of genes (*SLC17A6*, *GAD1*, *ADCYAP1*, *OTP*, *SIM1*). Arrowhead indicates cluster of neurons that are classified as GABAergic, yet do not express *GAD1* and do express *SLC17A6*. (C) Cartoon of embryonic migration of *Otp*<sup>+</sup> cells that migrate from proliferative zones around the third ventricle to periventricular hypothalamus (vzPZH) and medial amygdala (MeA), following data in (51). (D) ISH for *Otp* in sagittal section of P28 mouse brain. Dotted outline indicates borders of medial amygdala (MeA). (E) snRNA-seq from mouse amygdala neurons. Heatmap shows normalized, scaled expression ( $n = 25,930$  nuclei, 3 replicates pooled). Dendrogram shows hierarchical clustering of neuron types based on all expressed genes. Dotted outline shows the presence of three *Slc17a6*<sup>+</sup> subtypes that preferentially cluster with GABAergic (*GAD1*<sup>+</sup>/*GAD2*<sup>+</sup>) subtypes and that express *Sim1* and/or *Adcyap1*. (F) Marmoset P0 coronal section showing ISH staining for *SIM1* from the marmoset BRAIN/Minds atlas (59, 60). (G) Cartoon of amygdala in sagittal plane of adult marmoset. FISH staining for *GAD1* (magenta), *SLC17A6* (cyan), and *SIM1* (yellow) in the medial amygdala (Me). Magenta arrows highlight *GAD1*-expressing nuclei. Green arrows highlight *SLC17A6* and *SIM1* coexpression. Scale bars, 20  $\mu\text{m}$  (top row) and 5  $\mu\text{m}$  (middle and bottom rows). (H) Marmoset neuronal HCA PCA dendrogram with branches colored by *TAC3* expression. Unexpectedly, a primate-specific striatal *TAC3* population (10) clusters with hypothalamic and basal forebrain populations. Arrowhead indicates clade containing *TAC3*<sup>+</sup> striatal type. (I) Heatmap of genes expressed in the clade indicated by green arrowhead in (H). Expression values of each gene are normalized to its max within the clade.

expression of *OXT* and *AVP* in the hypothalamic type, and *DRD2* expression exclusively in the striatal type), ruling out dissection contamination (Fig. 4I). The three types remained direct neighbors when the dendrogram was recomputed using other distance metrics (e.g., correlation-based or PCA scores using TF expression; see Fig. 3).

As the similarity of the *TAC3*<sup>+</sup> striatal type to hypothalamic and basal forebrain neurons was unexpected, we omitted the two basal forebrain and hypothalamic *TAC3*<sup>+</sup> metacells and recomputed the dendrogram using HCA PCA (retaining 286 of the original 288 neuronal types) to determine whether *TAC3*<sup>+</sup> striatal types were globally most similar to hypothalamic neurons (fig. S4C). The global structure of the dendrogram was essentially the same as the original,

except that the striatal  $TAC3^+$  type joined the major telencephalic GABAergic clade with other striatal interneurons (neighboring the striatal  $PVALB^+$  subtype) (fig. S4C), recapitulating our original assessment that considered their similarity only to other striatal interneurons (10). This suggests that  $TAC3^+$  striatal types are not globally most similar to hypothalamic resident neurons, but that they share a particularly close identity to a specific subset of hypothalamic and basal forebrain types that also express  $TAC3$ .

Considering their unexpected transcriptional similarity to both a telencephalic (basal forebrain) and a diencephalic (hypothalamus) type, the  $TAC3^+$  striatal type may arise from a telencephalic progenitor (46) that also gives rise to sister diencephalic (hypothalamic) types, or else shows striking transcriptional convergence with diencephalic types that have distinct developmental origins. Favoring a telencephalic origin as suggested by previous report (46), the hypothalamic, basal forebrain, and striatal  $TAC3^+$  types are all  $FOXG1^+$ , a TF associated with telencephalic origin (Fig. 4I). They also express  $LHX6^+$  and  $NKX2-1^+$  (Fig. 2), consistent with an MGE origin; however, we note that in mice some hypothalamic types also express  $Nkx2-1$  and  $Lhx6$  (30, 61, 62). Ultimately, lineage tracing of the striatal  $TAC3^+$  type in a primate would resolve whether a shared progenitor gives rise to both telencephalic and diencephalic types.

### Neocortical expression fingerprints differ between neurons and glia

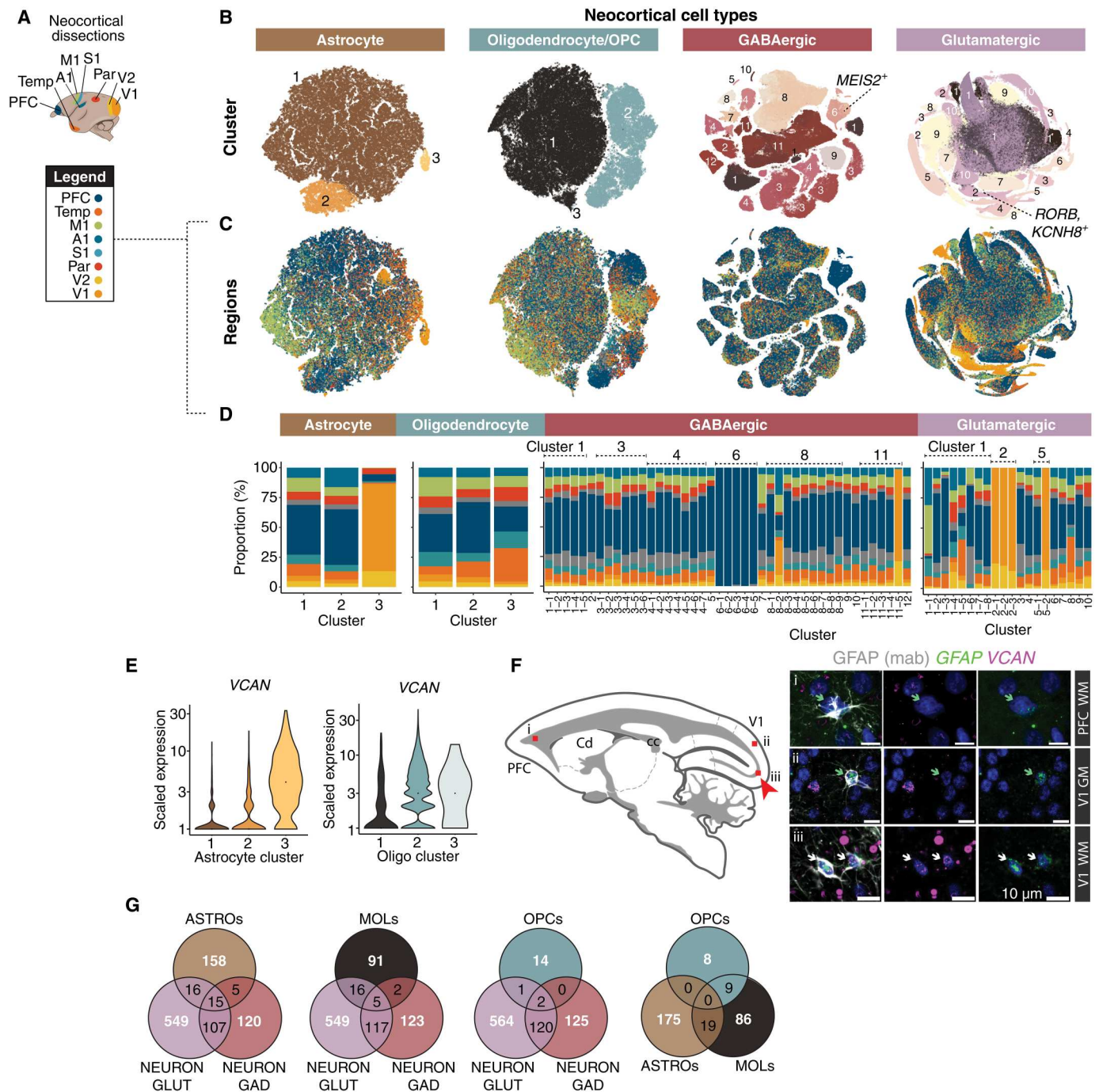
Our cross-structure analysis suggests that in the marmoset brain, global neuronal taxonomies are influenced by broad developmental factors such as cephalic origin (Figs. 1 to 4). However, there is also considerable within-structure heterogeneity of cell types and proportions, particularly in highly patterned structures like the neocortex. We previously showed that primate cortical interneurons regionally varied in their gene expression considerably more than mice, suggesting an evolutionary distinction (10). Within the human neocortex, molecular divergence across functionally specialized areas is already present in cortical progenitors, suggesting that early developmental processes shape intra-cortical cell identities (63, 64). However, it is unclear whether intra-cortical regional changes are shared or distinct across distinct cell classes such as neurons and glia.

We examined regional distinctions in proportions and gene expression of neocortical cell types from eight locations (Fig. 5A). Consistent with prior studies across a range of mammalian species (4, 10, 20, 65), cell subtypes identified in one cortical region were generally present in all other cortical regions, though in different proportions (fig. S1B). There were two main exceptions in neurons. GABAergic  $MEIS2^+$  cells (GABAergic cluster 6; Fig. 5B) were far overrepresented in PFC samples (particularly in dissections of medial and medio-orbital PFC; fig. S1C), a compositional distinction that is not observed in mouse (4, 65), and that may reflect the differential routing of  $MEIS2^+$  neurons into primate neocortex instead of olfactory bulb as they do in mice (46). The second exception was a cluster of  $RORB^+$ ,  $KCNH8^+$  glutamatergic neurons in V1 (and to a lesser extent V2) that was distinct from  $RORB^+$  populations found in the other cortical regions (glutamatergic cluster 2; Fig. 5B). The expansion and divergence of  $RORB^+$  populations in visual cortex is consistent with the elaboration and subspecialization of primary visual cortex layer IV in primates (66).

Glutamatergic neurons sampled from different neocortical locations show regionally distinct gene expression (65). In primates, this is also true of neocortical GABAergic neurons (10). However, the extent to which glial cells are locally customized in distinct regions of neocortex is less well understood (5, 8, 17, 18, 67). Studies in mice have revealed layer-specific astrocyte subpopulations in the cortex (68) as well as variation between neocortical and hippocampal astrocytes (69). We found that one astrocyte type, marked by high expression of  $VCAN$ , was highly enriched in marmoset occipital lobe dissections (V1 and V2) (astrocyte cluster 3; Fig. 5B).  $VCAN$  is also expressed in OPCs, but we did not observe higher  $VCAN$  expression in V1 OPCs (Fig. 5E). Using smFISH, we validated higher colocalization of  $VCAN$  and  $GFAP$  in V1-adjacent white matter compared with PFC white matter and V1 gray matter (Fig. 5F), suggesting that within the neocortex, the regional variation in  $VCAN$  expression is specific to white matter-resident occipital astrocytes.

To address whether cortical regional variation in gene expression is shared across cell types, we performed pairwise comparisons between major clusters of cortical excitatory neurons, inhibitory neurons, astrocytes, and oligodendrocyte lineage types across all eight neocortical locations (Fig. 5E and table S3). Because primate cell types show considerable interindividual variability in gene expression from genetic and other sources (6, 70), we calculated regionally differentially expressed genes (rDEGs) for each of the three donors for which we had sampled the complete set of eight cortical locations, and took the consensus set of rDEGs across individuals. Each cell class displayed rDEGs across neocortical regions (Fig. 5G), an effect that could not be attributed to ambient RNA contamination (fig. S3A). Similar to what has been described in the mouse cortex (68, 69), astrocytes within the marmoset cortex exhibited considerable regional transcriptional variation, but overall neurons had more rDEGs than macroglia (Fig. 5G and fig. S3B). Forty-nine percent of rDEGs in interneurons were also rDEGs in excitatory neurons (Fig. 5G). Although cortical astrocytes and oligodendrocyte lineage cells arise from a closer common lineage with cortical excitatory neurons [dorsal radial glia (71)], they shared a lower percentage of rDEGs in common with excitatory neurons (16% astrocytes, 18% mature oligodendrocytes, 17% OPCs) than the fraction of rDEGs that were shared between excitatory neurons and interneurons, which arise from distinct progenitor populations (72) [though see (73, 74)].

rDEGs within a cell class (glutamatergic, GABAergic, astrocyte, oligodendrocyte lineage) tend to be biased in the same regions across subtypes within that class. However, certain subtypes and regions accumulated more rDEGs than others. For example, across all cell types and particularly within neurons, higher-order temporal association cortex and PFC tended to be most distinct from V1 and V2 (fig. S3B). There was also cell type variability in the extent to which rDEGs were private to individual donors (table S3): 55% and 39% of glutamatergic and GABAergic rDEGs, respectively, were shared among donors, compared to 24% and 14% shared astrocyte and oligodendrocyte rDEGs. These results may indicate lower interindividual consistency in regional gene expression in marmoset nonneuronal cells, although larger sample sizes are needed to determine underlying genetic, environmental, or technical sources (6, 70).



**Fig. 5. Regional variation in neocortical cell types and expression patterns.** (A) Cortical regions sampled. (B) t-SNE embeddings of neocortical macroglia (astrocytes, oligodendrocyte lineage types) and neurons (GABAergic interneurons, glutamatergic neurons). Colors represent clusters (numbered). (C) Same as (B) but cells colored by neocortical dissection. (D) Regional proportions of each cluster; colors same as (A). (E) *VCAN* expression across astrocyte clusters and oligodendrocyte lineage clusters. Colors as in (B). (F) (Left) Cartoon of sagittal section imaged; red boxes (i to iii) correspond to (right) tissue validation of increased abundance of *VCAN<sup>+</sup>* astrocytes in adult marmoset V1-adjacent white matter (iii) compared with PFC-adjacent white matter (i) and V1 gray matter (ii). GFAP antibody (gray) combined with smFISH probes against *VCAN* (magenta) and *GFAP* (green). Green arrows correspond to *GFAP<sup>+</sup>* (antibody) and *GFAP<sup>+</sup>* (smFISH) cells. White arrows correspond to *GFAP<sup>+</sup>* (antibody), *GFAP<sup>+</sup>* (smFISH), and *VCAN<sup>+</sup>* cells. V1, visual cortex V1; PFC, prefrontal cortex; GM, gray matter; WM, white matter. Red arrow highlights locale of *VCAN<sup>+</sup>* *GFAP<sup>+</sup>* cells. Scale bar, 10  $\mu$ m. (G) Venn diagrams showing overlap of neocortical regionally differentially expressed genes (rDEGs) across GABAergic neurons, glutamatergic neurons, astrocytes, and oligodendrocyte lineage cells (MOLs and OPCs). rDEGs are defined as >3-fold expression difference in homologous cell types across pairs of cortical regions.

## Interneurons distribute distinctly across primate neocortex and striatum

Our analysis of RNA expression from the single-nucleus dataset indicates that developmentally linked telencephalic GABAergic populations (48, 49) retain globally similar identities in adulthood. For example, *SST*<sup>+</sup> striatal interneurons are more similar to *SST*<sup>+</sup> hippocampal and neocortical interneurons than they are to other striatal interneuron types (Fig. 2). Within each brain structure, developmentally linked populations become functionally specialized and follow distinct spatial rules for their allocation. In the mouse, quantitative imaging of molecularly-identified interneuron types has revealed that densities of *Sst*<sup>+</sup>, *Vip*<sup>+</sup>, and *Pvalb*<sup>+</sup> cortical interneurons vary across the cortical mantle; their relative proportions relate to unique functional and microcircuit properties of different cortical areas (22). In primates, overall neuron densities are more spatially variable than in mice: They vary by as much as fivefold across the cortical sheet, with highest neuron proportions and densities found in occipital cortex and particularly in V1 (75). However, quantitative mapping of molecularly-defined subtypes of neurons across areas has not been performed in a primate, and it is not known if they have conserved or distinct spatial distributions compared to mice.

We used smFISH to image the distributions of major cortical and striatal interneuron types across sagittal sections in marmoset (Figs. 6 and 7; figs. S5 to S7). We quantified proportions and densities of each type relative to all cells [4',6-diamidino-2-phenylindole (DAPI); fig. S5]. In the neocortex, we binned these using an areal parcellation of marmoset neocortex (76) to determine whether interneuron proportions varied systematically by brain area. In total, we quantified 377,554 neocortical interneurons across 30 sections by smFISH from one marmoset donor (Cj 18-134; table S1). In striatum, we quantified 6848 interneurons across 32 sections. Each series sampled sagittal sections ~160  $\mu$ m apart, beginning 1184 to 1584  $\mu$ m from the midline up to 6384  $\mu$ m laterally, including most of the striatum with the exception of the lateral-most portion of the putamen. We used a cell segmentation algorithm to count positive cells across sagittal sections and expressed interneuron proportions as a percent of all cells (DAPI<sup>+</sup>).

Neocortical types were identified with probes for *SST*, *PVALB*, *CXCL14*, *VIP*, and *LAMP5* (table S4), which collectively account for all major cortical interneuron populations (Fig. 6 and figs. S5 and S6). *CXCL14* is a marker for caudal ganglionic eminence (CGE)-derived cortical interneurons. It is expressed in most *VIP*<sup>+</sup> and *LAMP5*<sup>+</sup> cortical neurons, as well as in a smaller population of *VIP*<sup>-</sup>, *LAMP5*<sup>-</sup> types, some of which are *PAX6*<sup>+</sup> [and which correspond to the *SNCG*<sup>+</sup> population in humans and mice (10, 17)]. As *LAMP5* is also expressed in subsets of excitatory neurons, we performed dual-labeling smFISH with *GAD1* to avoid counting glutamatergic types. Major striatal interneuron types were identified with probes for *SST*, *PVALB*, *CHAT/SLC5A7*, *TH*, *CCK*, and *TAC3* (table S4). These together account for most major populations of non-SPN neurons in the striatum, with the exception of a population of *MEIS2*<sup>+</sup> GABAergic striatal neurons that cluster together with non-SPN GABAergic interneurons (Figs. 1C and 2) but are difficult to distinguish uniquely using smFISH, since several other markers are also expressed at variable levels in other striatal cell types.

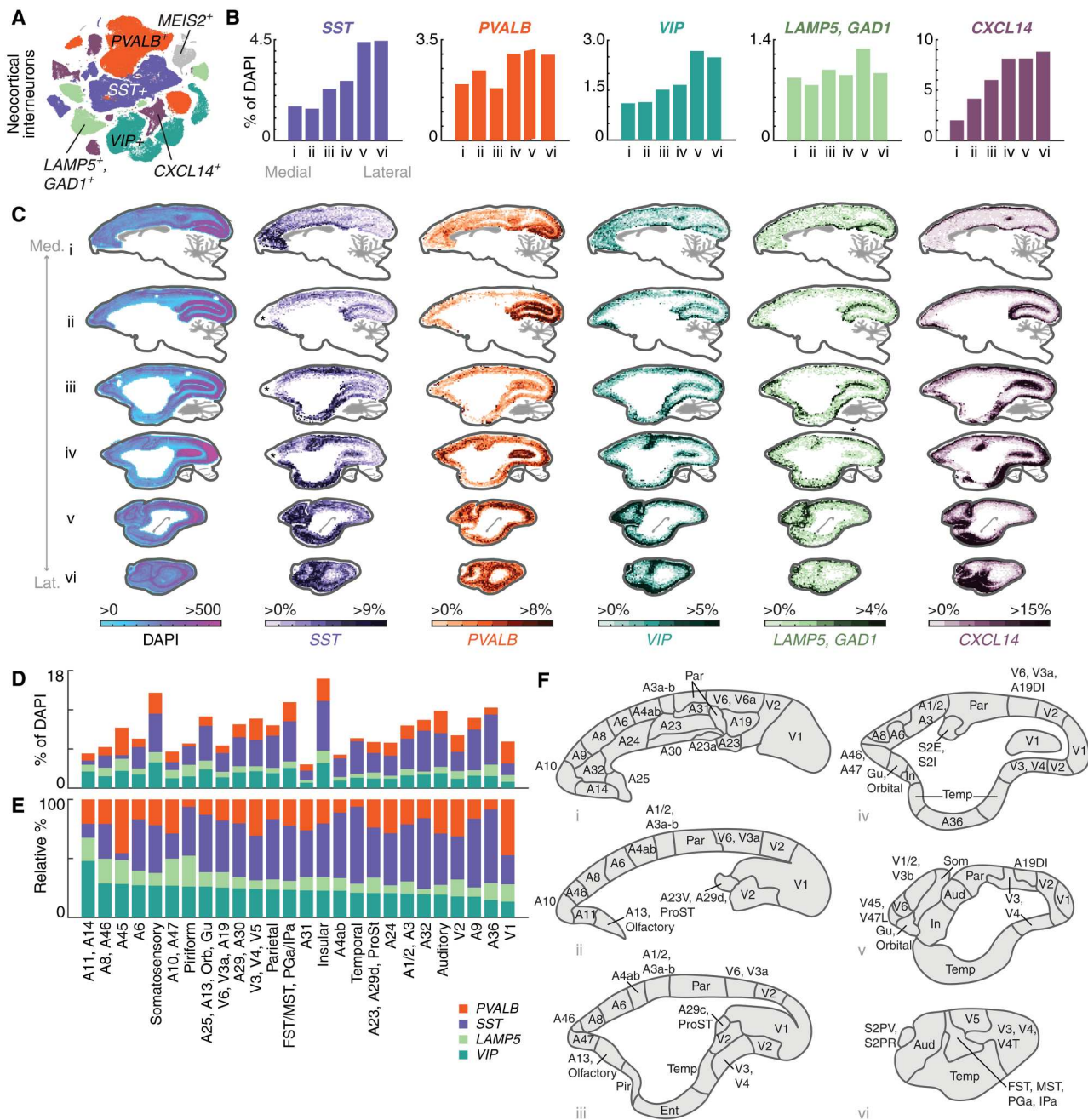
## Neocortical interneuron types have highly focal biodistributions

Quantitative analysis of smFISH revealed highly focal and variable distributions of interneuron subtypes across the marmoset neocortex (Fig. 6, B to E, and fig. S5 and S6). In both absolute numbers and relative proportions, *PVALB*<sup>+</sup> interneurons were strongly enriched in the occipital lobe, particularly along the calcarine sulcus in the medial sections, as well as the occipital pole more laterally (Fig. 6C, fig. S5 and S6, and data S2). *SST*<sup>+</sup> interneurons in the neocortex increase mediolaterally (Fig. 6, B and C; fig. S5; data S2), but closer inspection revealed that this is driven not by a spatial gradient so much as by highly focal enrichments around primary motor area (M1) and primary somatosensory cortex (S3, S1/2), the cingulate cortex, entorhinal cortex, and medial PFC (Fig. 6C). *CXCL14*<sup>+</sup> neurons are enriched along the calcarine sulcus medially, as well as in ventral aspects of the occipital cortex more laterally. There are higher proportions dorsomedially in the parietal cortex (Fig. 6C). *VIP*<sup>+</sup> neurons were enriched in PFC and increased laterally at or near somatosensory cortex and posterior parietal cortex (Fig. 6C). *LAMP5*<sup>+</sup> interneurons showed a bias to the top of the cortical layers, consistent with dominant composition of this class as neurogliaform layer 1 types (10, 77, 78), although this class also contains the *LAMP5/LHX6* type that is found in deeper layers (Fig. 6C) (10, 20).

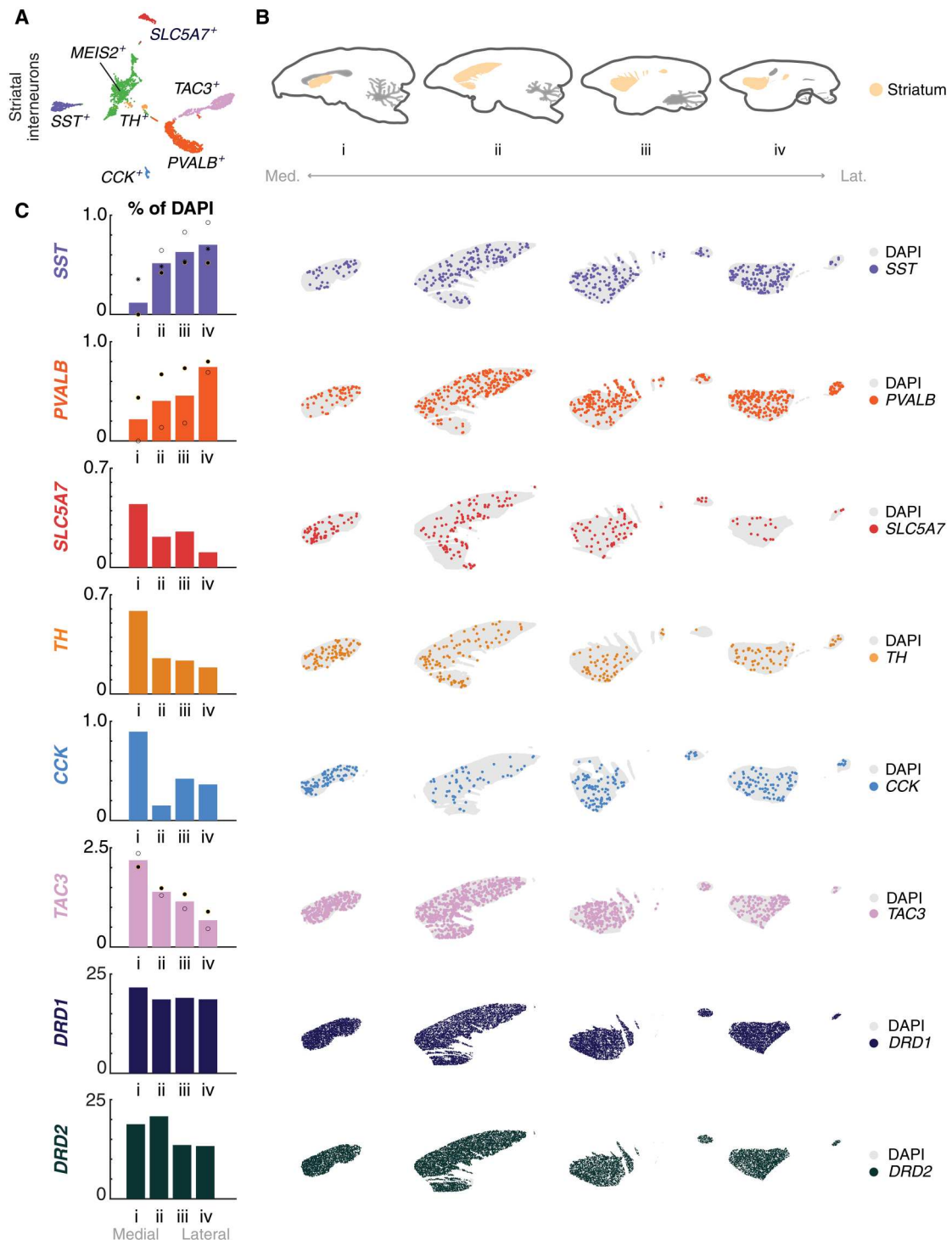
To better appreciate how these distributions relate to neocortical areas, we used a histologically based marmoset neocortical parcellation (76) (Fig. 6, D to F) to bin smFISH interneuron proportions by cortical area (Fig. 6, D and E, and fig. S6 and S7). Small adjacent areas were merged, resulting in 26 areal groupings (Fig. 6F). Overall interneuron proportions relative to all cells varied by 4.5-fold across areal groupings (Fig. 6, D and E, and fig. S6B). As a fraction of all cells (DAPI<sup>+</sup>), A31 (dorsal posterior cingulate) had the lowest overall proportion of interneurons (3%), while insular cortex (In) had the highest (17%) (Fig. 6D).

In mice, quantitative mapping of interneuron densities has revealed higher *Sst*<sup>+</sup> densities and lower *Pvalb*<sup>+</sup> densities in neocortical areas involved in higher cognitive functions, such as medial frontal and lateral association areas (22). This local circuit feature follows cortico-cortical connectivity network topography in mouse (79). Conversely, lower *Sst/Pvalb* ratios are associated with mouse primary motor and sensory areas, which are associated with less distributed (and more local) cortical connectivity (22). To determine whether primate neocortex followed similar or distinct rules of interneuron allocation, we examined normalized proportions of the four largely mutually exclusive interneuron classes (*PVALB*<sup>+</sup>, *SST*<sup>+</sup>, *LAMP5*<sup>+</sup>, *VIP*<sup>+</sup>) within our marmoset areal groupings (Fig. 6, D and E, and fig. S7A). Lateral temporal cortex, including A36, had the highest *SST/PVALB* ratio. Other areas with high *SST/PVALB* ratios included piriform cortex (Pir), M1, and several medial/orbitofrontal areas (A25, A13, A32). This suggests that some marmoset association areas, notably medial frontal area and lateral temporal cortex, have high *SST/PVALB* ratio composition consistent with higher-order association networks in mice (fig. S7A) (22).

Strikingly, other marmoset higher-order areas, including polar and lateral prefrontal areas (particularly A8, A46, A10, A47, and A45), which are thought to be among the most evolutionarily divergent relative to other prefrontal areas (80), and which in primates are characterized by long-range cortico-cortical connectivity to



**Fig. 6. Cell type-specific distributions of interneurons in marmoset neocortex using quantitative smFISH.** (A) t-SNE of GABAergic neocortical interneurons, colored by subclass marker [PVALB, SST, VIP, LAMP5 (GAD1<sup>+</sup>), CXCL14]. Gray points are the MEIS2<sup>+</sup> population that is restricted to orbitomedial PFC (fig. S1, B and C) and was not spatially profiled. (B) Medial-lateral proportions of each major class as percentage of all cells (DAPI<sup>+</sup>). Barplots quantify positive cells as proportion of all (DAPI<sup>+</sup>) cells from medial to lateral sections of smFISH performed on six thin (16 μm) sagittal sections of marmoset neocortex, each section being 1600 μm apart and covering 9600 μm of neocortex. Colors as in (A). (C) smFISH for neocortical interneuron subclass markers showing locations of cells positive for each marker across six sagittal sections of the marmoset neocortex. First column shows density of all DAPI<sup>+</sup> nuclei per unit area (approximately 387 μm per bin) profiled from one series. Heatmap scale in subsequent columns shows percentage of marker-positive cells relative to DAPI<sup>+</sup> cells. Average proportions across section shown in (B). Med., medial; Lat., lateral. Asterisks in (ii) and (iii) of the SST series and (iv) of the LAMP5 series denote regions where tissue or staining artifacts caused loss of signal. These can be seen in the DAPI-only series of these experiments, which show lower overall cell counts at these locations (fig. S5). (D) Quantitation of interneuron proportions by cortical area parcellated according to (F). (E) Relative percentages of interneuron proportions by cortical area shown in (F), sorted by max relative proportion of VIP<sup>+</sup> interneurons. (F) Cartoons of cortical areas and areal groupings used to bin smFISH neocortical interneuron proportions from (A) in (D) and (E). Neocortical parcellation from (76).



**Fig. 7. Cell type-specific distributions of interneurons in marmoset striatum using quantitative smFISH.** (A) t-SNE of striatal, cholinergic neurons (*SLC5A7*) and GABAergic interneurons (*SST*, *PVALB*, *TH*, *CCK*, *TAC3*). Green points correspond to the *MEIS2*<sup>+</sup> population and were not spatially profiled. (B) Cartoon of marmoset striatum illustrating the area profiled, medial to lateral. (C) (Left) Medial-lateral gradients of striatal cell type proportions across four sagittal sections. Dots, individual replicates. Colors as in (A). *DRD1*, dark blue; *DRD2*, dark green. (Right) Scatterplot of smFISH for striatal cholinergic and interneuron subclass markers, as well as dopaminergic striatal populations [*DRD1*<sup>+</sup> (dark blue) and *DRD2*<sup>+</sup> (dark green)], showing locations of cells positive for each subclass marker across four (16 μm) sagittal sections of the marmoset striatum, each 1600 μm apart covering 6400 μm in total.

other association cortices (12, 81, 82), do not exhibit high *SST/PVALB* ratios (Fig. 6, D and E, and fig. S7A). Of all cortical areal groupings we measured, the lowest *SST/PVALB* ratio was found in A45, a higher-order lateral prefrontal area with extensive projections to all lobes of the neocortex (82). These results suggest that marmoset lateral prefrontal areas may be characterized not by exceptional *SST/PVALB* ratios (Fig. 6, D and E, and fig. S7A). Instead, lateral prefrontal areas have the highest total fraction of *VIP*<sup>+</sup> and *LAMP5/GAD1*<sup>+</sup> interneurons (all above 50% of all interneurons); both of these populations predominately arise from the CGE, a progenitor zone that has expanded in primate evolution (83). These results suggest that primate lateral prefrontal areas do not follow the compositional principles that typify frontal areas in the mouse (22) or medial frontal areas in marmoset (Fig. 6, D and E, and fig. S7A).

### Interneuron proportions in marmoset striatum follow mediolateral gradients

The striatum, a crucial brain structure involved in motor control, reward, and decision-making, displays complex functional topography and connectivity. In humans, antero-ventromedial portions containing nucleus accumbens/ventral striatum are functionally coupled to limbic and higher-order cognitive networks, while lateral subdivisions are more coupled to sensory and motor networks (84). Prior work relating bulk gene expression measurements in human and macaque striatum to cortico-striatal network organization found a relationship between functional domain (e.g., somato/motor versus limbic network) and regional differences in gene expression, including enrichment of *PVALB* in lateral portions of the striatum associated with sensory and motor processing (85). These results suggest that differences in local cell type composition across striatum may underlie aspects of its functional topography.

In mice, striatal interneuron subtypes have distinct spatial distributions: Cholinergic (*Chat*<sup>+</sup>) neuron proportions increase dorsally and anteriorly (86), *Pvalb*<sup>+</sup> interneurons are more abundant in dorsolateral striatum than in dorsomedial striatum, and *Sst*<sup>+</sup> interneurons are spatially homogeneous (fig. S7C) (87). Primates retain the major populations of striatal interneurons found in mice (4, 88), and additionally have gained the type distinguished by *TAC3* expression (10, 46) described in previous sections (Fig. 7A). A systematic quantification of striatal interneuron types has not been performed comprehensively in a primate, and it is unknown if they follow similar or distinct spatial distributions observed in mice. We therefore used smFISH to investigate distributions of the major types of conserved striatal interneurons (*SST*, *PVALB*, *SLC5A7/CHAT*, *TH*, *CCK*, *TAC3*; probes in table S4) in serial sagittal sections of marmoset striatum (Fig. 7B).

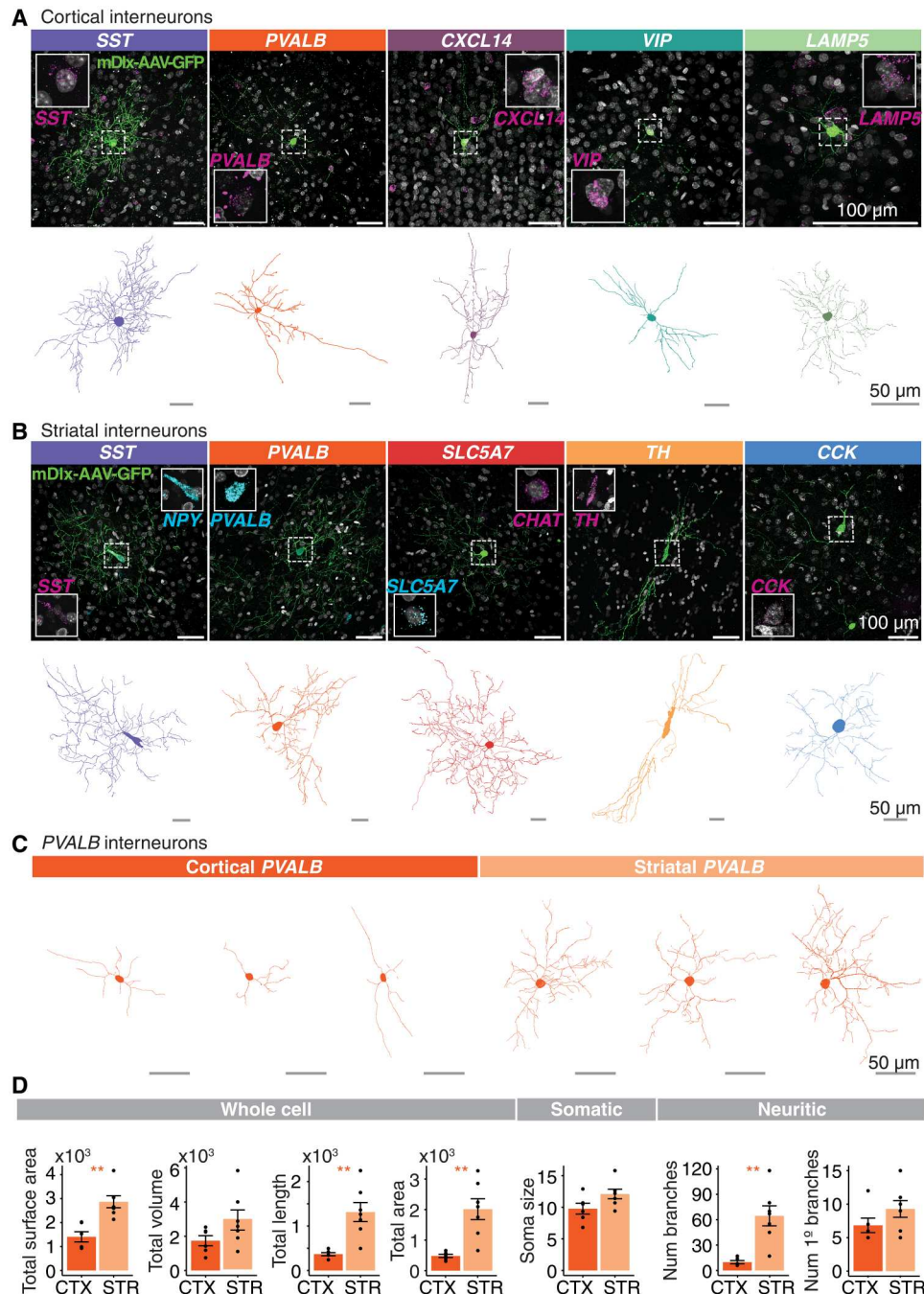
Each striatal interneuron population exhibited a nonuniform distribution across the marmoset striatum, particularly in the medial-lateral axis. Similar to mice, the proportion of striatal *PVALB*<sup>+</sup> interneurons increases in lateral sections, from ~0% to 0.8% of all cells (Fig. 7C and data S2). Unlike mice, marmoset *SST*<sup>+</sup> interneuron distribution is nonuniform, appearing sparse near the midline and increasing in proportion (0.1 to 0.7%; Fig. 7C and data S2). Cholinergic neurons (*CHAT*<sup>+</sup>) show the opposite medial-lateral gradient (0.45 to 0.1%; Fig. 7C and data S2). Similar to *CHAT*<sup>+</sup> neurons, *TH*<sup>+</sup> striatal interneurons, which are transcriptionally similar to the *PVALB*<sup>+</sup> type, exhibited a decreasing medial-lateral gradient (Fig. 7C and data S2). *CCK*<sup>+</sup> striatal

interneurons, which are a minority population in marmoset (10) and mouse (88), are enriched close to the midline and become much sparser laterally (Fig. 7C and data S2). The *TAC3*<sup>+</sup> interneurons showed a decreasing medial-lateral gradient, similar to *CHAT*<sup>+</sup> neurons (2 to 0.5%; Fig. 7C and data S2). No striatal population exhibited anterior-posterior or dorsal-ventral gradients with the exception of *PVALB* interneurons, which showed a modest dorsal-ventral gradient. Unlike striatal interneurons, SPN proportions were largely uniform across the major axes: *DRD1* and *DRD2*, which distinguish direct and indirect SPNs, respectively, were slightly enriched medially but otherwise exhibited largely uniform distributions across the striatum (Fig. 7C). Overall, these findings suggest that the spatial distribution of interneuron subtypes within the primate striatum may play a role in shaping its functional topography and connectivity. While mice and primates share similar striatal interneuron populations, there are systematic differences in their spatial distributions.

### Reconstructions reveal regional variation in interneuron morphology

The morphology of interneuron types relates essentially to their function and contribution to neural circuits. While methods such as biocytin filling and Golgi staining are the gold standard for morphological reconstructions, low-titer AAVs carrying membrane-bound fluorescent reporters can be used to sparsely label cells for morphological reconstruction. We used a reporter AAV under the control of the forebrain interneuron-specific mDlx enhancer (29) to label neocortical and striatal interneurons, and then performed smFISH (probes in table S4) on thick sections (120 μm) to confirm the molecular identity of GFP<sup>+</sup> cells. Reconstructions using viral labeling are more challenging than with single-cell filling methods, because GFP expression from neighboring cells and passing fibers has to be distinguished from signal attributable to the target cell. In some cases, the GFP signal appears punctate, making it challenging to follow discontinuous processes. For these reasons, our reconstructions are conservative: We aimed to avoid reconstructing false-positive fibers (fibers originating from other cells) and, in some cases, likely under-ascertain the full dendritic arborization of target cells.

We imaged 1203 GFP<sup>+</sup> neurons in the striatum and 4374 GFP<sup>+</sup> neurons in the neocortex. We used NeuTube to reconstruct the top 216 telencephalic neurons that had the qualitatively highest signal to noise and most complete soma and processes based on GFP signal, did not have other cells labeled in the field of view (FOV), and were positive for at least one smFISH probe (table S5; examples in fig. S8). Raw image stacks and NeuTube reconstructions are available from the BIL. Using combinations of one to two probes for marker genes of different types, we identified GFP<sup>+</sup> neocortical and striatal interneurons, respectively, with smFISH (neocortex: *SST*, *PVALB*, *CXCL14*, *VIP*, *LAMP5*; striatum: *SST*, *PVALB*, *SLC5A7*, *TH*, *CCK*; table S4) (Fig. 8, A and B, top rows). We used a second tracing and reconstruction method (Imaris) on the 41 most complete GFP<sup>+</sup> cells. Of these, we discarded 5 due to incomplete soma or highly discontinuous processes in the image stack, retaining 36 cells (Fig. 8, A and B, bottom rows, and table S6). Morphological parameters were measured using the Surface function, which detects surface area and volume based on the fluorescence of the mDlx-AAV-GFP expression, and the Filament Tracer function, which traces structural features starting from the soma to the terminal



**Fig. 8. Morphological characteristics of marmoset cortical and striatal interneurons.** (A and B) Top rows: Examples of AAV9-hDlx5/6-GFP-fGFP labeled neocortical (A) and striatal (B) interneurons. Insets show magnified cell nucleus of GFP<sup>+</sup> cell along with smFISH staining for interneuron type marker to confirm molecular identity of labeled cell. Scale bar, 100 μm. Bottom rows: Reconstructed skeletonized morphology (Imaris) of GFP<sup>+</sup> cells depicted in the top rows. Scale bar, 50 μm. (C) Representative reconstructed neocortical and striatal PVALB<sup>+</sup> interneurons from one marmoset (Cj 17-154; tables S1 and S6). Scale bar, 50 μm. (D) Quantification of morphological characteristics of PVALB<sup>+</sup> cells from Cj 17-154. Means ± SEM. \*P < 0.05 and \*\*\*P < 0.01. See Materials and Methods for detailed statistical information.

processes based on the diameter of the soma and the thinnest projection of the cell (table S6). To assess whether there are region-dependent morphological differences within molecularly similar types, we compared several parameters between a collection of neocortical and striatal PVALB<sup>+</sup> GFP<sup>+</sup> cells that were reconstructed from the same marmoset (Cj 17-154; Fig. 8, C and D) to avoid

variable tissue shrinkage arising from different tissue storage conditions across the animals (see Materials and Methods). Striatal PVALB<sup>+</sup> GFP<sup>+</sup> cells were larger than cortical PVALB<sup>+</sup> GFP<sup>+</sup> cells in terms of length and area, but not soma diameter or the volume of GFP<sup>+</sup> fluorescence (Fig. 8D). While the number of branches issued by their somas also did not differ, the striatal PVALB<sup>+</sup>



GFP<sup>+</sup> interneurons exhibited more dendritic branch points than their cortical counterparts (Fig. 8D), suggesting that marmoset striatal *PVALB*<sup>+</sup> interneurons have greater arborization than *PVALB*<sup>+</sup> cortical interneurons.

### An enhancer-AAV labels a primate-specific interneuron type

Given that the primate-specific *TAC3*<sup>+</sup> type comprises almost 30% of striatal interneurons in marmoset (10), we expected a sizable proportion of striatal GFP<sup>+</sup> cells labeled by AAV9-h*Dlx5/6*-GFP-fGFP to be *TAC3*<sup>+</sup>. However, while 55 GFP<sup>+</sup> striatal cells were imaged across all smFISH *TAC3* probe-treated sections, we failed to find any colocalization of GFP with *TAC3* transcripts. In these experiments, which were double-labeled with *TAC3* and *PVALB* probes, 9 of 55 were GFP<sup>+</sup>, *PVALB*<sup>+</sup> (16% *PVALB*<sup>+</sup>, versus expected 21% of interneurons expected from snRNA-seq proportions), while 0 of 55 were GFP<sup>+</sup>, *TAC3*<sup>+</sup> (0% versus 28% expected from snRNA-seq proportions).

The mDlx enhancer is a regulatory element specific to forebrain interneurons (89). Although the regulatory element, and the forebrain-interneuron expressed genes flanking it, *Dlx5* and *Dlx6*, are highly conserved in vertebrates, we wondered whether the lack of accessibility of the mDlx locus in the *TAC3* interneurons could explain our inability to find colocalization of *TAC3* expression and GFP. To assess this, we generated snATAC-seq data (69,808 nuclei) from fresh marmoset striatum (Fig. 9A). We used Signac (90) to integrate our previously annotated striatal snRNA-seq data and identify major striatal types. We then examined accessibility of the marmoset sequence homologous to the mDlx locus across interneuron types. While other striatal interneuron types (particularly the *SST*<sup>+</sup> and *PVALB*<sup>+</sup> types) showed accessibility (ATAC-seq peaks, reflecting chromatin accessibility) at the mDlx locus, the *TAC3*<sup>+</sup> type did not (Fig. 9B).

To develop a viral tool that could transduce the striatal *TAC3*<sup>+</sup> cell type, we next nominated candidate regulatory elements specific to the *TAC3*<sup>+</sup> interneuron type. We used Signac to identify differentially accessible peaks in the *TAC3*<sup>+</sup> cluster relative to all others and filtered the set by fold change, percent accessibility across the target cell type population, and peak size (Fig. 9C). To maximize the likelihood of obtaining a functional reporter while minimizing the number of marmosets used, and because several of our top candidates were very short, we selected four top regulatory element candidates for the *TAC3*<sup>+</sup> type for tandem packaging (see Materials and Methods). The four candidates were located on four different chromosomes and spanned between 94 and 215 base pairs (bp). One site was in an exon of *CDH13*, one was in the first exon of *TAC3*, one was in an exon of *LOC108592287*, and one was intergenic (closest gene 200 kb distance). These four elements were packaged in tandem in an AAV9 vector (AAV9-tandemE-*TAC3*-EGFP) containing a cytoplasmic enhanced GFP (EGFP) reporter (Fig. 9D). We delivered the virus via magnetic resonance imaging (MRI)-guided local injection into the anterior striatum of two marmosets (Cj 19-207 and 17-B111; Fig. 9E, fig. S9A, and table S1) and imaged coronal sections of striatum for EGFP-positive cells after 6 and 10 weeks of incubation time (Fig. 9F and fig. S9B). smFISH confirmed colocalization of *TAC3* transcripts in EGFP<sup>+</sup> neurons in striatum (Fig. 9G and fig. S9, C and D). In one animal, the virus diffused beyond the boundary of the striatum. In this animal, we also detected strongly labeled EGFP<sup>+</sup> cells in a border zone between striatum, BNST (bed

nucleus of the stria terminalis), and the globus pallidus. smFISH confirmed that these cells, too, were *TAC3*<sup>+</sup> (fig. S9, B to E).

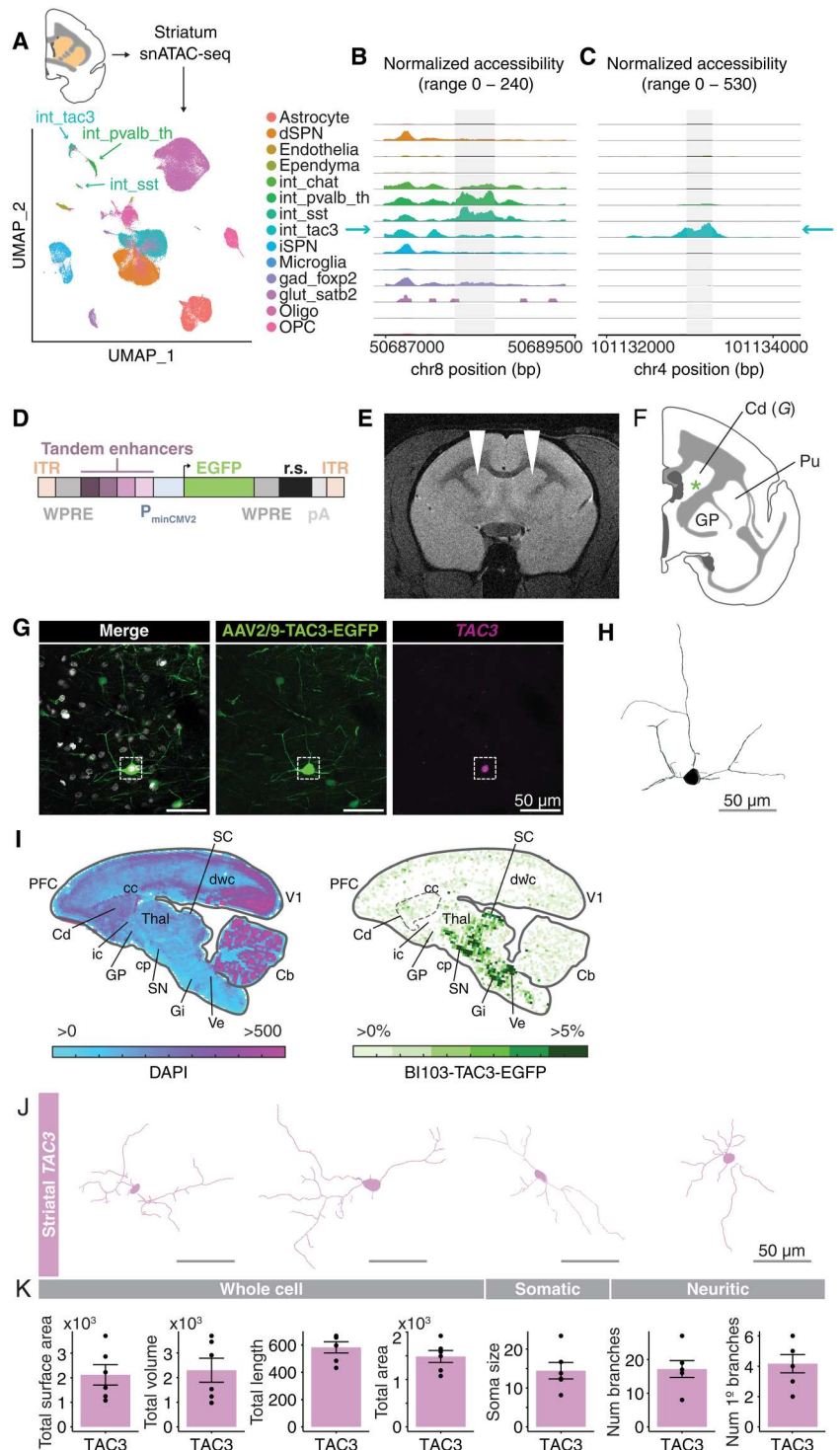
To determine the broader biodistribution of cells transduced by the tandemE-*TAC3* enhancer, we next packaged the same enhancers in an AAV capsid (BI103) capable of efficient transduction of brain cell types after systemic intravenous delivery in marmosets. We delivered this AAV (AAV-BI103-tandemE-*TAC3*-EGFP) to one marmoset (Cj 20-214; table S1). After a 4-week incubation, the brain was perfused, sliced, and stained with anti-GFP antibody to amplify EGFP signal, and with smFISH probes against *TAC3*<sup>+</sup> to confirm colocalization. We found EGFP<sup>+</sup>/*TAC3*<sup>+</sup> colocalized cells in striatum as well as several extra-striatal locations including hypothalamus, substantia nigra, superior colliculus, brainstem, and neocortex (Fig. 9I and fig. S9, F and G). To assess the morphology of *TAC3*<sup>+</sup> striatal interneurons produced by viral labeling, we reconstructed several of the most complete cells from both the local injections (AAV9-tandemE-*TAC3*-EGFP) and the systemic injection (AAV-BI103-tandemE-*TAC3*-EGFP) using Imaris (Fig. 9, H, J, and K, and table S6). The cells tended to have two to three thick branches that extended from the soma, and which bifurcated close to the soma and became thinner thereafter. Reconstructed cells had median 13.1  $\mu\text{m}$  ( $\pm$  5.1 SD) soma diameter, 2221.3 volume ( $\mu\text{m}^3$ ;  $\pm$  1192.8 SD), and 16.5 ( $\pm$  6.11) total dendritic branch points. These experiments show that whereas systemic injections of AAV9 under the mDlx enhancer could not transduce the striatal *TAC3*<sup>+</sup> type, presumably due to lack of accessibility at the mDlx locus in this cell type (Fig. 9B), local injections of AAV9 as well as systemic injections with alternative capsid with the cell type-targeted enhancers were successful. Broadly, the use of cell type-specific regulatory elements, coupled with viral engineering, enables a new frontier for the study of highly divergent primate brain cell types.

### DISCUSSION

Here, we generated a molecular and cellular census of the adult marmoset brain using snRNA-seq, quantitative smFISH, and cell type-specific AAV labeling. Our study reveals the complex repertoire of cell types in the marmoset brain. Our snRNA-seq dataset of over 2.4 million brain cells across 18 brain regions in the marmoset indicates that lineage is an important factor shaping adult transcriptomic identity of neuronal types, apparently more so than neurotransmitter utilization (Figs. 1 to 5). Using quantitative smFISH, we revealed, for the first time in a primate, the spatial distributions of molecularly resolved GABAergic interneuron types (Figs. 6 and 7). Using GFP delivered by interneuron-specific AAVs, we generated morphological reconstructions of all major interneuron types in the neocortex and striatum (Figs. 8 and 9). We generated a viral genetic tool under the control of cell type-specific regulatory elements to successfully label a previously described putative primate-specific striatal interneuron type (Fig. 9). These datasets, generated as part of the BICCN, complement other recent and extensive cellular profiling studies in other species (1, 2, 30–32) and will enable insights in cell type innovations and modifications in future comparative studies.

Telencephalic glutamatergic and GABAergic neurons strongly segregate in mammals as well as in homologous structures in amphibians (8), suggesting an evolutionarily conserved distinction. An initial atlas in mouse indicated that glutamatergic and GABAergic neurons from diencephalic and midbrain structures also partition

**Fig. 9. Development of an enhancer-AAV for TAC3primate-specific striatal interneurons.** **(A)** snATAC-seq (69,808 nuclei) from fresh marmoset striatum (Cj 18-153; table S1). Uniform Manifold Approximation and Projection (UMAP) with labels transferred from striatal snRNA-seq. **(B)** TAC3<sup>+</sup> interneurons lack accessibility at the locus corresponding to the mDlx sequence (89) in marmoset. **(C)** Chromatin accessibility of a candidate enhancer for the TAC3<sup>+</sup> type. **(D)** Construct design for TAC3 interneuron AAV. Four candidate regulatory elements [example in (C)] were packaged in tandem upstream of a minimal promoter driving EGFP expression. ITR, inverted terminal repeats; WPRE, Woodchuck hepatitis virus posttranscriptional regulatory element; P<sub>minCMV2</sub>, minimal CMV2 promoter; EGFP, enhanced green fluorescent protein; r.s., 300 bp random sequence; pA, polyadenylation sequence. **(E)** MRI showing injection location (white arrowheads) of the AAV9-tandemE-TAC3-EGFP virus into dorsal striatum in one animal (Cj 19-207; table S1). **(F)** Location of cell shown (G). **(G)** Main: EGFP antibody–amplified image of a labeled cell [position in (F)]. Insets: smFISH of TAC3 colocalization. Scale bar, 50 μm. **(H)** Morphological reconstruction (Imaris) of cell in (G). Scale bar, 50 μm. **(I)** Heatmaps showing the density of cells transduced by AAV-BI103-tandemE-TAC3-EGFP after systemic intravenous injection. Left, density of all DAPI<sup>+</sup> nuclei per unit area profiled (approximately 344 μm per bin); right, density of EGFP<sup>+</sup> cells relative to DAPI<sup>+</sup> cells. PFC, prefrontal cortex; Cd, caudate; cc, corpus callosum; ic, internal capsule; GP, globus pallidus; Thal, thalamus; cp, cerebral peduncle; SN, substantia nigra; Gi, gigantocellular reticular nucleus; SC, superior colliculus; dwc, deep cortical white matter; Ve, vestibular nucleus; V1, visual cortex V1; Cb, cerebellum. **(J)** Reconstructions (Imaris) of striatal EGFP<sup>+</sup>/TAC3<sup>+</sup> cells. **(K)** Morphological parameters of reconstructed cells from striatal EGFP<sup>+</sup>/TAC3<sup>+</sup> cells from local injection (E) and systemic injection (I) (table S6).



almost perfectly by neurotransmitter usage (5), suggesting that this could be a general rule. However, more recent transcriptomic censuses in mouse (1, 30) and human (2) indicate that across mammalian species, many nontelencephalic glutamatergic and GABAergic types tend to form highly intermixed clades that do not separate clearly by neurotransmitter identity. Our results in the adult marmoset (and mouse; Fig. 3) concord with the notion that gene

expression distinctions between telencephalic glutamatergic and GABAergic neurons do not hold for neurons in other brain structures (Figs. 1 to 3). This has implications for generalizing transcriptomic associations to other phenotypes. For example, transcriptomic changes in glutamatergic or GABAergic neurons have been associated to diseases such as autism and schizophrenia (91–93). Such associations may not generalize to GABAergic or

glutamatergic types outside of the sampled brain region (usually neocortex), consistent with observations that “global” changes to glutamatergic or GABAergic neurons in relation to disease actually often only surface in a few brain regions (94).

A previous analysis, based on shared patterns of several key TFs, proposed that telencephalic GABAergic neurons are developmentally and evolutionarily related to diencephalic GABAergic neurons (95). Our results indicate that when profiled at adulthood, only a limited number of telencephalic GABAergic types are transcriptionally similar to diencephalic types, some of which may actually arise from cephalic boundary crossings earlier in development. Most neocortical, hippocampal, and some amygdalar and striatal GABAergic types are so distinct from diencephalic GABAergic types that they share more gene expression in common with telencephalic glutamatergic types (Fig. 1C). In general, the hierarchical arrangement of neuron types (the neuronal taxonomy) across brain structures was preserved across marmoset and mouse datasets (Fig. 3C), suggesting an evolutionarily conserved arrangement. An important future task is to leverage recent brain cell census datasets in macaque (31) and human (2) to investigate the consistency of these observations over a broader range of species.

Developmental origin or shared lineage plays a strong role in shaping the adult transcriptomic identity of neurons, but phenotypic convergence, whereby adult cell types converge on similar transcriptomic identities despite a nonshared developmental origin, may also drive apparent similarities among cell types. Recent advances in lineage tracing coupled with single-cell RNA sequencing demonstrates that phenotypic convergence is surprisingly common between transcriptomically defined types (19, 74). As such, it is difficult to disambiguate between these possibilities in the absence of data that confirm lineage directly (8, 19). For example, in our data, the clade of GABAergic projection neurons that contains striatal SPNs and amygdala and basal forebrain *GAD1*<sup>+</sup>, *FOXP2*<sup>+</sup> neurons also contained several subtypes of hypothalamic *GAD1*<sup>+</sup>, *FOXP2*<sup>+</sup> neurons (Figs. 1C and 4 and fig. S2D). Either these hypothalamic *FOXP2*<sup>+</sup> subtypes have a convergent expression identity with long-range GABAergic projection neurons of the telencephalon or they arise from a common lineage.

Our analysis of nonneuronal cells revealed striking regional heterogeneity of astrocytes across major brain divisions, to a greater extent than oligodendrocytes, OPCs, microglia, ependymal cells, and vascular cells (Fig. 1D and fig. S2). Regional heterogeneity of astrocytes is well established in the mouse (5, 96–98) and has recently been demonstrated in marmoset (18) and human (2) transcriptomic studies. The division of astrocyte populations across cephalic boundaries, and their differential expression of early patterning genes such as *FOXG1* [fig. S2, F to H; also (18)], suggests a developmental origin for astrocyte regional heterogeneity [discussed in (67)], perhaps as early as the first trimester in humans (99). Our atlas builds upon these findings by providing increased resolution of intra-cephalic regional heterogeneity in astrocytes, which we explored in the cortex. We found moderate proportional differences in cortical area representation between astrocyte and oligodendrocyte clusters (Fig. 5, C to E), except for an astrocyte subtype (cluster 3) marked by high *VCAN* expression that localized to V1-adjacent white matter. Notably, astrocytes and oligodendrocytes within the cortex exhibited unique regional gene expression signatures between cortical areas that were not shared with neurons, despite

a common developmental origin between cortical glutamatergic neurons and macroglia (71, 99–102).

Our data support known instances of cross-cephalic migrations in thalamus and amygdala, and also suggest new ones (Fig. 4). The similarity of the primate-specific *TAC3*<sup>+</sup> striatal type to hypothalamic and basal forebrain *TAC3*<sup>+</sup> types is unexpected (Fig. 4H) and suggests another possible cross-cephalic vesicle migration event. Recent work (46) proposed that the striatal *TAC3*<sup>+</sup> type has a ventral ventricular telencephalic origin, similar to most other GABAergic interneuron types destined for striatum. While phenotypic convergence remains possible, another possibility is that a ventral telencephalic progenitor gives rise to both telencephalic and diencephalic types. This is supported by the expression of *FOXG1*, a TF necessary for ventral telencephalic fate (103), in all three transcriptionally similar *TAC3*<sup>+</sup> populations. While cross-cephalic migration events may be nominated on the basis of unexpected transcriptional similarity, as we did here using hierarchical clustering analyses (Figs. 3 and 5 and fig. S4), ultimately these and other instances of unexpected transcriptional similarity should be investigated using lineage tracing or other methods capable of inferring developmental relationships directly.

Developmentally linked interneuron populations in striatum and neocortex (e.g., *PVALB*<sup>+</sup> types in both structures) displayed distinct spatial distributions measured by cell counting of smFISH imaged across serial sections of the marmoset brain (Figs. 6 and 7). While all interneuron subtype distributions in the striatum followed an increasing or decreasing gradient along a medial-lateral axis, the interneurons in the neocortex followed more complex distributions that were not captured by simple gradients. In the mouse, *Sst*<sup>+</sup>/*Pvalb*<sup>+</sup> ratios are a hallmark of higher-order cortex (22), but we found a different pattern across much of the higher-order association cortex in marmosets. Areas in the marmoset lateral PFC were unique in their relatively high proportions of *VIP*<sup>+</sup> neurons (and had unexceptional or even low ratios of *SST*<sup>+</sup>/*PVALB*<sup>+</sup> neurons) (Fig. 6 and figs. S5 and S6). These results suggest that primate lateral prefrontal areas may not follow the same local cortical circuit organizing principles that pertain to mouse association cortex.

Morphological characterization of the striatal and neocortical interneuron populations suggests variation in overall size and dendritic arborization among subtypes (Fig. 8). Striatal *PVALB*<sup>+</sup> cells were found to be larger in length, volume, and surface area and consisted of more dendritic branch points than the cortical cells. Our data altogether suggest that the striatal *PVALB*<sup>+</sup> interneurons exhibit higher dendritic complexity compared to their respective counterparts in the neocortex (Fig. 8, C and D). These morphological differences could result in differing electrophysiological properties. For example, it is possible that *PVALB*<sup>+</sup> cells in the striatum have higher capacitance and receive more synaptic inputs onto them, thereby impacting local signal integration and computation, than cortical *PVALB*<sup>+</sup> interneurons. The functional importance of morphological differences across the interneuron subtypes identified in our study will ultimately require studying how these cells affect the circuits in which they reside (i.e., cellular/subcellular targeting biases and functional properties) (27, 104, 105).

Recent bioengineering advances have improved delivery of viruses across the blood-brain barrier to enable cell type-specific access in nonhuman primates (17, 29, 106–109). To maximize translatability across species, many approaches that use cell type-

specific regulatory elements to restrict expression rely on evolutionary conservation (e.g., between mice and humans) as a screening criterion. However, the evolutionarily conserved mDlx enhancer, which generally labels telencephalic interneurons in a range of mammalian species, did not drive expression in the primate-specific striatal  $TAC3^+$  cell type (Fig. 9B). This could be due to distinct developmental origins (Fig. 4, H and I), or because the  $TAC3^+$  type evolved along the primate lineage and accumulated a divergent regulatory landscape. However, we also observed that the mDlx enhancer selectively under-ascertained several other interneuron populations, and often in region and cell type-specific ways. For example, it systematically under-labeled  $VIP^+$  and  $SST^+$  types in the neocortex (expected 22% and 26% of interneurons, obtained 2% and 3%, respectively), as well as  $SST^+$  interneurons in the striatum (14%, obtained 1.8%). This under-labeling could in part be due to the titer and systemic delivery approach we adopted, which was necessary to achieve sparse labeling for morphology.

The lack of transduction of  $TAC3^+$  striatal interneurons using the AAV-mDlx virus prompted us to develop an AAV (AAV9-tandemE-TAC3-EGFP) under the control of candidate  $TAC3^+$  interneuron striatal regulatory elements (Fig. 9, D to H). This enabled initial reconstructions of striatal  $TAC3^+$  interneuron morphology (Fig. 9, H and J). When delivered systemically using a capsid that has efficient transduction across the blood-brain barrier in marmosets (Bi103-tandemE-TAC3-EGFP), the virus also labeled  $TAC3^+$  neurons elsewhere in the brain, including in the hypothalamus, neocortex, and substantia nigra. Neocortical  $TAC3^+$  cells bear little resemblance to  $TAC3^+$  striatal interneurons in terms of their global gene expression profiles (Fig. 4H). Therefore, the widespread transduction of  $TAC3^+$  neurons suggests that one or more of the regulatory elements that we used (originally defined only relative to other striatal cell types) may endogenously regulate the  $TAC3$  gene itself, or else may control another gene that tends to be commonly expressed in developmentally and transcriptionally disparate  $TAC3^+$  populations. Further refinements to these AAVs can optimize for specificity or completeness for the striatal  $TAC3^+$  type alone according to experimental goals and delivery strategy.

Together, our study reveals the complex landscape, biodistribution, and morphological variation of transcriptionally defined cell types in the marmoset brain. The persistent fingerprint of developmental origin present in neuronal gene expression underscores important roles for novel neurogenic niches, developmental patterning, and altered mechanisms guiding proliferation and cell migration in primate brain evolution that could be the subject of future study. These resources will enable further comparative studies of the evolution and development of primate brain cell types.

## MATERIALS AND METHODS

### Animals used for study

#### Marmosets

Marmosets were pair-housed in spacious holding rooms with environmental control of temperature (23° to 28°C), humidity (40 to 72%), and 12-hour light/dark cycle. Their cages were equipped with a variety of perches and enrichment devices, and they received regular health checks and behavioral assessment from Massachusetts Institute of Technology (MIT) Division of Comparative

Medicine (DCM) veterinary staff and researchers. All animal procedures were conducted with prior approval by the MIT Committee for Animal Care (CAC) and following veterinary guidelines.

#### Mice

Experimental mice were purchased from The Jackson Laboratory company and housed at the McLean hospital animal facility (three to five mice per cage) on a 12:12-hour light-dark cycle in a temperature-controlled colony room with unrestricted access to food and water. All procedures were conducted in accordance with policy guidelines set by the National Institutes of Health and were approved by the McLean Institutional Animal Care and Use Committee (IACUC).

### Tissue processing for single-nucleus sequencing and smFISH

#### Marmoset specimens for snRNA-seq

Marmoset experiments were approved by and in accordance with MIT CAC protocol number 051705020. Adult marmosets (2 to 14.5 years old, 12 individuals; table S1) were deeply sedated by intramuscular injection of ketamine (20 to 40 mg kg<sup>-1</sup>) or alfaxalone (5 to 10 mg kg<sup>-1</sup>), followed by intravenous injection of sodium pentobarbital (10 to 30 mg kg<sup>-1</sup>). When the pedal withdrawal reflex was eliminated and/or the respiratory rate was diminished, animals were transcardially perfused with ice-cold sucrose-Hepes buffer (4, 10). Whole brains were rapidly extracted into fresh buffer on ice. Sixteen 2-mm coronal blocking cuts were rapidly made using a custom-designed marmoset brain matrix. Slabs were transferred to a dish with ice-cold dissection buffer (4, 10). All regions were dissected using a marmoset atlas as reference (110), snap-frozen in liquid nitrogen or dry ice-cooled isopentane, and stored in individual microcentrifuge tubes at -80°C.

#### Marmoset specimens for snATAC-seq

Tissue from one marmoset (female, 2 years old; table S1) was used for both snRNA-seq and snATAC-seq (table S1). Fresh tissue was dissected from anterior striatum (including caudate and putamen) and used immediately for snATAC-seq.

#### Mouse specimens for amygdala snRNA-seq

Three adult (P80 to P90) male wild-type mice were deeply anesthetized with isoflurane and sacrificed by decapitation. Brains were quickly excised, washed in ice-cold sterile 0.1 M phosphate-buffered saline (PBS), and dissected onto an ice-cold glass surface. Amygdala nuclei were identified and isolated using "The Allen Mouse Brain Atlas" (<https://mouse.brain-map.org/static/atlas>; table S7) as a reference for anatomical landmarks. The basolateral amygdaloid nucleus was exposed by performing two coronal cuts using the borders of primary somatosensory cortex and primary visual cortex as landmarks. Dissected specimens were collected in 1.5-ml microcentrifuge tubes, snap-frozen on dry ice, and stored at -80°C until use.

#### Marmoset specimens for smFISH

Two marmosets were deeply sedated by intramuscular injection of alfaxalone (5 to 10 mg kg<sup>-1</sup>) (table S1), followed by intravenous overdose of sodium pentobarbital (10 to 30 mg kg<sup>-1</sup>). When the pedal withdrawal reflex was eliminated and/or the respiratory rate was diminished, animals were transcardially perfused with ice-cold saline. The brain was immediately removed, embedded in optimal cutting temperature (OCT) freezing medium, and flash-frozen in an isopropyl ethanol-dry ice bath. Samples were cut on a cryostat (Leica CM 1850) at a thickness of 16 μm, adhered to SuperFrost

Plus microscope slides (VWR, 48311-703), and stored at  $-80^{\circ}\text{C}$  until use. Portions of the brain that were not cut were recoated in OCT and stored again for future use. Samples were immediately fixed in 4% paraformaldehyde (PFA) and stained on the slide according to the Molecular Instruments HCR generic sample in solution RNA-FISH protocol (Molecular Instruments, <https://files.molecularinstruments.com/MI-Protocol-RNAFISH-GenericSolution-Rev7.pdf>; table S7) or the Advanced Cell Diagnostics RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (ACD; 323100, [https://acdbio.com/sites/default/files/UM%20323100%20Multiplex%20Fluorescent%20v2\\_RevB.pdf](https://acdbio.com/sites/default/files/UM%20323100%20Multiplex%20Fluorescent%20v2_RevB.pdf)) protocol (table S7).

## snRNA-seq, library preparation, sequencing

### 10x RNA-seq

Unsorted single-nucleus suspensions from frozen marmoset and mouse samples were generated as in (10). Gel-Bead-in-emulsion (GEM) generation and library preparation followed the manufacturer's protocol (10X Chromium single-cell 3' v.3, protocol number #CG000183\_ChromiumSingleCell3'\_v3\_UG\_Rev-A). Raw sequencing reads were aligned to the National Center for Biotechnology Information (NCBI) CJ1700 reference (marmoset) or GRCm38 (mouse). Reads that mapped to exons or introns were assigned to annotated genes. Libraries were sequenced to a median read depth of 8 reads per unique molecular identifier (UMI; or transcript), obtaining a median 7262 UMIs per cell.

### RNA sequencing data processing, curation, clustering

Processing and alignment steps follow those outlined in the Drop-seq alignment workflow (table S7). Raw BCL files were processed using IlluminaBasecallsToSam, and reads with a barcode quality score below Q10 were discarded. Cell barcodes (CBCs) were filtered using the 10X CBC whitelist, followed by template switching oligo (TSO) and poly(A) trimming. Reads were aligned using STAR and tagged with their gene mapping [exonic, intronic, or untranslated region (UTR)] and function (strand). The reads were then processed through GATK BQSR and tabulated in a digital gene expression (DGE) matrix containing all CBCs with at least 20 transcripts aligning to coding, UTR, or intronic regions. Cell selection was performed based on CellBender remove-background non-empties (111), % intronic (% of a CBC's reads that are intronic), and number of UMIs for a CBC. A new filtered DGE containing only these selected CBCs was then generated. Finally, a gene-metagene DGE was created by merging the selected CBCs DGE with a metagene DGE (made by identifying reads from the selected CBCs that have a secondary alignment mapping to a different gene than its primary alignment).

Cell type classification models were trained using our annotations and scPred R package version 1.9.2 (35). Detection of cell-cell doublets was performed using a two-step process based on the R package DoubletFinder (112). DoubletFinder implements a nearest-neighbors approach to doublet detection. First, artificial doublets are simulated from input single-cell data and are coclustered with true libraries. True doublet libraries are identified by their relative fraction of artificial doublet nearest-neighbors in gene expression space. In our workflow, this process is run twice, once with high stringency to identify and remove clear doublets, and again with a lower threshold to identify remaining subtler doublet libraries. We used the following parameters for round 1:  $\text{PN} = 0.4$ ,  $\text{PK} = 10 \wedge \text{seq}(-4, -1, \text{length.out} = 50)$ ,  $\text{NUM\_PCS} = 5$ .

For the second round:  $\text{PN} = 0.45$ ,  $\text{NUM\_PCS} = 10$ ,  $\text{PK} = 10 \wedge \text{seq}(-4, -1, \text{length.out} = 50)$ . Because doublet libraries are initially categorized as true libraries in the nearest-neighbor search, we find that this two-step process improves the sensitivity and accuracy of doublet detection.

Clustering was performed using independent component analysis (ICA; R package fastICA) dimensionality reduction followed by Louvain. Cells assigned to one of the major glial types (oligodendrocyte lineage, astrocytes, vascular/endothelia, microglia/macrophages) by scPred were collected across all brain regions and clustered together. Neurons from most telencephalic structures (neocortex, hippocampus, striatum, amygdala) confidently assigned to the categories "GABAergic" and "glutamatergic" and so were clustered separately by neurotransmitter usage for each brain structure. Striatal neuron categories were "SPN" and "GABAergic interneuron." Neurons in nontelencephalic brain structures were clustered separately by brain structure. All clusterings were performed in two stages: First-round clustering was based on the top 60 ICs and a default resolution (res) of 0.1, nearest neighbors (nn) = 25. Following the process outlined in (4), each was manually reviewed for skew and kurtosis of gene loadings on factors and cells to identify ICs that loaded on outliers, doublets, or artifactual signals. These were discarded, and reclustering was performed on the remaining ICs. Each resulting cluster was then subjected to second-round clustering, during which ICs were again curated. Second-round clustering explored a range of parameters: nn = 10, 20, 30; res = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0. Final parameter values were chosen to optimize concordance, when possible, between the final number of clusters and the number of included ICs such that each cluster was defined by one primary IC. Metacells for each cluster are generated by summing transcript counts for each cell across all cells in the cluster, normalizing by total number of transcripts, and scaling to counts per 100,000. Hierarchical clustering of metacells was performed in two ways: (i) on the top 100 PCA components (HCA PCA) (using the HCPC package), and (ii) using Pearson correlation as the distance function (HCA Corr). Trees were visualized using the ggtree package suite (113). See associated code for plotting functions (table S7).

### Identification of rDEGs

We computed rDEGs for neocortical glutamatergic neurons, GABAergic neurons, astrocytes, and oligodendrocyte lineage types. We used individual cell level cluster assignments from the initial neocortical clustering (which contained all regions together) to create per-region metacells for each cluster. Normalized metacells were  $\log_{10}$ -transformed, and pairwise differences across regions within the same cluster (cell type) were examined. Genes with >3-fold difference in the same cluster between two regions were considered rDEGs. For this analysis, we omitted from comparisons any region-cluster metacell generated from fewer than 50 cells, but retained comparisons between regions that had >50 cells in that cluster. Genes that were consistently rDEGs in at least three individuals for each cluster-region pair are reported in table S3.

### Ancestral state reconstruction

Ancestral state reconstruction (ASR) is a method to infer hidden ancestral traits from extant observations. For example, given a phylogenetic tree of species and genomic sequences thought to be homologous across those species, the reconstruction takes into

account branch lengths to reconstruct the most likely ancestral sequence. We applied a maximum likelihood–based ASR approach (R package: fastAnc) to the scaled, normalized metacells of cell types and the dendrogram of their similarity to produce estimates of expression of each gene at the internal nodes of the tree. These enabled comparisons of leaf nodes to internal nodes as well as internal nodes to each other. For example, compared to the parent node of amygdala, basal forebrain, and hypothalamic SPN-like GABAergic projection neurons, the reconstructed parent node of striatal SPNs had higher expression of known markers of striatal projection neurons such as *DACH1*. We used these reconstructions of internal node gene expression to compare major clades of the tree, and used a chi-square test to determine whether TFs were overrepresented among differentially expressed genes (threshold: threefold difference) between pairs of internal nodes.

### Spatial smFISH experiments

All probes are listed in table S4. All smFISH validation experiments were carried out on distinct biological replicates from those used for snRNA-seq or single-cell ATAC-seq experiments.

#### smFISH tissue processing and quantification

Two marmosets (Cj 18-134 and 19-212) were euthanized and perfused with PBS (table S1). The brain was removed, embedded rapidly in OCT, and stored in the  $-80^{\circ}\text{C}$  freezer. Tissue was then cut to  $16\ \mu\text{m}$  on a cryostat and stored in the  $-80^{\circ}\text{C}$  freezer until needed. In situ hybridization was performed for genes of interest (see table S4) with HCR or ACD antisense probes and incubated with TrueBlack Lipofuscin Autofluorescence Quencher (Biotium, 23007) for 10 s at room temperature to eliminate confounding lipofuscin autofluorescence present in the tissue. Samples were then coverslipped with ProLong Diamond Antifade Mountant (Invitrogen, P36970). Z-stack serial images were taken through the whole depth across striatum, hypothalamus, and basal forebrain regions, and several regions of neocortex, on the TissueGnostic TissueFAXS SL slide-scanning, spinning disk confocal microscope (Hamamatsu Orca Flash 4.0 v3) using a  $20\times/0.8$  NA (numerical aperture) air objective for ACD stains or a  $40\times/1.2$  NA water-immersion objective for HCR stains.

Series images were segmented using StrataQuest, a software package from TissueGnostics, which enables the quantification of signals within segmented images (similar to CellProfiler). Nuclei objects were generated using the DAPI channel, and artifacts were removed based on size and intensity. Exclusion regions of interest (ROIs) were manually drawn to avoid areas of white matter, large artifacts, and autofluorescence before computing intensity and other statistical and morphological measurements (20 parameters) for each channel. Specifically, 50 cells were hand-labeled as positive or negative for the markers of interest to identify the appropriate threshold for feature selection using the parameters that best discriminated this binary. Parameters included mean intensity, maximum intensity, SD of intensity, range of intensity, and equivalent diameter. These filters were then applied to the unlabeled data to identify positive cells.

Segmented cells were further analyzed using in-house code (<https://doi.org/10.5281/zenodo.8364941>; table S7). Spatial locations of the cells were visualized by plotting the  $x$ - $y$  coordinates associated with each nucleus. These were then binned into 2D histograms across the  $x$  and  $y$  axis (corresponding to the rostrocaudal plane and to the dorsoventral plane, respectively). A size of 100

bins was chosen for the first (medial-most) slice of a series across the  $x$  axis, and the calculated bin size was then used across the  $y$  axis of the first slice and across other slices in the series. Positive events for a gene in a given bin were either normalized to the number of detected nuclei in that bin (DAPI) and plotted in 2D as a relative heatmap or simply plotted as a density heatmap without being normalized to DAPI. Whole slice normalizations across the mediolateral axis (all positive events for a gene in a given slice relative to DAPI) were plotted as bar graphs. DAPI and marker of interest counts are available in data S1 and S2.

### Neocortical areal proportions

In addition to the mediolateral subdivisions of marmoset neocortex, we further parcellated the neocortex into subregions referencing the Brain/MINDS 3D Digital Marmoset Atlas (table S7) (76). Slice selection was based upon visual recognition of prominent landmarks (white matter, striatal boundary, as well as DAPI nuclei staining) within our tissue and matched with the nearest atlas slice. ROIs for each parcellated region were created using the StrataQuest software in the anterior-posterior and dorsal-ventral axes (Fig. 6F). The corresponding feature selection that was previously set within the full neocortical sections was carried across gene and slice. The parameters used were mean intensity, maximum intensity, SD of intensity, range of intensity, and equivalent diameter. These filters were then applied to the unlabeled data to identify positive cells within each individual ROI. The parcellated neocortical areas were then analyzed using in-house code (<https://doi.org/10.5281/zenodo.8364941>; table S7). As above, positive events for a gene in a given ROI were normalized to DAPI and plotted in a (stacked) bar graph. Relative percentages between genes were calculated and plotted stacked.

### Morphology and smFISH experiments

AAV9-hDlx5/6-GFP-fGFP virus was generated as in (29). Virus was systemically intravenously injected ( $400$  to  $700\ \mu\text{l}$  at  $1.7 \times 10^{10}$  to  $2.4 \times 10^{10}$  titers) in five marmosets. The virus was allowed to incubate for approximately 2 months. After systemic intravenous injection with AAV9-hDlx5/6-GFP-fGFP, marmosets were euthanized and perfused with saline followed by 4% PFA. Brains were removed, and  $120\text{-}\mu\text{m}$  sections were cut on a vibratome into PBS-azide and stored at  $4^{\circ}\text{C}$  or moved into 70% ethanol for storage at  $-20^{\circ}\text{C}$ . The 70% ethanol storage prevents RNA degradation at this temperature without considerable tissue shrinkage for short storage times. Because of laboratory shutdowns during the pandemic, sections from two marmosets were stored in 70% ethanol for approximately 4 months. These samples exhibited considerable shrinkage, measured by DAPI-stained nuclei diameters (tables S5 and S6); therefore, we only compared morphology parameters within-donor. Sections were taken as needed, and in situ hybridization was performed with HCR antisense probes, following the generic sample in solution HCR protocol with a twofold increase in concentration of probe to hybridization buffer (Molecular Instruments, <https://files.molecularinstruments.com/MI-Protocol-RNAFISH-GenericSolution-Rev8.pdf>; table S7), for markers of interest that corresponded with RNA-seq defined clusters. The sections were then stained with anti-GFP antibody and a secondary antibody (table S4) to amplify the endogenous GFP signal (<https://www.protocols.io/view/marmoset-nhp-free-floating-anti-gfp-antibody-stain-3byl47nb2lo5/v1>; table S7). Sections were incubated in

TrueBlack (Biotium, 23007) for 3 to 5 min to mask confounding lipofuscin autofluorescence throughout the section. Sections were then mounted onto a slide and coverslipped with ProLong Diamond Antifade Mountant (Invitrogen, P36970) for imaging.

#### Imaging for morphology

Sections prepared for morphology were imaged on a Nikon Ti Eclipse inverted microscope with an Andor CSU-W1 confocal spinning disc unit and Andor DU-888 EMCCD using a 40×/1.15 NA water-immersion objective, and later on a TissueGnostic TissueFAXS SL slide-scanning, spinning disk confocal microscope (with Hamamatsu Orca Flash 4.0 v3) using a 40×/1.2 NA water-immersion objective. With the TissueFAXS, overview images were taken to select GFP<sup>+</sup> cells for imaging at 40× and to highlight the exact location of the cell. GFP<sup>+</sup> cells were imaged for stained markers of interest. Selected sections were imaged on an upright confocal laser scanning microscope (Olympus Fluoview FV3000) using a 40×/0.95 NA air objective or a 60×/1.50 NA oil-immersion objective and cooled Gallium Arsenide Phosphide Photomultiplier Tubes (GaAsP PMTs).

#### Morphological reconstruction and feature quantification

##### - Imaris

Without preprocessing the confocal images, three-dimensional (3D) reconstruction and surface rendering of striatal and neocortical interneurons were performed using Imarisx64 9.9 software (Oxford Instruments) based on GFP<sup>+</sup> signal. Surface-rendered images were used to determine the soma diameter, total volume, and total surface area for each z-stack image. 3D-skeleton diagrams (Figs. 8, A to C, and 9, H and J), corresponding to each surface-rendered image, were generated using the Filament Tracing wizard in Imaris and then pseudo-colored in Adobe Illustrator. The total number of primary dendritic branches, dendritic branch points, area, volume, and length of the 3D-skeleton diagrams was calculated using the AutoPath (no loops) algorithm in the filament tracing wizard in Imaris. The total number of primary dendritic branches for each cell is defined by the number of dendrite branches in the filament trace one distance value away from the soma. The distance value is calculated automatically by the AutoPath (no loops) algorithm based on the diameter of the soma and the diameter of the thinnest cellular process. All data were exported to CSV, and data collected from exemplar cells (Figs. 8, A and B, and 9G) were reported in table S6.

A separate dataset containing GFP<sup>+</sup> interneurons from one animal (Cj 17-154; table S1) was first blinded using a custom Python script and then reconstructed with Imaris (Fig. 8, C and D). All results are presented as mean ± SEM. Comparisons between soma diameter, surface area, surface volume, and the number of primary branches were carried out using unpaired *t* tests. Unpaired Mann-Whitney *U* tests were used to statistically compare the total length, area, volume, and total number of dendritic branch point measurements. Statistically significant analyses were denoted as follows: \**P* < 0.05; \*\**P* < 0.01.

#### Morphological reconstruction and feature quantification

##### - NeuTube

Automatic or semi-automatic tracing algorithms are challenged by some data, perhaps due to the low signal-to-noise ratio (SNR) of a given image. To overcome this, we manually reconstructed GFP<sup>+</sup> neurons via Neutube (114) tracing software (table S5). With the software, we (i) create 3D volume rendering of the GFP-AAV marmoset neuron, (ii) use the signal transfer function (e.g., histogram

equalization) for overall intensity and opacity values to optimize the SNR by manually examining the clearest visualization of the dendrites, (iii) build the neuron skeleton over the 3D volume by tracing the processes, (iv) scan through the 3D volume to make sure no parts of the neuron are missed, (v) check the raw 2D images to see if any of the branches were not presented well in 3D due to volume rendering artifacts, and (vi) label axon, dendrites, and soma parts of the skeleton model. Reconstructed neurons were saved as SWC format.

#### snATAC-seq, library preparation, sequencing

snATAC-seq (10x Genomics Single Cell ATAC v1) was conducted on fresh marmoset tissue (1 female, 2 years old) dissected from anterior striatum. Nuclei suspensions were generated following the 10x Genomics recommended protocol (CG000212 Rev. B). Library preparation followed 10x Genomics Single Cell ATAC v1 Guide (CG000168 Rev. D). Library was sequenced on an Illumina NovaSeq (RRID:SCR\_016387) using 100-bp paired-end reads to a median per-cell fragment depth of 25,472. Alignment and fragment counting was conducted using Cell Ranger (RRID:SCR\_01734), aligned to marmoset genome cj1700 ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_009663435.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_009663435.1/); GCA\_009663445.2). snATAC-seq data were integrated with marmoset striatal snRNA-seq data from two independent animals (bi005, bi006) using Signac (RRID:SCR\_021158) with the following parameters: integration method: cca; weight.reduction = lsi. Differentially accessible peaks for each cluster were calculated with a threefold change cutoff and the minimum fraction of expressed cells (in target cluster) = 0.2. Enhancer locations were as follows (CJ1700 coordinates): Eh14 = chr1-122696345-122696560 (215 bp); Eh15 = chr17-32919427-32919564 (137 bp); Eh16 = chr20-38770509-38770619 (110 bp); Eh17 = chr9-60258694-60258788 (94 bp).

#### TAC3-AAV viral design and production

AAVs were produced in accordance with the protocol (115). Construct DNA (5.7 µg/150-mm dish) was transfected with pAAV9 capsid plasmid (22.8 µg/150-mm dish) and pAddDeltaF6 helper plasmid (11.45 µg/150-mm dish) using polyethylenimine 25K MW (Polysciences, 23966-1). Collection of cells and media for AAV harvesting began 72 hours following transfection before iodixanol gradient ultracentrifugation purification using a Type 70 Ti Fixed-Angle Titanium Rotor (Beckman-Coulter, 337922). Titer was calculated using digital droplet polymerase chain reaction (ddPCR) according to the protocol "ddPCR Titration of AAV Vectors" (<https://www.addgene.org/protocols/aav-ddpcr-titration/>) using the QX200 AutoDG Droplet Digital PCR System (BioRad, 1864100). gBlock fragments containing the tandem TAC3 enhancers (E14E15E16E17) were synthesized by Integrated DNA Technologies (IDT). The enhancers were cloned into AAV plasmid vector backbones containing a minimal cytomegalovirus (CMV) promoter, the reporter, and a barcode unique to each enhancer sequence using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB-E5520S), following the standard protocol. Two microliters of Gibson assembly product was used to transform 50 µl of homemade Stb13 cells following a standard transformation protocol. Mini-preparation of plasmids was carried out using the NucleoSpin mini kit (Macherey-Nagel). Positive clones were identified by restriction enzyme digestion and sequencing. The positive clones were grown in 300 ml of LB cultures, and the plasmids were

extracted using the NucleoBond Xtra Midi EF kit (Macherey-Nagel). The plasmids were sent out for sequencing again before being packaged into AAVs.

### TAC3-AAV local injection procedure

#### Structural MRI scanning

In preparation for MR imaging under anesthesia, an animal (two marmosets, Cj 19-207 and 17-B111; table S1) checked for robust health was fasted overnight with ad libitum access to water. Before scanning, the animal was sedated with alfaxalone (5 to 10 mg/kg) or a combination of alfaxalone (4 to 8 mg/kg) and ketamine (5 to 10 mg/kg) given intramuscularly. It was then transferred to a custom-designed MRI cradle equipped with a 3D-printed nose cone for delivering anesthetic gases. The animal's head was securely held to minimize motion artifacts and stereotaxically align the rostral-caudal axis with the MR scanner bore using contrast-filled MR-compatible ear bars, an adjustable palette bar, and eye bars. During scanning, anesthesia was maintained at 1 to 2% isoflurane in a mixture of oxygen. A temperature-controlled water circulating heat pad was used for thermal support. Heart rate and blood oxygen saturation levels were continuously monitored and recorded every 10 min with an MR-compatible monitoring system (Nonin, MN). MR scans consisted of high-resolution 3D T1 mapping using a magnetization-prepared rapid gradient-echo (MPRAGE) sequence (116) with repetition time (TR) = 6000 ms, TE (echo time) = 4.6 ms, flip angle (FA) = 12°, FOV = 80, matrix size = 256 × 256. A high-resolution T2-weighted anatomical scan was also obtained using a RARE pulse sequence with effective TE = 35.5 ms, TR = 2500 ms, RARE factor = 8, spatial resolution = 156 μm × 156 μm × 0.5 mm, matrix size = 256 × 256, RARE factor = 4, FOV = 80 mm.

#### Virus injection surgery

Following sedation with alfaxalone (5 to 10 mg/kg), the animal was intubated and maintained at 1.5 to 2% isoflurane in a mixture of oxygen throughout the rest of the procedure. An intravenous catheter was placed in the saphenous vein for infusion of fluids and drugs during the surgery and the recovery period. Heart rate, SpO<sub>2</sub>, electrocardiography, end-tidal CO<sub>2</sub>, and rectal temperature were continuously monitored and logged every 5 min throughout. Once the animal acquired a stable plane of anesthesia, it was placed in a stereotaxic apparatus (Narishige, Japan). A thin layer of sterile eye lubricant was applied to protect against corneal drying. Animal was also provided with a single bolus dose of intravenous dexmedetomidine (0.4 mg/kg) to guard against brain swelling. The scalp and fascia were removed in layers via blunt dissection to expose the injection sites. Using MRI atlas guidance, bilateral craniotomies were performed to expose the medial anterior caudate and ventral posterior putamen region. The tandem enhancer virus AAV9-tandemE-TAC3-EGFP was injected in left caudate (high titer, 10<sup>13</sup>; low volume, 0.5 μl) and left putamen (low titer, 10<sup>11</sup>; high volume, 1.5 μl). Each injection was delivered with a controlled syringe pump (100 nl/min). After each injection, the needle was kept in place for 10 min and then slowly retracted over 1 to 2 min to minimize injection backflow. Following withdrawal, the cranial openings were covered with bonewax, and fascia and scalp were sutured in layer-wise manner with 3-0 vicryl sutures. The suture site was cleaned with warm sterile saline and covered with hypafix. For recovery, the animal was transferred in an incubator and monitored closely until it was able to move effortlessly and accepted treats before returning to its home cage. Post-op medications were provided under

veterinary advice for pain and inflammation control until full recovery. The animals were euthanized 6 to 10 weeks after viral injection and were perfused transcardially with 4% PFA. The brain was extracted for histology, smFISH, and confocal microscopy.

### TAC3-AAV-BI103 systemic injection procedure

One 2.5-year-old female marmoset (table S1) was sedated with alfaxalone (5 to 10 mg/kg) given intramuscularly. A catheter was placed into the tail vein, and 200 μl of purified viral particles was injected at a dose of 7.85 × 10<sup>13</sup> viral genomes (vg)/kg, followed by injection of 1.5 ml of saline to flush the line. The animal was euthanized 4 weeks after viral injection and was perfused transcardially with 4% PFA. The brain was extracted for histology, smFISH, and confocal microscopy.

### TAC3-AAV histology

For AAV9-tandemE-TAC3-EGFP experiments, coronal sections around the injection sites were cut at 100 μm on a vibratome (Leica, VT1000S) and then stored in PBS-NaN<sub>3</sub> at 4°C until use. For AAV-BI103-tandemE-TAC3-EGFP experiments, coronal and sagittal sections were processed. Sections were stained for TAC3 (Molecular Instruments, PRC843; table S4) using the HCR v.3.0 protocol (Molecular Instruments) described above and a rabbit anti-GFP antibody (Invitrogen, A11122; table S4) with subsequent secondary goat anti-rabbit conjugated with AF488 (Thermo Fisher Scientific, A-11008). Stained sections were then incubated for 1 to 3 min in TrueBlack to eliminate confounding lipofuscin autofluorescence. The striatum was imaged for each section using the Tissue-FAXS SL with a 40×/1.2 NA water-immersion lens through the whole thickness of the tissue.

### Supplementary Materials

This PDF file includes:

Figs. S1 to S9

Legends for tables S1 to S7

Legends for data S1 and S2

Other Supplementary Material for this

manuscript includes the following:

Tables S1 to S7

Data S1 and S2

### REFERENCES AND NOTES

- Z. Yao, C. T. J. van Velthoven, M. Kunst, M. Zhang, D. McMillen, C. Lee, W. Jung, J. Goldy, A. Abdelhak, P. Baker, E. Barkan, D. Bertagnolli, J. Campos, D. Carey, T. Casper, A. B. Chakka, R. Chakraborty, S. Chavan, M. Chen, M. Clark, J. Close, K. Crichton, S. Daniel, T. Dolbear, L. Ellingwood, J. Gee, A. Glandon, J. Gloe, J. Gould, J. Gray, N. Guilford, J. Guzman, D. Hirschstein, W. Ho, K. Jin, M. Kroll, K. Lathia, A. Leon, B. Long, Z. Maltzer, N. Martin, R. McCue, E. Meyerdierks, T. N. Nguyen, T. Pham, C. Rimorin, A. Ruiz, N. Shapovalova, C. Slaughterbeck, J. Sulc, M. Tieu, A. Torkelson, H. Tung, N. V. Cuevas, K. Wadhvani, K. Ward, B. Levi, C. Farrell, C. L. Thompson, S. Mufti, C. M. Pagan, L. Kruse, N. Dee, S. M. Sunkin, L. Esposito, M. J. Hawrylycz, J. Waters, L. Ng, K. A. Smith, B. Tasic, X. Zhuang, H. Zeng, A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *bioRxiv* 2023.03.06.531121 [Preprint]. 2023 March 6. <https://doi.org/10.1101/2023.03.06.531121>.
- K. Siletti, R. Hodge, A. M. Albiach, L. Hu, K. W. Lee, P. Lönnerberg, T. Bakken, S.-L. Ding, M. Clark, T. Casper, N. Dee, J. Gloe, C. Dirk Keene, J. Nyhus, H. Tung, A. M. Yanny, E. Arenas, E. S. Lein, S. Linnarsson, Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
- J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac, X. Zhuang, Molecular, spatial, and functional



- single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaa05324–80 (2018).
- A. Saunders, E. Z. Mocosko, A. Wysoker, M. Goldman, F. M. Krienen, H. de Rivera, E. Bien, M. Baum, L. Bortolin, S. Wang, A. Goeva, J. Nemesh, N. Kamitaki, S. Brumbaugh, D. Kulp, S. A. McCarroll, Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018).
  - A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E. Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J. Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
  - K. L. Chiou, A. R. DeCasien, K. P. Rees, C. H. Spurrell, A. A. Gogate, H. A. Pliner, S. Tremblay, A. Mercer, C. J. Whalen, J. E. N.-D. Valle, M. C. Janiak, S. E. B. Surratt, O. González, N. R. Compo, M. K. Stock, A. V. Ruiz-Lambides, M. I. Martínez, C. B. R. Unit, M. A. Wilson, A. D. Melin, S. C. Antón, C. S. Walker, J. Sallet, J. M. Newbern, L. M. Starita, J. Shendure, J. P. Higham, L. J. N. Brent, M. J. Montague, M. L. Platt, N. Snyder-Mackler, Multiregion transcriptomic profiling of the primate brain reveals signatures of aging and the social environment. *Nat. Neurosci.* **25**, 1714–1723 (2022).
  - D. Hain, T. Gallego-Flores, M. Klinkmann, A. Macias, E. Ciirdaeva, A. Arends, C. Thum, G. Tushev, F. Kretschmer, M. A. Tosches, G. Laurent, Molecular diversity and evolution of neuron types in the amniote brain. *Science* **377**, eabp8202 (2022).
  - J. Woych, A. Ortega Gurrola, A. Deryckere, E. C. B. Jaeger, E. Gumnit, G. Merello, J. Gu, A. Joven Arous, N. D. Leigh, M. Yun, A. Simon, M. A. Tosches, Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science* **377**, eabp9186 (2022).
  - P. Theodoni, P. Majka, D. H. Reser, D. K. Wójcik, M. G. P. Rosa, X.-J. Wang, Structural attributes and principles of the neocortical connectome in the marmoset monkey. *Cereb. Cortex* **32**, 15–28 (2021).
  - F. M. Krienen, M. Goldman, Q. Zhang, R. C. H. del Rosario, M. Florio, R. Machold, A. Saunders, K. Levandowski, H. Zaniewski, B. Schuman, C. Wu, A. Lutservitz, C. D. Mullally, N. Reed, E. Bien, L. Bortolin, M. Fernandez-Otero, J. D. Lin, A. Wysoker, J. Nemesh, D. Kulp, M. Burns, V. Tkachev, R. Smith, C. A. Walsh, J. Dimidschstein, B. Rudy, L. S. Kean, S. Berretta, G. Fishell, G. Feng, S. A. McCarroll, Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).
  - X. Tian, Y. Chen, P. Majka, D. Szczupak, Y. S. Perl, C. C.-C. Yen, C. Tong, F. Feng, H. Jiang, D. Glen, G. Deco, M. G. P. Rosa, A. C. Silva, Z. Liang, C. Liu, An integrated resource for functional and structural connectivity of the marmoset brain. *Nat. Commun.* **13**, 7416 (2022).
  - R. L. Buckner, D. S. Margulies, Macroscale cortical organization and a default-like apex transmodal network in the marmoset monkey. *Nat. Commun.* **10**, 1976 (2019).
  - D. C. Van Essen, C. J. Donahue, T. S. Coalson, H. Kennedy, T. Hayashi, M. F. Glasser, Cerebral cortical folding, parcellation, and connectivity in humans, nonhuman primates, and mice. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26173–26180 (2019).
  - I. Kelava, I. Reillo, A. Y. Murayama, A. T. Kalinka, D. Stenzel, P. Tomancak, F. Matsuzaki, C. Lebrand, E. Sasaki, J. C. Schwamborn, H. Okano, W. B. Huttner, V. Borrell, Abundant occurrence of basal radial glia in the subventricular zone of embryonic neocortex of a lissencephalic primate, the common marmoset *Callithrix jacchus*. *Cereb. Cortex* **22**, 469–481 (2012).
  - F. A. C. Azevedo, L. R. B. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. L. Ferretti, R. E. P. Leite, W. Jacob Filho, R. Lent, S. Herculano-Houzel, Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513**, 532–541 (2009).
  - S. Ma, M. Skarica, Q. Li, C. Xu, R. D. Risgaard, A. T. N. Tebbenkamp, X. Mato-Blanco, R. Kovner, Ž. Kršnik, X. de Martin, V. Luria, X. Martí-Pérez, D. Liang, A. Karger, D. K. Schmidt, Z. Gomez-Sanchez, C. Qi, K. T. Gobeske, S. Pochareddy, A. Debnath, C. J. Hottman, J. Spurrier, L. Teo, A. G. Boghdadi, J. Homman-Ludie, J. J. Ely, E. W. Daadi, D. Mi, M. Daadi, O. Marin, P. R. Hof, M.-R. Rasin, J. Bourne, C. C. Sherwood, G. Santpere, M. J. Girgenti, S. M. Strittmatter, A. M. M. Sousa, N. Sestan, Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).
  - T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D. Bertagnoli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dombrow, D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R. Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D. Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu, J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed, C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M. Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz, C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte, S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M. Hawrylycz, B. Tasic, E. Z. Mocosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R. Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
  - J.-P. Lin, H. M. Kelly, Y. Song, R. Kawaguchi, D. H. Geschwind, S. Jacobson, D. S. Reich, Transcriptomic architecture of nuclei in the marmoset CNS. *Nat. Commun.* **13**, 5531 (2022).
  - R. C. Bandler, I. Vitali, R. N. Delgado, M. C. Ho, E. Dvoretzskova, J. S. Ibarra Molinas, P. W. Frazel, M. Mohammadkhani, R. Machold, S. Maedler, S. A. Liddelow, T. J. Nowakowski, G. Fishell, C. Mayer, Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2021).
  - R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, Z. Yao, J. Eggermont, T. Holt, B. P. Levi, S. I. Shehata, B. Aevermann, A. Beller, D. Bertagnoli, K. Brouner, T. Casper, C. Cobbs, R. Dalley, N. Dee, S.-L. Ding, R. G. Ellenbogen, O. Fong, E. Garren, J. Goldy, R. P. Gwinn, D. Hirschstein, C. D. Keene, M. Keshk, A. L. Ko, K. Lathia, A. Mahfouz, Z. Maltzer, M. McGraw, T. N. Nguyen, J. Nyhus, J. G. Ojemann, A. Oldre, S. Parry, S. Reynolds, C. Rimorin, N. V. Shapovalova, S. Somasundaram, A. Szafer, E. R. Thomsen, M. Tieu, G. Quon, R. H. Scheuermann, R. Yuste, S. M. Sunkin, B. Lelieveldt, D. Feng, L. Ng, A. Bernard, M. Hawrylycz, J. W. Phillips, B. Tasic, H. Zeng, A. R. Jones, C. Koch, E. S. Lein, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
  - N. L. Jorstad, J. Close, N. Johansen, A. M. Yanny, E. R. Barkan, K. J. Travaglini, D. Bertagnoli, J. Campos, T. Casper, K. Crichton, N. Dee, S.-L. Ding, E. Gelfand, J. Goldy, D. Hirschstein, M. Kroll, M. Kunst, K. Lathia, B. Long, N. Martin, D. McMillen, T. Pham, C. Rimorin, A. Ruiz, N. Shapovalova, S. Shehata, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, K. Ward, E. M. Callaway, P. R. Hof, C. Dirk Keene, B. P. Levi, S. Linnarsson, P. P. Mitra, K. Smith, R. D. Hodge, T. E. Bakken, E. S. Lein, Transcriptomic cytoarchitecture reveals principles of human neocortex organization. *Science* **382**, ead6f812 (2023).
  - Y. Kim, G. R. Yang, K. Pradhan, K. U. Venkataraju, M. Bota, L. C. G. del Molino, G. Fitzgerald, K. Ram, M. He, J. M. Levine, P. Mitra, Z. J. Huang, X.-J. Wang, P. Osten, Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171**, 456–469.e22 (2017).
  - S. M. Ford, Callitrichids as phyletic dwarfs, and the place of the callitrichidae in platyrrhini. *Primates* **21**, 31–43 (1980).
  - J. Hill, T. Inder, J. Neil, D. Dierker, J. Harwell, D. Van Essen, Similar patterns of cortical expansion during human development and evolution. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13135–13140 (2010).
  - B. Lee, R. Dalley, J. A. Miller, T. Chartrand, J. Close, R. Mann, A. Mukora, L. Ng, L. Alfiler, K. Baker, D. Bertagnoli, K. Brouner, T. Casper, E. Csajbok, N. Dee, N. Donadio, S. L. W. Driessens, T. Egdorf, R. Enstrom, A. A. Galakhova, A. Gary, E. Gelfand, J. Goldy, K. Hadley, T. S. Heistek, D. Hill, N. Johansen, N. Jorstad, L. Kim, A. K. Kocsis, L. Kruse, M. Kunst, G. Leon, B. Long, M. Mallory, M. Maxwell, M. McGraw, D. McMillen, E. J. Melief, G. Molnar, M. T. Mortrud, D. Newman, J. Nyhus, X. Opitz-Araya, T. Pham, A. Pom, L. Potekhina, R. Rajanbabu, A. Ruiz, S. M. Sunkin, I. Szots, N. Taskin, B. Thyagarajan, M. Tieu, J. Trinh, S. Vargas, D. Vumbaco, F. Waleboer, N. Weed, G. Williams, J. Wilson, S. Yao, T. Zhou, P. Barzo, T. Bakken, C. Cobbs, R. G. Ellenbogen, L. Esposito, M. Ferreira, N. W. Gouwens, B. Grannan, R. P. Gwinn, J. S. Hauptman, R. Hodge, T. Jarsky, C. Dirk Keene, A. L. Ko, B. Levi, J. G. Ojemann, A. Patel, J. Ruzevick, D. L. Silbergeld, K. Smith, J. Waters, H. Zeng, J. Berg, N. A. Giourounova, B. Kalmbach, C. P. J. de Kock, H. D. Mansvelde, S. A. Sorensen, G. Tamas, E. S. Lein, J. T. Ting, Signature morpho-electric properties of diverse GABAergic interneurons in the human neocortex. *Science* **382**, ead6f484 (2023).
  - N. W. Gouwens, S. A. Sorensen, F. Baftizadeh, A. Budzillo, B. R. Lee, T. Jarsky, L. Alfiler, K. Baker, E. Barkan, K. Berry, D. Bertagnoli, K. Bickley, J. Bomben, T. Braun, K. Brouner, T. Casper, K. Crichton, T. L. Daigle, R. Dalley, R. A. de Frates, N. Dee, T. Desta, S. D. Lee, N. Dotson, T. Egdorf, L. Ellingwood, R. Enstrom, L. Esposito, C. Farrell, D. Feng, O. Fong, R. Gala, C. Gamlin, A. Gary, A. Glandon, J. Goldy, M. Gorham, L. Graybuck, H. Gu, K. Hadley, M. J. Hawrylycz, A. M. Henry, D. Hill, M. Hupp, S. Kebede, T. K. Kim, L. Kim, M. Kroll, C. Lee, K. E. Link, M. Mallory, R. Mann, M. Maxwell, M. McGraw, D. McMillen, A. Mukora, L. Ng, L. Ng, K. Ngo, P. R. Nicovich, A. Oldre, D. Park, H. Peng, O. Penn, T. Pham, A. Pom, Z. Popović, L. Potekhina, R. Rajanbabu, S. Ransford, D. Reid, C. Rimorin, M. Robertson, K. Ronellenfitch, A. Ruiz, D. Sandman, K. Smith, J. Sulc, S. M. Sunkin, A. Szafer, M. Tieu, A. Torkelson, J. Trinh, H. Tung, W. Wakeman, K. Ward, G. Williams, Z. Zhou, J. T. Ting, A. Arkhipov, U. Sümbül, E. S. Lein, C. Koch, Z. Yao, B. Tasic, J. Berg, G. J. Murphy, H. Zeng, Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 935–953.e19 (2020).
  - J. DeFelipe, P. L. López-Cruz, R. Benavides-Piccione, C. Bielza, P. Larrañaga, S. Anderson, A. Burkhalter, B. Cauli, A. Fairén, D. Feldmeyer, G. Fishell, D. Fitzpatrick, T. F. Freund, G. González-Burgos, S. Hestrin, S. Hill, P. R. Hof, J. Huang, E. G. Jones, Y. Kawaguchi, Z. Kisvárdy, Y. Kubota, D. A. Lewis, O. Marin, H. Markram, C. J. McBain, H. S. Meyer, H. Momyer, S. B. Nelson, K. Rockland, J. Rossier, J. L. R. Rubenstein, B. Rudy, M. Scanziani, G. M. Shepherd, C. C. Sherwood, J. F. Staiger, G. Tamás, A. Thomson, Y. Wang, R. Yuste,

- G. A. Ascoli, New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* **14**, 202–216 (2013).
28. Petilla Interneuron Nomenclature Group, G. A. Ascoli, L. Alonso-Nanclares, S. A. Anderson, G. Barrionuevo, R. Benavides-Piccione, A. Burkhalter, G. Buzsáki, B. Cauli, J. Defelipe, A. Fairén, D. Feldmeyer, G. Fishell, Y. Fregnac, T. F. Freund, D. Gardner, E. P. Gardner, J. H. Goldberg, M. Helmstaedter, S. Hestrin, F. Karube, Z. F. Kisvárdy, B. Lambold, D. A. Lewis, O. Marin, H. Markram, A. Muñoz, A. Packer, C. C. H. Petersen, K. S. Rockland, J. Rossier, B. Rudy, P. Somogyi, J. F. Staiger, G. Tamas, A. M. Thomson, M. Toledo-Rodriguez, Y. Wang, D. C. West, R. Yuste, Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
  29. J. Dimidschstein, Q. Chen, R. Tremblay, S. L. Rogers, G.-A. Saldi, L. Guo, Q. Xu, R. Liu, C. Lu, J. Chu, J. S. Grimley, A.-R. Krostag, A. Kaykas, M. C. Avery, M. S. Rashid, M. Baek, A. L. Jacob, G. B. Smith, D. E. Wilson, G. Kosche, I. Kruglikov, T. Rusielewicz, V. C. Kotak, T. M. Mowery, S. A. Anderson, E. M. Callaway, J. S. Dasen, D. Fitzpatrick, V. Fossati, M. A. Long, S. Noggle, J. H. Reynolds, D. H. Sanes, B. Rudy, G. Feng, G. Fishell, A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nat. Neurosci.* **19**, 1743–1749 (2016).
  30. J. Langlieb, N. Sachdev, K. Balderrama, N. Nadaf, M. Raj, E. Murray, J. Webber, C. Vanderburg, V. Gazestani, D. Tward, C. Mezas, X. Li, D. Cable, T. Norton, P. P. Mitra, F. Chen, E. Macosko, The cell type composition of the adult mouse brain revealed by single cell and spatial genomics. *bioRxiv* 2023.03.06.531307 [Preprint]. 13 March 2023. <https://doi.org/10.1101/2023.03.06.531307>.
  31. K. L. Chiou, X. Huang, M. O. Bohlen, S. Tremblay, A. R. Decasien, D. R. O'Day, C. H. Spurrell, A. A. Gogate, T. M. Zintel, Cayo Biobank Research Unit, M. G. Andrews, M. I. Martinez, L. M. Starita, M. J. Montague, M. L. Platt, J. Shendure, N. Snyder-Mackler, A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *Sci. Adv.* **9**, eadh1914 (2023).
  32. N. L. Jorstad, J. H. T. Song, D. Exposito-Alonso, H. Suresh, N. Castro, F. M. Krienen, A. M. Yannay, J. C. Close, E. Gelfand, K. J. Travaglini, S. Basu, M. Beaudin, D. Bertagnolli, M. Crow, S.-L. Ding, J. Eggermont, A. Glandon, J. Goldy, T. Kroes, B. Long, D. McMillen, T. Pham, C. Rimorin, K. Siletti, S. Somasundaram, M. Tieu, A. Torkelson, K. Ward, G. Feng, W. D. Hopkins, T. Höllt, C. Dirk Keene, S. Linnarsson, S. A. McCarroll, B. P. Lelieveldt, C. C. Sherwood, K. Smith, C. A. Walsh, A. Dobin, J. Gillis, E. S. Lein, R. D. Hodge, T. E. Bakken, Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023).
  33. N. R. Zemke, E. J. Armand, W. Wang, S. Lee, J. Zhou, Y. E. Li, H. Liu, W. Tian, J. R. Nery, R. G. Castanon, A. Bartlett, J. K. Osteen, D. Li, X. Zhuo, V. Xu, M. Miller, F. M. Krienen, Q. Zhang, N. Taskin, J. Ting, G. Feng, S. A. McCarroll, E. M. Callaway, T. Wang, M. M. Behrens, E. S. Lein, J. R. Ecker, B. Ren, Comparative single cell epigenomic analysis of gene regulatory programs in the rodent and primate neocortex. *bioRxiv* 2023.04.08.536119 [Preprint]. 8 April 2023. <https://doi.org/10.1101/2023.04.08.536119>.
  34. B. R. Herb, H. J. Glover, A. Bhaduri, C. Colantuoni, T. L. Bale, K. Siletti, S. Linnarsson, R. Hodge, E. Lein, A. R. Kriegstein, C. A. Doege, S. A. Ament, Single-cell genomics reveals region-specific developmental trajectories underlying neuronal diversity in the human hypothalamus. *bioRxiv* 2021.07.20.453090 [Preprint]. 19 October 2022. <https://doi.org/10.1101/2021.07.20.453090>.
  35. J. Alquicira-Hernandez, A. Sathe, H. P. Ji, Q. Nguyen, J. E. Powell, scPred: Accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).
  36. H. Zeng, J. R. Sanes, Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
  37. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
  38. H. Zeng, What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
  39. D. L. Kaufman, C. R. Houser, A. J. Tobin, Two forms of the gamma-aminobutyric acid synthetic enzyme glutamate decarboxylase have distinct intraneuronal distributions and cofactor interactions. *J. Neurochem.* **56**, 720–723 (1991).
  40. G. Meyer, Building a human cortex: The evolutionary differentiation of Cajal-Retzius cells and the cortical hem. *J. Anat.* **217**, 334–343 (2010).
  41. P. Jager, G. Moore, P. Calpin, X. Durmishi, I. Salgarella, L. Menage, Y. Kita, Y. Wang, D. W. Kim, S. Blackshaw, S. R. Schultz, S. Brickley, T. Shimogori, A. Delogu, Dual midbrain and forebrain origins of thalamic inhibitory interneurons. *eLife* **10**, e59272 (2021).
  42. K. Letinic, P. Rakic, Telencephalic origin of human thalamic GABAergic neurons. *Nat. Neurosci.* **4**, 931–936 (2001).
  43. L. W. Swanson, G. D. Petrovich, What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).
  44. T. Hirata, P. Li, G. M. Lanuza, L. A. Cocas, M. M. Huntsman, J. G. Corbin, Identification of distinct telencephalic progenitor pools for neuronal diversity in the amygdala. *Nat. Neurosci.* **12**, 141–149 (2009).
  45. T. Kaoru, F. C. Liu, M. Ishida, T. Oishi, M. Hayashi, M. Kitagawa, K. Shimoda, H. Takahashi, Molecular characterization of the intercalated cell masses of the amygdala: Implications for the relationship with the striatum. *Neuroscience* **166**, 220–230 (2010).
  46. M. T. Schmitz, K. Sandoval, C. P. Chen, M. A. Mostajo-Radji, W. W. Seeley, T. J. Nowakowski, C. J. Ye, M. F. Paredes, A. A. Pollen, The development and evolution of inhibitory neurons in primate cerebrum. *Nature* **603**, 871–877 (2022).
  47. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
  48. C. C. Harwell, L. C. Fuentealba, A. Gonzalez-Cerrillo, P. R. L. Parker, C. C. Gertz, E. Mazzola, M. T. Garcia, A. Alvarez-Buylla, C. L. Cepko, A. R. Kriegstein, Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* **87**, 999–1007 (2015).
  49. C. Mayer, C. Hafemeister, R. C. Bandler, R. Machold, R. B. Brito, X. Jaglin, K. Allaway, A. Butler, G. Fishell, R. Satija, Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
  50. C. Neyt, M. Welch, A. Langston, J. Kohtz, G. Fishell, A short-range signal restricts cell movement between telencephalic proliferative zones. *J. Neurosci.* **17**, 9194–9203 (1997).
  51. F. García-Moreno, M. Pedraza, L. G. Di Giovannantonio, M. Di Salvio, L. López-Mascaraque, A. Simeone, J. A. De Carlos, A neuronal migratory pathway crossing from diencephalon to telencephalon populates amygdala nuclei. *Nat. Neurosci.* **13**, 680–689 (2010).
  52. A. J. Granger, M. L. Wallace, B. L. Sabatini, Multi-transmitter neurons in the mammalian central nervous system. *Curr. Opin. Neurobiol.* **45**, 85–91 (2017).
  53. N. X. Tritsch, A. J. Granger, B. L. Sabatini, Mechanisms and functions of GABA co-release. *Nat. Genet.* **17**, 139–145 (2016).
  54. M. Bupesh, I. Legaz, A. Abellán, L. Medina, Multiple telencephalic and extratelencephalic embryonic domains contribute neurons to the medial extended amygdala. *J. Comp. Neurol.* **519**, 1505–1525 (2011).
  55. L. Zhang, V. S. Hernandez, C. R. Gerfen, S. Z. Jiang, L. Zavala, R. A. Barrio, L. E. Eiden, Behavioral role of PACAP signaling reflects its selective distribution in glutamatergic and GABAergic neuronal subpopulations. *eLife* **10**, e61718 (2021).
  56. N. Bozadjieva-Kramer, R. A. Ross, D. Q. Johnson, H. Fenselau, D. L. Haggerty, B. Atwood, B. Lowell, J. N. Flak, The role of mediobasal hypothalamic PACAP in the control of body weight and metabolism. *Endocrinology* **162**, bqab012 (2021).
  57. J.-H. Cho, K. Zushida, G. P. Shumyatsky, W. A. Carlezon Jr., E. G. Meloni, V. Y. Bolshakov, Pituitary adenylate cyclase-activating polypeptide induces postsynaptically expressed potentiation in the intra-amygdala circuit. *J. Neurosci.* **32**, 14165–14177 (2012).
  58. E. Y. Lee, L. C. Chan, H. Wang, J. Lieng, M. Hung, Y. Srinivasan, J. Wang, J. A. Waschek, A. L. Ferguson, K.-F. Lee, N. Y. Yount, M. R. Yeaman, G. C. L. Wong, PACAP is a pathogen-inducible resident antimicrobial neuropeptide affording rapid and contextual molecular host defense of the brain. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1917623117 (2021).
  59. Y. Kita, H. Nishibe, Y. Wang, T. Hashikawa, S. S. Kikuchi, M. U. A. C. Yoshida, C. Yoshida, T. Kawase, S. Ishii, H. Skibbe, T. Shimogori, Cellular-resolution gene expression profiling in the neonatal marmoset brain reveals dynamic species- and region-specific differences. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020125118 (2021).
  60. T. Shimogori, A. Abe, Y. Go, T. Hashikawa, N. Kishi, S. S. Kikuchi, Y. Kita, K. Niimi, H. Nishibe, M. Okuno, K. Saga, M. Sakurai, M. Sato, T. Serizawa, S. Suzuki, E. Takahashi, M. Tanaka, S. Tatsumoto, M. Toki, M. U. Y. Wang, K. J. Windak, H. Yamagishi, K. Yamashita, T. Yoda, A. C. Yoshida, C. Yoshida, T. Yoshimoto, H. Okano, Digital gene atlas of neonate common marmoset brain. *Neurosci. Res.* **128**, 1–13 (2018).
  61. D. P. Orquera, M. B. Tavella, F. S. J. de Souza, S. Nasif, M. J. Low, M. Rubinstein, The homeodomain transcription factor NKX2.1 is essential for the early specification of melanocortin neuron identity and activates pomc expression in the developing hypothalamus. *J. Neurosci.* **39**, 4023–4035 (2019).
  62. R. Murcia-Ramón, V. Company, I. Juárez-Leal, A. Andreu-Cervera, F. Almagro-García, S. Martínez, D. Echevarría, E. Puelles, Neuronal tangential migration from Nkx2.1-positive hypothalamus. *Brain Struct. Funct.* **225**, 2857–2869 (2020).
  63. A. A. Pollen, T. J. Nowakowski, J. Chen, H. Retallack, C. Sandoval-Espinosa, C. R. Nicholas, J. Shuga, S. J. Liu, M. C. Oldham, A. Diaz, D. A. Lim, A. A. Leyrat, J. A. West, A. R. Kriegstein, Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
  64. A. Bhaduri, C. Sandoval-Espinosa, M. Otero-García, I. Oh, R. Yin, U. C. Eze, T. J. Nowakowski, A. R. Kriegstein, An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
  65. B. Tasic, Z. Yao, L. T. Graybiel, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M. N. Economou, S. Viswanathan, O. Penn, T. Bakken, V. Menon, J. Miller, O. Fong, K. E. Hirokawa, K. Lathia, C. Rimorin, M. Tieu, R. Larsen, T. Casper, E. Barkan, M. Kroll, S. Parry, N. V. Shapovalova, D. Hirschstein, J. Pendergraft, H. A. Sullivan, T. K. Kim, A. Szafer, N. Dee, P. Groblewski, I. Wickersham, A. Cetin, J. A. Harris, B. P. Levi, S. M. Sunkin, L. Madisen, T. L. Daigle, L. Looger, A. Bernard, J. Phillips, E. Lein, M. Hawrylycz, K. Svoboda, A. R. Jones,

- C. Koch, H. Zeng, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
66. P. Balaram, J. H. Kaas, Towards a unified scheme of cortical lamination for primary visual cortex across primates: Insights from NeuN and VGLUT2 immunoreactivity. *Front. Neuroanat.* **8**, 81 (2014).
67. O. A. Bayraktar, L. C. Fuentealba, A. Alvarez-Buylla, D. H. Rowitch, Astrocyte development and heterogeneity. *Cold Spring Harb. Perspect. Biol.* **7**, a020362 (2015).
68. O. A. Bayraktar, T. Bartels, S. Holmqvist, V. Kleshchevnikov, A. Martirosyan, D. Polioudakis, L. Ben Haim, A. M. H. Young, M. Y. Batiuk, K. Prakash, A. Brown, K. Roberts, M. F. Paredes, R. Kawaguchi, J. H. Stockley, K. Sabeur, S. M. Chang, E. Huang, P. Hutchinson, E. M. Ullian, M. Hemberg, G. Coppola, M. G. Holt, D. H. Geschwind, D. H. Rowitch, Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat. Neurosci.* **23**, 500–509 (2020).
69. M. Y. Batiuk, A. Martirosyan, J. Wahis, F. de Vin, C. Marneffe, C. Kusserow, J. Koepfen, J. F. Viana, J. F. Oliveira, T. Voet, C. P. Ponting, T. G. Belgard, M. G. Holt, Identification of region-specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220 (2020).
70. N. Johansen, S. Somasundaram, K. J. Travaglini, A. M. Yanny, M. Shumyatcher, T. Casper, C. Cobbs, N. Dee, R. Ellenbogen, M. Ferreira, J. Goldy, J. Guzman, R. Gwinn, D. Hirschstein, N. L. Jorstad, C. Dirk Keene, A. Ko, B. P. Levi, J. G. Ojemann, T. Pham, N. Shapovalova, D. Silbergeld, J. Sulc, A. Torkelson, H. Tung, K. Smith, E. S. Lein, T. E. Bakken, R. D. Hodge, J. A. Miller, Inter-individual variation in human cortical cell type abundance and expression. *Science* **382**, eadf2359 (2023).
71. A. Kriegstein, A. Alvarez-Buylla, The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).
72. R. C. Bandler, C. Mayer, G. Fishell, Cortical interneuron specification: The juncture of genes, time and geometry. *Curr. Opin. Neurobiol.* **42**, 17–24 (2017).
73. Z. Petanjek, I. Kostovic, M. Esclapez, Primate-specific origins and migration of cortical GABAergic neurons. *Front. Neuroanat.* **3**, 26 (2009).
74. R. N. Delgado, D. E. Allen, M. G. Keefe, W. R. Mancia Leon, R. S. Zifra, E. E. Crouch, A. Alvarez-Buylla, T. J. Nowakowski, Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022).
75. C. E. Collins, D. C. Airey, N. A. Young, D. B. Leitch, J. H. Kaas, Neuron densities vary across and within cortical areas in primates. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15927–15932 (2010).
76. A. Woodward, T. Hashikawa, M. Maeda, T. Kaneko, K. Hikishima, A. Iriki, H. Okano, Y. Yamaguchi, The Brain/MINDS 3D digital marmoset brain atlas. *Sci Data.* **5**, 180009 (2018).
77. B. Schuman, R. P. Machold, Y. Hashikawa, J. Fuzik, G. J. Fishell, B. Rudy, Four unique interneuron populations reside in neocortical layer 1. *J. Neurosci.* **39**, 125–139 (2019).
78. M. Valero, T. J. Viney, R. Machold, S. Mederos, I. Zutshi, B. Schuman, Y. Senzai, B. Rudy, G. Buzsáki, Sleep down state-active ID2/Nkx2.1 interneurons in the neocortex. *Nat. Neurosci.* **24**, 401–411 (2021).
79. B. Zingg, H. Hintiryan, L. Gou, M. Y. Song, M. Bay, M. S. Bienkowski, N. N. Foster, S. Yamashita, I. Bowman, A. W. Toga, H.-W. Dong, Neural networks of the mouse neocortex. *Cell* **156**, 1096–1111 (2014).
80. H. B. M. Uylings, H. J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex? *Behav. Brain Res.* **146**, 3–17 (2003).
81. M. K. Lin, Y. S. Takahashi, B.-X. Huo, M. Hanada, J. Nagashima, J. Hata, A. S. Tolpygo, K. Ram, B. C. Lee, M. I. Miller, M. G. Rosa, E. Sasaki, A. Iriki, H. Okano, P. Mitra, A high-throughput neurohistological pipeline for brain-wide mesoscale connectivity mapping of the common marmoset. *eLife* **8**, e40042 (2019).
82. P. Majka, S. Bai, S. Bakola, S. Bednarek, J. M. Chan, N. Jermakow, L. Passarelli, D. H. Reser, P. Theodoni, K. H. Worthy, X.-J. Wang, D. K. Wójcik, P. P. Mitra, M. G. P. Rosa, Open access resource for cellular-resolution analyses of corticocortical connectivity in the marmoset monkey. *Nat. Commun.* **11**, 1133 (2020).
83. A. Hladnik, D. Džaja, S. Darmopil, N. Jovanov-Milošević, Z. Petanjek, Spatio-temporal extension in site of origin for cortical calretinin neurons in primates. *Front. Neuroanat.* **8**, 50 (2014).
84. E. Y. Choi, B. T. T. Yeo, R. L. Buckner, The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **108**, 2242–2263 (2012).
85. K. M. Anderson, F. M. Krienen, E. Y. Choi, J. M. Reinen, B. T. T. Yeo, A. J. Holmes, Gene expression links functional networks across cortex and striatum. *Nat. Commun.* **9**, 1428 (2018).
86. M. Matamalas, J. Götz, J. Bertran-Gonzalez, Quantitative imaging of cholinergic interneurons reveals a distinctive spatial organization and a functional gradient across the mouse striatum. *PLOS ONE* **11**, e0157682 (2016).
87. E. Fino, M. Vandecasteele, S. Perez, F. Saudou, L. Venance, Region-specific and state-dependent action of striatal GABAergic interneurons. *Nat. Commun.* **9**, 3339 (2018).
88. A. B. M. Manchado, C. B. Gonzales, A. Zeisel, H. Munguba, B. Bekkouche, N. G. Skene, P. Lonnerberg, J. Ryge, K. D. Harris, S. Linnarsson, J. Hjerling-Leffler, Diversity of interneurons in the dorsal striatum revealed by single-cell RNA sequencing and PatchSeq. *Cell Rep.* **24**, 2179–2190.e7 (2018).
89. T. Zerucha, T. Stühmer, G. Hatch, B. K. Park, Q. Long, G. Yu, A. Gambarotta, J. R. Schultz, J. L. Rubenstein, M. Ekker, A highly conserved enhancer in the *Dlx5/Dlx6* intergenic region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain. *J. Neurosci.* **20**, 709–721 (2000).
90. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
91. D. Velmeshev, L. Schirmer, D. Jung, M. Haeussler, Y. Perez, S. Mayer, A. Bhaduri, N. Goyal, D. H. Rowitch, A. R. Kriegstein, Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
92. W. B. Ruzicka, S. Mohammadi, J. F. Fullard, J. Davila-Velderrain, S. Subburaju, D. R. Tso, M. Hourihan, S. Jiang, H.-C. Lee, J. Bendl, G. Voloudakis, V. Haroutunian, G. E. Hoffman, P. Roussos, M. Kellis, PsychENCODE Consortium, Single-cell multi-cohort dissection of the schizophrenia transcriptome. medRxiv 2022.08.31.22279406 [Preprint]. 2 September 2022. <https://doi.org/10.1101/2022.08.31.22279406>.
93. C. Nagy, M. Maitra, A. Tanti, M. Suderman, J.-F. Théroux, M. A. Davoli, K. Perlman, V. Yerko, Y. C. Wang, S. J. Tripathy, P. Pavlidis, N. Mechawar, J. Ragoussis, G. Turecki, Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat. Neurosci.* **23**, 771–781 (2020).
94. A. Sequeira, F. Mamdani, C. Ernst, M. P. Vawter, W. E. Bunney, V. Lebel, S. Rehal, T. Klempan, A. Gratton, C. Benkelfat, G. A. Rouleau, N. Mechawar, G. Turecki, Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLOS ONE* **4**, e6585 (2009).
95. D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
96. H.-H. Tsai, H. Li, L. C. Fuentealba, A. V. Molofsky, R. Taveira-Marques, H. Zhuang, A. Tenney, A. T. Murnen, S. P. J. Fancy, F. Merkle, N. Kessaris, A. Alvarez-Buylla, W. D. Richardson, D. H. Rowitch, Regional astrocyte allocation regulates CNS synaptogenesis and repair. *Science* **337**, 358–362 (2012).
97. L. Morel, M. S. R. Chiang, H. Higashimori, T. Shoneye, L. K. Iyer, J. Yelick, A. Tai, Y. Yang, Molecular and functional properties of regional astrocytes in the adult brain. *J. Neurosci.* **37**, 8706–8717 (2017).
98. H. Chai, B. Diaz-Castro, E. Shigetomi, E. Monte, J. C. Octeau, X. Yu, W. Cohn, P. S. Rajendran, T. M. Vondriska, J. P. Whitelegge, G. Coppola, B. S. Khakh, Neural circuit-specialized astrocytes: Transcriptomic, proteomic, morphological, and functional evidence. *Neuron* **95**, 531–549.e9 (2017).
99. E. Braun, M. Danan-Gotthold, L. E. Borm, E. Vinsland, K. W. Lee, P. Lönnerberg, L. Hu, X. Li, X. He, Ž. Andrusivová, J. Lundeberg, E. Arenas, R. A. Barker, E. Sundström, S. Linnarsson, Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
100. S. C. Noctor, V. Martínez-Cerdeño, L. Ivic, A. R. Kriegstein, Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).
101. G. La Manno, K. Siletti, A. Furlan, D. Gyllborg, E. Vinsland, A. Mossi Albiach, C. Mattsson Langseth, I. Khven, A. R. Lederer, L. M. Dratva, A. Johnsson, M. Nilsson, P. Lönnerberg, S. Linnarsson, Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021).
102. D. J. Di Bella, E. Habibi, R. R. Stickels, G. Scalia, J. Brown, P. Yadollahpour, S. M. Yang, C. Abbate, T. Biancalani, E. Z. Maccosio, F. Chen, A. Regev, P. Arlotta, Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).
103. B. Martynoga, H. Morrison, D. J. Price, J. O. Mason, Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. *Dev. Biol.* **283**, 113–127 (2005).
104. J. M. Tepper, F. Tecuapetla, T. Koós, O. Ibáñez-Sandoval, Heterogeneity and diversity of striatal GABAergic interneurons. *Front. Neuroanat.* **4**, 150 (2010).
105. R. Tremblay, S. Lee, B. Rudy, GABAergic Interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
106. D. Vormstein-Schneider, J. D. Lin, K. A. Pelkey, R. Chittajallu, B. Guo, M. A. Arias-Garcia, K. Allaway, S. Sakopoulos, G. Schneider, O. Stevenson, J. Vergara, J. Sharma, Q. Zhang, T. P. Frankel, J. Smith, L. A. Ibrahim, K. J. M. Astro, E. Sabri, S. Huang, E. Favuzzi, T. Burbridge, Q. Xu, L. Guo, I. Vogel, V. Sanchez, G. A. Saldi, B. L. Gorissen, X. Yuan, K. A. Zaghoul, O. Devinsky, B. L. Sabatini, R. Batista-Brito, J. Reynolds, G. Feng, Z. Fu, C. J. McBain, G. Fishell, J. Dimidschstein, Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nat. Neurosci.* **23**, 1–21 (2020).
107. J. K. Mich, L. T. Graybuck, E. E. Hess, J. T. Mahoney, Y. Kojima, Y. Ding, S. Somasundaram, J. A. Miller, B. E. Kalmbach, C. Radaelli, B. B. Gore, N. Weed, V. Omstead, Y. Bishaw, N. V. Shapovalova, R. A. Martinez, O. Fong, S. Yao, M. Mortrud, P. Chong, L. Loftus,

- D. Bertagnolli, J. Goldy, T. Casper, N. Dee, X. Opitz-Araya, A. Cetin, K. A. Smith, R. P. Gwinn, C. Cobbs, A. L. Ko, J. G. Ojemann, C. D. Keene, D. L. Silbergeld, S. M. Sunkin, V. Gradinaru, G. D. Horwitz, H. Zeng, B. Tasic, E. S. Lein, J. T. Ting, B. P. Levi, Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).
108. D. Goertzen, N. C. Flytzanis, N. Goeden, M. R. Chuapoco, A. Cummins, Y. Chen, Y. Fan, Q. Zhang, J. Sharma, Y. Duan, L. Wang, G. Feng, Y. Chen, N. Y. Ip, J. Pickel, V. Gradinaru, AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).
109. M. R. Chuapoco, N. C. Flytzanis, N. Goeden, J. C. Oceau, K. M. Roxas, K. Y. Chan, J. Scherrer, J. Winchester, R. J. Blackburn, L. J. Campos, K. N. M. Man, J. Sun, X. Chen, A. Lefevre, V. P. Singh, C. M. Arokiaraj, T. F. Shay, J. Vendemiatti, M. J. Jang, J. K. Mich, Y. Bishaw, B. B. Gore, V. Omstead, N. Taskin, N. Weed, B. P. Levi, J. T. Ting, C. T. Miller, B. E. Deverman, J. Pickel, L. Tian, A. S. Fox, V. Gradinaru, Adeno-associated viral vectors for functional intravenous gene transfer throughout the non-human primate brain. *Nat. Nanotechnol.* 10.1038/s41565-023-01419-x (2023).
110. G. Paxinos, C. Watson, M. Petrides, M. Rosa, H. Tokuno, *The Marmoset Brain in Stereotaxic Coordinates* (Elsevier Academic Press, 2012).
111. S. J. Fleming, M. D. Chaffin, A. Arduini, A.-D. Akkad, E. Banks, J. C. Marioni, A. A. Philippakis, P. T. Ellinor, M. Babadi, Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
112. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
113. S. Xu, Z. Dai, P. Guo, X. Fu, S. Liu, L. Zhou, W. Tang, T. Feng, M. Chen, L. Zhan, T. Wu, E. Hu, Y. Jiang, X. Bo, G. Yu, ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Mol. Biol. Evol.* **38**, 4039–4042 (2021).
114. L. Feng, T. Zhao, J. Kim, neuTube 1.0: A new design for efficient neuron reconstruction software based on the SWC format. *eNeuro* **2**, ENEURO.0049 (2015).
115. R. C. Challis, S. Ravindra Kumar, K. Y. Chan, C. Challis, K. Beadle, M. J. Jang, H. M. Kim, P. S. Rajendran, J. D. Tompkins, K. Shivkumar, B. E. Deverman, V. Gradinaru, Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat. Protoc.* **14**, 379–414 (2019).
116. J. V. Liu, N. A. Bock, A. C. Silva, Rapid high-resolution three-dimensional mapping of T1 and age-dependent variations in the non-human primate brain using magnetization-prepared rapid gradient-echo (MPRAGE) sequence. *Neuroimage* **56**, 1154–1163 (2011).

**Acknowledgments:** We thank T. Blosser for his early involvement in developing the smFISH workflows. We thank A. Takahashi for assistance with MRI scanning for animals receiving local

viral injections. We thank the MIT veterinarian staff for animal husbandry and for their assistance with surgical procedures. We thank M. Burns and Y. Hou for assistance with AAV intravenous injections and animal perfusions. We acknowledge K.X.L.'s undergraduate thesis (March 2022, Division of Life Sciences, Harvard University), which motivated subsequent analyses included in this manuscript. We thank M. DeBerardine for assistance with cellgene deposition. This publication was supported by and coordinated through the Brain Initiative Cell Census Network (BICCN). This publication is part of the Human Cell Atlas ([www.humancellatlas.org/publications/](http://www.humancellatlas.org/publications/)). **Funding:** This work was supported by NIH grants U01MH114819 (to G.F., S.A.M., and E.B.) and 1F31MH133329 (to M.E.S.), National Institute of Neurological Disorders and Stroke grant UG3NS111689 (to B.E.D.), NSF GRFP #1745302 (to M.E.S.), MathWorks Science Fellowship (to M.E.S.), Collamore-Rogers Fellowship at MIT (M.E.S.), NSF GRFP #1122374 (T.W.S.), Broad Institute's Stanley Center for Psychiatric Research (to S.A.M., G.F., and B.E.D.), Dean's Innovation Award (Harvard Medical School) (to S.A.M.), Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT (to G.F.), Poitras Center for Psychiatric Disorders Research at MIT (to G.F.), and McGovern Institute for Brain Research at MIT (to G.F.). **Author contributions:** RNA/ATAC data generation: F.M.K., M.G., A.L., and G.M. Spatial/smFISH data generation: K.M.L., H.Z., M.E.S., D.M., and K.X.L. Data analysis: F.M.K., K.M.L., H.Z., R.C.H.d.R., M.E.S., M.G., A.L., K.X.L., V.F.B.-G., D.M., T.W.S., and S.A.M. Data interpretation: F.M.K., K.M.L., H.Z., R.C.H.d.R., M.E.S., M.G., K.X.L., V.F.B.-G., S.B., E.B., D.M., S.A.M., and G.F. Tissue samples/tissue processing: F.M.K., K.M.L., H.Z., M.G., A.L., Q.Z., G.C., and S.B. AAV design, generation, and experiments: M.W., C.C., Q.Z., J.S., J.V., J.D., A.H., E.N., K.Y.C., and B.E.D. Morphology data generation/analysis: K.M.L., H.Z., V.F.B.-G., T.W.S., A.M., and E.B. Software/data management: F.M.K., K.M.L., R.C.H.d.R., M.G., A.W., J.N., and S.K. Writing: F.M.K., K.M.L., H.Z., M.E.S., V.F.B.-G., S.A.M., and G.F. **Competing interests:** B.E.D. and K.Y.C. are listed as inventors on a patent application filed by the Broad Institute related to the AAV-BI103 capsid and methods of its use. All other authors declare no competing interests. **Data and materials availability:** Raw sequence data produced as part of the BICCN are available for download from the Neuroscience Multi-omics Archive (<https://assets.nemoarchive.org/dataset/1je0mn3>) and the Brain Cell Data Center (<https://biccn.org/data>). Processed data can be explored and downloaded from the cellxgene data repository (RRID:SCR\_021059; <https://cellxgene.cziscience.com/collections/0fd39ad7-5d2d-41c2-bda0-c55bde614bdb>). Morphological reconstructions and single-molecule FISH of interneuron types are available for download through the Brain Image Library (<https://doi.brainimaginglibrary.org/doi/10.35077/g.609>). The AAV-BI103 rep-cap plasmid is available through Addgene (Addgene ID: 203537) and through direct requests to B.E.D. (bdeverma@broadinstitute.org). Code sources are in table S7.

Submitted 19 August 2023

Accepted 26 September 2023

Published 12 October 2023

10.1126/sciadv.adk3986

## A marmoset brain cell census reveals regional specialization of cellular identities

Fenna M. Krienen, Kirsten M. Levandowski, Heather Zaniewski, Ricardo C.H. del Rosario, Margaret E. Schroeder, Melissa Goldman, Martin Wienisch, Alyssa Lutservitz, Victoria F. Beja-Glasser, Cindy Chen, Qiangge Zhang, Ken Y. Chan, Katelyn X. Li, Jitendra Sharma, Dana McCormack, Tay Won Shin, Andrew Harrahill, Eric Nyase, Gagandeep Mudhar, Abigail Mauermann, Alec Wysoker, James Nemes, Seva Kashin, Josselyn Vergara, Gabriele Chelini, Jordane Dimidschstein, Sabina Berretta, Benjamin E. Deverman, Ed Boyden, Steven A. McCarroll, and Guoping Feng

*Sci. Adv.* **9** (41), eadk3986. DOI: 10.1126/sciadv.adk3986

### View the article online

<https://www.science.org/doi/10.1126/sciadv.adk3986>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).