A marmoset brain cell census reveals persistent influence of developmental origin on neurons

Fenna M. Krienen[1,2]*, Kirsten M. Levandowski[2,3,4], Heather Zaniewski[2,3], Ricardo C.H. del Rosario[2], Margaret E. Schroeder[2,3,4], Melissa Goldman[2], Alyssa Lutservitz[1], Qiangge Zhang[2,3,4], Katelyn X. Li[3,4], Victoria F. Beja-Glasser[2,3,4], Jitendra Sharma[3,4], Tay Won Shin[3,4,5], Abigail Mauermann[3,4,5], Alec Wysoker[2], James Nemesh[2], Seva Kashin[2], Josselyn Vergara[2], Gabriele Chelini[6], Jordane Dimidschstein[2], Sabina Berretta[7,8], Ed Boyden[3,4,5], Steven A. McCarroll[1,2#]*, Guoping Feng[2,3#]*

[1] Department of Genetics, Harvard Medical School; Boston, Massachusetts 02115, USA.
[2] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard; Cambridge, Massachusetts 02142, USA.
[3] McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; Cambridge, Massachusetts 02142, USA.
[4] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; Cambridge, Massachusetts 02139, USA.
[5] HHMI, Cambridge, MA, USA.
[6] Center for Mind/Brain Sciences, University of Trento. Piazza della Manifattura n.1, Rovereto (TN). 38068, IT.
[7] Basic Neuroscience Division, McLean Hospital, Belmont MA USA.
[8] Dept. of Psychiatry, Harvard Medical School, Boston, MA USA.

[#]These authors contributed equally to this work
*Corresponding authors. Email: fkrienen@broadinstitute.org, mccarroll@genetics.med.harvard.edu, fengg@mit.edu

Abstract:

Within the vertebrate neocortex and other telencephalic structures, molecularly-defined neurons tend to segregate at first order into inhibitory (GABAergic) and excitatory (glutamatergic) types. We used single-nucleus RNA sequencing, analyzing over 2.4 million brain cells sampled from 16 locations in a primate (the common marmoset) to ask whether (1) neurons generally segregate by neurotransmitter status, and (2) neurons expressing the same neurotransmitters share additional molecular features in common, beyond the few genes directly responsible for neurotransmitter synthesis and release. We find the answer to both is "no": there is a remarkable degree of transcriptional similarity between GABAergic and glutamatergic neurons found in the same brain structure, and there is generally little in common between glutamatergic neurons residing in phylogenetically divergent brain structures. The origin effect is permanent: we find that cell types that cross cephalic boundaries in development retain the transcriptional identities of their birthplaces. GABAergic interneurons, which migrate widely, follow highly specialized and distinct distributions in striatum and neocortex. We use interneuron-restricted AAVs to reveal the morphological diversity of molecularly defined types. Our analyses expose how lineage and functional class sculpt the transcriptional identity and biodistribution of primate neurons.

One-Sentence Summary: Primate neurons are primarily imprinted by their region of origin, more so than by their functional identity.

Main Text:

The complex functional heterogeneity of the primate brain has its basis, in part, in the diversity of its cellular and molecular repertoire. Here, we provide a census of cell types across major cortical and subcortical structures in the adult marmoset brain. Previous single cell sequencing studies of the marmoset brain focused on single brain regions (*1*, *2*), or on specific cell classes across regions (*3*, *4*). However, inclusion of both closely and distantly related brain structures and cell types can yield insights into the developmental and ontological relationships between them (*5*). While comprehensive transcriptomic atlases exist for the mouse (*6*, *7*), less is known about the landscape of brain cell types in primates.

We generated single-nucleus RNA sequencing (snRNA-seq; 10x Genomics 3' v3.1) data from 2.4 million unsorted nuclei across 8 cortical and 8 subcortical locations from 10 young adult marmosets (4 M, 6 F), and resolved clusters from all major neuronal and non-neuronal cell classes. snRNA-seq data are available on the Brain Initiative Cell Census Network (BICCN) website as well as the NeMO archive (https://nemoarchive.org/data/). We used smFISH to spatially profile major interneuron types in striatum and neocortex in marmosets that had received systemic delivery of AAVs carrying a reporter (GFP) under an interneuron-selective regulatory element (*8*). This enabled molecular identification of morphologically reconstructed neurons, which are available for download from the Brain Image Library (BIL; https://www.brainimagelibrary.org/).

All brain structures in the central nervous system contain excitatory and inhibitory neuronal populations, though proportions may vary widely. We found that the transcriptional identities of excitatory and inhibitory neurons within telencephalic brain structures segregate strongly. In contrast, there is much greater transcriptional similarity between GABAergic and glutamatergic neuronal types in non-telencephalic compartments. Moreover, few distinctions present in telencephalic glutamatergic neurons are shared with glutamatergic neurons in non-telecephalic brain regions. The adult transcriptomic identity of a neuron is shaped much more by its developmental identity than by its signaling repertoire.

Compared to mouse or other outgroup species, few primate lineage-gained cell types have emerged; likely more common is the redirection, repurposing, or elaboration of conserved types (*2*, *9–12*). We find that a recently discovered, primate-specific striatal interneuron type (*3*) is molecularly similar to other *TAC3* expressing types in the basal forebrain and hypothalamus. The similarity between diencephalic and telencephalic subtypes suggests that this could be an example of cross-cephalic vesicle migration. In general, we find that neurons that cross cephalic boundaries retain the transcriptional identities of their origins.

A transcriptomic taxonomy of marmoset brain cell types

snRNA-seq from each brain region (**Fig. 1A**) was first analyzed independently to identify major cell types and their proportions (**Fig. 1B-D**). Briefly, using linear discriminant analysis (scPred; (*13*)) 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). For non-neuronal cells, we aggregated

cells across brain structures for each class and clustered them together to resolve their cellular taxonomies within and across brain structures. GABAergic and glutamatergic neurons from cortical regions were classified and clustered separately. Each clustering analysis involved additional curation of doublets and outlier cells, followed by a round of sub-clustering of major clusters. For each cluster, a "metacell" was generated that represented scaled and normalized expression of cells in the cluster, which was the starting point for downstream, cross-region analyses.

Regional expression fingerprints differ between neocortical neurons and glia

The mammalian neocortex is partitioned into functionally, connectionally, and cytoarchitectonically distinct regions, called areas. A longstanding question is the ontological relationship between cell types in different cortical brain areas. We examined regional distinctions in proportions and gene expression of cell types from 2.4 million nuclei sampled across 8 cortical locations (**Fig. 1A**). Consistent with prior reports (*3*, *6*, *10*, *14*), neuronal subtypes identified in one cortical region were generally present in all other cortical regions, though in different proportions (**Fig. 1C,D**). There were two main exceptions. GABAergic *MEIS2*+ cells were far overrepresented in PFC samples (specifically in dissections of medial and medio-orbital PFC, **Fig. 1E,F**), a compositional distinction not observed in mouse (*6*, *14*). These cells likely correspond to the recently described population of LGE-derived *MEIS2*+ neurons that populate the olfactory bulb in mice, which are instead directed to medial prefrontal cortex in macaques and humans (*12*). The second exception was a cluster of *RORB+, KCNH8+* glutamatergic neurons in V1 (and to a lesser extent V2) that diverged strongly from *RORB*+ populations in the other cortical regions (**Fig. 1E,F**). The expansion and divergence of *RORB*+ populations in visual cortex is consistent with the elaboration and sub-specialization of primary V1 layer IV in primates (*15*).

We resolved 4 clusters of cortical astrocytes, including 2 protoplasmic (AST01-Ctx_RORA and AST02_Ctx_APOE) and 2 fibrous clusters (AST03-Ctx_GFAP and AST04-Ctx_GRIK2), determined to be fibrous or protoplasmic based on expression of *GFAP* and other marker genes (**Fig. 1F**) (*10*). The AST04-Ctx_GRIK2 cluster was 5-fold enriched in V1 (enrichment calculated as the proportion of cluster in region / proportion of cluster in the whole cortex) and 2.5-fold enriched in V2 (**Fig. S1A**). In contrast to astrocytes, the oligodendrocyte lineage was more regionally homogenous (**Fig. S1A**). We resolved 1 oligodendrocyte precursor cell (OPC) cluster (OPC-Ctx_PTPRZ1), 1 committed oligodendrocyte precursor (COP) / newly-formed oligodendrocyte (NFOL) cluster (COP-NFOL-Ctx_PCDH7), and 3 mature or myelin-forming oligodendrocyte clusters (MOL1-Ctx_PEX5L, MOL2-Ctx_PLP1, and MOL3-Ctx_PCDH7) (**Fig. 1F**). All clusters were present in similar (fold change < 2) proportions across cortical regions, except for the OPC cluster, which was ~4-fold depleted (relative proportion of OPC-Ctx_PTPRZ1 in V1 to cluster in whole cortex < 0.25) and the MOL3 cluster, which was >2-fold increased in V1 (**Fig. S1A**). This prompted us to investigate the normalized abundance (to astrocytes, as their abundance is relatively consistent across regions) of OPCs across cortical regions, which averaged 30% across all regions. We observed an increased OPC:astrocyte ratio in temporal association and prefrontal regions (temp_tpo = 48.1%, temp = 38.2%, pfcdl = 35.6%), and decreased OPC:astrocyte ratio in V1 (24.7%), V2 (25.0%), M1 (25.8%), and A1 (26.7%). These differences may represent regional specializations in the population of OPCs on hand to perform functions such as neuron-OPC synaptic communication or adaptive myelination, which play important roles in neuronal signaling and cognitive functions (*16*, *17*).

Gene expression varies in broad macroscale spatial gradients across the neocortex, and spatial proximity is a strong predictor of transcriptional similarity (*18*). Broad functional classes, such as the distinction between sensory and higher order areas, also configures neocortical gene expression (*18*, *19*). While spatial gradients may emerge at the aggregate level, it is unclear whether cell types within a cortical region vary uniformly, or whether different types follow distinct spatial rules. Differences in cortical regional expression could be shared across cell types or could be private to each type. Glutamatergic neurons sampled from distinct cortical locations have regionally distinct gene expression (*14*), and primate cortical GABAergic interneurons also harbor regional gene expression differences (*3*). Glial types – particularly astrocytes – are heterogenous across the major cephalic subdivisions in mouse (*2*, *7*, *20*, *21*) and primate brain (*4*); but the extent to which they are locally customized in distinct regions of neocortex is less well understood. Transcriptomic profiling in the mouse has revealed layer-specific astrocyte subpopulations in the cortex (*22*) and heterogeneity within cortical and hippocampal astrocytes (*23*), but these studies did not systematically compare astrocyte transcriptomes within distinct cortical subregions.

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 (11 types), inhibitory neurons (8 types), astrocytes (2 types) and oligodendrocyte lineage types (2 types), sampled in 8 neocortical locations. All cell types displayed regionally differentially expressed genes (rDEGs) (**Fig. S1-S6**). Similar to what has been described in the mouse cortex (*22*, *23*), astrocytes within the marmoset cortex exhibited regional heterogeneity (**Fig. S1E**), but overall neurons had far more rDEGs than macroglia (**Fig. 1G,H; Fig. S1B**). 62% of rDEGs in interneurons are also rDEGs in excitatory neurons. Interestingly, though cortical astrocytes and oligodendrocytes arise from a common lineage with cortical excitatory neurons (*24*), they shared a lower percentage of rDEGs in common with excitatory neurons (38% for protoplasmic astrocytes, 28% for fibrous astrocytes, 33% for MOLs, 54% for OPCs). Regionally differentially expressed genes within a cell class (glutamatergic, GABAergic or 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. Across all cell types and particularly within neurons, higher-order temporal association cortex and prefrontal cortex tended to be most distinct from V1 and V2 (**Fig. S1B,C**).

Developmental lineage, not neurotransmitter usage, defines transcriptional identities

Transcriptomic diversity arises from a number of sources, including a cell's function, developmental origin or lineage, and regional context. The neurotransmitters employed by a neuron are assumed to be fundamental to its identity and function. In the vertebrate central nervous system, glutamate is the primary excitatory neurotransmitter, while GABA or glycine are inhibitory. Within the neocortex and other telencephalic structures, transcriptomically-defined neuron types hierarchically group first into GABAergic types and glutamatergic types (*2*, *7*, *20*), a distinction that reflects these types' distinct developmental origins. In the mouse, other structures such as those found in the diencephalon and mesencephalon similarly segregate initially by GABAergic and glutamatergic status (*7*).

To evaluate how neurotransmitter utilization, developmental origin, brain structure, or other dimensions exert influence on adult primate neuronal transcriptomes, nuclei were clustered by brain region for each of 9 structures: cerebellum, brainstem, hypothalamus, thalamus, amygdala, striatum (including caudate, putamen and nucleus accumbens/ventral striatum), hippocampus and neocortex (collapsing all 8

neocortical samples into one). We resolved 288 neuronal types after excluding doublets and artifactual signals. We summed transcript counts individual cells of each type together, and generated 288 normalized and scaled "metacells". Hierarchical clustering was used to position the neuronal types on a single dendrogram using a correlation-based distance metric across 5,904 genes expressed above nominal levels (>10 transcripts per 100,000) in these cell types (**Fig. 2A**).

One possibility is that neurotransmitter deployment is fundamental to a neuron's transcriptional identity, and constitutes a core transcriptional phenotype that is shared across all neuronal types that utilize that neurotransmitter. For example, inhibitory neurons, which use GABA as a primary neurotransmitter, may be globally more similar to each other transcriptionally than they are to glutamatergic neurons across all brain regions sampled. An alternative possibility is that since the molecular machinery to synthesize, package, and release different neurotransmitter types require only a modest number of genes, other aspects of a neuron's identity such as its origin, lineage, projection phenotype, or function may play a more dominant role in configuring global patterns of gene expression.

First, we plotted the brain region of origin for each cell type (**Fig. 2; Fig. S7**). During embryogenesis, the neural tube forms vesicles that eventually give rise to mature cephalic subdivisions (telencephalon, diencephalon, mesencephalon, metencephalon, and myelencephalon). Major branches of the dendrogram largely partition by cephalic vesicle: most telencephalic types (neocortex, amygdala, hippocampus, striatum) cluster distinctly from diencephalic and hindbrain types, suggesting that developmental lineage plays a major role in shaping the transcriptional identity of adult primate brain cell types. An exception are the basal forebrain types; despite being part of the telencephalon, basal forebrain neurons largely intermingle with hypothalamic types, suggesting closer transcriptional similarity of two structures of distinct developmental origins. Another exception are the cholinergic neurons. Despite residing in multiple cephalic partitions, cholinergic (*CHAT*+) types largely form a single clade, including forebrain GABAergic types and a hypothalamic type that expresses *SLC17A8* (*VGLUT3*) (**Fig. 2B**). (One small population of *CHAT*+ neurons in thalamus instead joined a clade consisting of thalamic glutamatergic types and brainstem GABAergic types.)

Strong distinctions between transcriptomic identities of GABAergic and glutamatergic neurons appear to be exception rather than rule. For example, neuronal types expressing genes that encode one or more of the vesicular glutamate transporters (*SLC17A6, SLC17A7, SLC17A8*) were not transcriptionally similar to one another when taking into account all expressed genes. Similarly, neuronal types expressing *GAD1* (or *GAD2*) were widely dispersed across the tree (**Fig. 2B**). Within parts of the tree dominated by non-telencephalic structures (hypothalamus, basal forebrain, thalamus, brainstem, cerebellum), proximal clades or even adjacent leaf nodes frequently intermixed GABAergic and glutamatergic types (**Fig. 2D,E**); such intermixing is not observed in mouse in these structures (*7*). In marmoset, only within the telencephalon was there strong segregation of neuronal clades based on their glutamatergic or GABAergic neurotransmitter utilization.

Even within the telencephalon, excitatory/inhibitory identity did not predict transcriptional similarity between major clades: GABAergic projection neurons such as the medium spiny projection neurons (MSNs) of the striatum were transcriptionally closer to glutamatergic neocortical, amygdalar, and hippocampal neurons than they were to telencephalic GABAergic interneurons (**Fig. 2C**). Borrowing

from ancestral state reconstruction methods (*25*), we applied a maximum likelihood-based approach (fastAnc) to the global gene expression of cell types and the dendrogram of their similarity to infer the transcriptomic state of internal nodes of the tree. This enabled comparisons of leaf nodes to internal nodes as well as internal nodes to each other. We compared gene expression in the parent node of the clade containing striatal MSNs as well as basal forebrain, hypothalamic, and amygdala *FOXP2*+ neurons with its neighbor, the parent node of telencephalic glutamatergic neurons, and found that transcription factors (TFs) were overrepresented among differentially expressed genes (X2 = 24.1, p < 8e-7). In contrast, TFs were not overrepresented in differential gene expression between the parent nodes of telencephalic glutamatergic neurons and telencephalic GABAergic neurons (X2 = 2.3, p = 0.12; **Fig. S8A**).

Although glutamatergic neurons from distinct cephalic origins do not cluster together, maintaining glutamatergic neurotransmission 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. 2F**). Surprisingly few genes are strongly positively correlated with both *SLC17A6* (VGLUT2) and *SLC17A7* (*VGLUT1*) expression, even those associated with glutamate synthesis and packaging (**Fig. S8B**). 116 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. 2G**). Only a few genes correlated above 0.5 to both *SLC17A6* and *SLC17A7,* including *BDNF, NRN1*, and *TAFA1.* Moreover, only 10 genes (*ARPP21, BDNF, CACNA2D1, CHN1, CHST8, CPNE4, LDB2, NRN1, PTPRK, TAFA1*) are differentially expressed (> 2.5 fold change) in both *SLC17A6*+ glutamatergic neurons and *SLC17A7*+ glutamatergic neurons relative to *GAD1*+ neurons.

Expression of *SLC5A7* (solute transporter for choline) and *NTRK1* (a neurotrophic tyrosine kinase receptor necessary for cholinergic neuron development) were among the genes most highly correlated with *CHAT* expression. The distribution of pairwise correlations to *CHAT* had a lower standard deviation (mean r = 0.002, std. dev = 0.116) relative to baseline gene-gene correlations (mean r = 0.02, std. dev = 0.199; F(5078, 17315809) = 0.33, p-adj < 1e-15). In contrast, pairwise correlations to *SLC17A7* were much broader relative to the background distribution of all gene-gene correlations (**Fig. 2F**). We used factor analysis to search for latent factor(s) associated with primary neurotransmitter status. Although this analysis reveals factors and gene loadings associated with neighborhood structure on the dendrogram (**Fig. S9**), no factors distinguish GABAergic from glutamatergic types. The bulk of gene expression in glutamatergic neurons appears incidental to excitatory neurotransmission itself.

Neurons are imprinted by their region of origin

While some cells can migrate long distances from proliferative zones to their mature destinations (*26*, *27*), neurons generally tend to respect cephalic boundaries and remain within the same subdivision as their progenitors. Mechanisms for restricting within-cephalic migration have also been revealed, demonstrating the tight control over migration potential (*28*). While cephalic boundary crossings do exist (*29–31*), they are believed to be relatively rare. It is not known to what extent mature 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.

We inspected the neuronal dendrogram for evidence of cephalic crossings. Some tissues gave rise to cell types that exclusively clustered within their cephalic domain (**Fig. 2C-E**). For example, hippocampal cell

types all remained within the telencephalic partition of the dendrogram. Similarly, cerebellar cell types were restricted to a single clade along with two GABAergic brainstem types (**Fig. 2D**, **Fig. S7**). Only two neocortical neuron types joined clades outside of the major telencephalic branch: (1) the long-range projecting *CHODL*+ GABAergic neurons (*14*), which formed a clade with other strongly *NOS1*+, *NPY*+ neurons and which collectively were more similar to hypothalamic and basal forebrain types, and (2) Cajal-Retzius neurons. Surprisingly, despite evidence that primate Cajal-Retzius cells mostly originate from telencephalic sources (*32*), Cajal-Retzius (C-R *LHX9*) neurons were transcriptionally more similar to *OTX2*+ GABAergic types found in thalamus, hypothalamus, brainstem, and basal forebrain, and not to telencephalic types (**Fig. S7**). Consistent with previous work suggesting that mammalian thalamus contains both midbrain-derived and forebrain-derived GABAergic interneurons (*31*), GABAergic, *OTX2*+ thalamic neurons joined a clade containing brainstem GABAergic neurons, while *OTX2*- GABAergic thalamic populations joined a clade with forebrain neurons (**Fig. S7**). Within the major telencephalon branch, a clade of GABAergic projection neurons including striatal MSNs, amygdala, and basal forebrain *GAD1+, FOXP2+* neurons also contained several subtypes of hypothalamic *GAD1+, FOXP2+* neurons. (Interestingly, this clade includes the recently described eccentric medium spiny neurons, eSPNs(*6*). In mice eSPNs form equal proportions of *Drd1*+ and *Drd2*- subtypes, but in marmosets the *DRD1*+ population is far more abundant.)

The amygdala, a telencephalic structure, contains several cell types that have greater transcriptomic similarity to extra-telencephalic types. The amygdala is a loosely associated, functionally diverse collection of nuclei whose cells have diverse phylogenetic and developmental origins (*33*, *34*). The basolateral amygdala, which contains the largest proportion of excitatory neurons, is thought to contain cell types with similar properties as cortical and claustrum neurons, while intercalated nuclei containing inhibitory *FOXP2*+ projection neurons share developmental origins with some striatal GABAergic populations such as MSNs (*5*, *12*, *35*). Consistent with expectations, glutamatergic amygdalar neurons associated with cortical and hippocampal glutamatergic neurons, while *FOXP2*+ GABAergic amygdalar neurons joined the clade containing striatal spiny projection neuron types, consistent with recent lineage tracing of these populations in mouse (*5*), (as well as *FOXP2*+ types in basal forebrain and hypothalamus as described above) (**Fig. 2C**).

Unexpectedly, three amygdalar neuron types joined a clade not with other telencephalic neurons, but rather with *SLC17A6*+ hypothalamic and basal forebrain types (**Fig. 3A**, **Fig. S7**). When clustered with amygdala neurons alone, they associated with amygdalar GABAergic, not glutamatergic types (**Fig. 3B**), although they expressed *SLC17A6* and lacked expression of *GAD1* and *GAD2* (**Fig. 3C**). They also lacked expression of other genes needed for GABA synthesis such as *SLC32A1* (VGAT), and lacked the molecular machinery for non-canonical GABA reuptake or release that has been observed in other cell populations (*36*, *37*). Thus, these amygdala neurons display a "cryptic" transcriptomic identity: they are glutamatergic, but relative to other telencephalic residents, have transcriptomic profiles more similar to GABAergic types.

"Cryptic" amygdalar subtypes exhibited additional unusual transcriptomic features relative to other amygdala neuronal types, such as sparse expression of *OTP* and *SIM1* (**Fig. 3C**). These transcription factors are expressed in neuronal lineages arising from proliferative zones around the 3rd ventricle. In mice, there is a migratory stream of diencephalic neurons into telencephalon, specifically into the medial

amygdala (*30*, *38*). This migration depends on expression of *Otp (30)*, suggesting that the cryptic "GABAergic-like" glutamatergic cells we found are part of the diencephalic lineage previously identified in mouse. We acquired mouse amygdala snRNA-seq data (53,745 nuclei) and confirmed that the homologous cell type in mouse clusters with GABAergic neurons, and expresses *Slc17a6*, but not *Gad1* or *Gad2.* (**Fig. 3D**). In mice, cryptic neurons comprise the majority population in amygdala that express *Adcyap1* (PACAP) (**Fig. 3D**), a neuropeptide expressed extensively (but not exclusively) in hypothalamic populations (*39*), and that is associated with energy homeostasis (*40*) as well as stress and anxiety (*41*) and immune responses (*42*). (Unlike mice, in marmosets *ADCYAP1* is additionally expressed in subsets of *SLC17A7*+ neurons that do not share the "cryptic" phenotype, indicating that *ADCYAP1* has a distinct distribution of expression in amygdala neurons in mice and marmosets.)

To localize the specific amygdalar nuclei that harbors the cryptic population, we examined expression of *SIM1* in neonatal marmoset (*43*, *44*). *SIM1* expression localizes to MeA (https://gene-atlas.brainminds.jp; **Fig. 3E, Table S3**), consistent with migration of *Sim1*+ neurons from the diencephalon in mice (*30*). Thus, the cryptic amygdala neurons are a conserved population in mouse and primate that likely have diencephalic origins, and that retain transcriptional identities more in common with diencephalic types than with the telencephalic types with which they reside.

Striatal interneuron types are distributed in medial-lateral gradients

The striatum – composed of caudate, putamen, and the nucleus accumbens – displays complex functional topography. In humans, anterio-medial 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 (*45*). In mice, *Chat* interneuron distribution increases dorsally and anteriorly (*46*), *Pvalb* interneurons are more abundant in dorsolateral striatum than in dorsomedial striatum, and *Sst*+ interneurons are spatially homogeneous (*47*). A systematic quantification of striatal interneuron types has not been performed in a primate, and it is unknown if they follow similar or distinct spatial distributions.

Primates retain the major populations of striatal interneurons found in mice (*6*, *48*), and additionally have gained a novel type distinguished by *TAC3* expression *(3, 12)* (**Fig. 4A**). We used single-molecule FISH (smFISH) to investigate distributions of the major types of conserved striatal interneurons (*SST+, PVALB+, CHAT+*, *TH+, CCK+,* probes in **Table S1**) in serial sagittal sections of marmoset striatum (**Fig. 4C-G**). In total, we quantified 5,778 striatal interneurons across 32 sections, and expressed interneuron proportions as a percent of all cells (DAPI+). Each series sampled sagittal sections ~160 µm apart, beginning 1,184-1,584 µm from the midline up to 6,384 µm laterally, including the majority of the striatum with the exception of the most-lateral portion of the putamen.

Each striatal interneuron population exhibited a non-uniform distribution across the marmoset striatum, particularly in the medial-lateral axis. Similar to mice, the proportion of striatal *PVALB*+ interneurons increases in lateral sections, from ~0% to 0.8% of all cells (**Fig. 4D, Data S2**). Unlike mice, marmoset *SST*+ interneuron distribution is non-uniform, appearing sparse near the midline and increasing in proportion (0.1-0.7%; **Fig. 4C, Data S2**). Cholinergic neurons (*CHAT*+) show the opposite medial-lateral trend (0.45%-0.1%; **Fig. 4E, Data S2**). Similar to *CHAT*+ neurons, *TH*+ striatal interneurons, which are transcriptionally similar to the *PVALB*+ type, exhibited a decreasing medial-lateral gradient (**Fig. 4F,**

Data S2). *CCK*+ striatal interneurons, which are a minority population in marmoset (*3*) and mouse (*48*) are enriched close to the midline and become much sparser laterally (**Fig. 4G, 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 (**Fig. S10**). *DRD1* and *DRD2*, which distinguish direct and indirect MSNs, respectively, had slight enrichments medially but otherwise exhibited largely uniform distributions across the striatum (**Fig. S11**).

To capture the morphology of molecularly-defined striatal interneuron types, we used a reporter AAV under the control of the forebrain interneuron-specific mDlx enhancer (*8*). We performed sparse labeling via systemic injections in marmosets and used smFISH on thick sections (120 µm) to simultaneously capture local morphology of labeled neurons and confirm the molecular identity of GFP+ cells. In total, we imaged 1,203 GFP+ neurons in the striatum, and, using combinations of 2-3 probes for marker genes of different types, were able to molecularly identify 48 with smFISH.

Three-dimensional (3D) reconstructions and tracing (Imaris) of the most complete instances of these cell subtypes (**Fig. 4C-G**) were used to calculate morphological parameters of each molecular type within the set of mDlx-AAV-GFP labeled neurons (**Table S2**). 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 processes based on the diameters of the soma and the thinnest projection of the cell. While the cholinergic *SLC5A7*+ cell is the largest striatal interneuron subtype based on volume, surface area, length, and total dendrite area, it displays the fewest dendritic branches and branch points. Interestingly, the *CCK*+ interneuron has the same number of primary dendritic branches as the *SLC5A7*+ cell and over 10-fold more dendritic branch points. The *TH*+ striatal interneuron has the highest number of branch points compared to the other subtypes (**Table S2**).

Primate-specific striatal TAC3+ interneurons are similar to diencephalic types

Previously, we discovered a *TAC3*+, *LHX6*+ interneuron subtype in the striatum of humans, macaques, and marmosets, that was absent from the striatum in mice and ferrets (*3*). The striatal *TAC3*+ type comprised ~30% of all interneurons in the primate and was detected in snRNA-seq datasets targeting nucleus accumbens, caudate, and putamen. However, this cell type's spatial distribution within the striatum has not been determined. We used smFISH with probes for *TAC3*, *DRD1*, and *DRD2* in sagittal sections of marmoset striatum. snRNA-seq data indicated that *TAC3* is exclusively expressed in this interneuron type in caudate and putamen, and is additionally expressed in a minority population of medium spiny neuron (MSN)-like neurons in nucleus accumbens (< 0.3% of MSNs). The *TAC3*+ interneuron type expresses *DRD2*, while the *TAC3*+ MSN-like type expresses *DRD1*, enabling the populations to be distinguished. Quantification of the smFISH images revealed that the *TAC3*+ type was present throughout the extent of the sampled sections, and showed a clear medial-lateral density gradient in striatum (**Fig. 5A, Data S3**). The same spatial trend is seen in *CHAT*+, *TH*+ and *CCK*+ interneurons (**Fig. 4E-G**) but not *SST*+ or *PVALB*+ striatal interneuron types, which show the opposite gradient (**Fig. 4C,D**).

We next investigated whether *TAC3* interneurons are transcriptionally unique relative to other cell types in the marmoset brain. We previously concluded that the primate striatal *TAC3*+ type was most similar to other, MGE-derived striatal interneurons such as the *PVALB*+ type, and that no homologous type was

present in hippocampus or neocortex, the two other telencephalic structures that we had sampled at the time (*3*). Here, our broader census of subcortical brain structures enabled us to revisit whether this novel type is similar to other, extra-telencephalic neuronal populations in the marmoset brain that were not originally sampled in our previous work. We first examined *TAC3* expression using smFISH across whole sagittal sections, finding that ventral striatum/nucleus accumbens neurons have high per-cell expression of *TAC3* relative to cortex and dorsal striatum, and that medial prefrontal/frontal polar cortex has the highest density of *TAC3*+ neurons in neocortex (**Fig. 5B**).

Our snRNA-seq data indicated that *TAC3* is expressed in 20 different cell types, including expected expression in cortical GABAergic neurons as well as several amygdala, basal forebrain, thalamic, and hippocampal types. However, most of these other *TAC3*+ populations are not transcriptionally similar to the *TAC3*+ striatal type and are distributed widely across the dendrogram (**Fig. 5C**). Further, the *TAC3*+ striatal type was most similar not to other striatal interneuron types (as we had previously concluded (*3*)), but rather to two distinct GABAergic types present in the basal forebrain and hypothalamus (**Fig. 5C,D**). Intriguingly, these populations are also *TAC3*+ and *LHX6*+ (**Fig. 5D**). Each of these three *TAC3*+ populations had qualitative and quantitative gene expression differences, ruling out contamination from dissection (**Fig. 5D**). For example, the population in basal forebrain expresses *AVP* (*AVP* is also detected in other basal forebrain subtypes); *DRD2* expression is uniquely present in the striatal population, and absent in the other basal forebrain and hypothalamus populations. Our *TAC3* smFISH experiments suggest the non-striatal populations could be localized to lateral hypothalamus, lateral preoptic area, ventrolateral hypothalamus, and the bed nucleus of the stria terminalis, though more comprehensive sampling may reveal additional populations within other subcortical nuclei (**Fig. 5B**).

Considering their unexpected transcriptional similarity to a telencephalic (basal forebrain) and a diencephalic (hypothalamus) type, the *TAC3*+ striatal type may either arise from a telencephalic progenitor (*12*) that also gives rise to diencephalic types, or else shows striking transcriptional convergence with diencephalic types that have distinct developmental origins. Favoring telencephalic origin, all three *TAC3*+ types are *FOXG1*+ (*49*).

Neocortical interneuron types have highly focal biodistributions

In the mouse, densities of *Sst*+, *Vip*+, and *Pvalb*+ cortical interneurons vary by cortical area (*50*). In primates, overall neuron densities vary by as much as 5 fold across the cortical sheet, with highest neuron proportions and densities found in occipital cortex and particularly in V1 (*51*). However, quantitative mapping of subtypes of neurons across large cortical volumes has not been performed in a primate, and it is not known if related subtypes that occupy different brain structures follow the same or different spatial distributions. Major classes of neocortical interneurons (*PVALB+, SST+, VIP+, LAMP5+*) are transcriptionally similar to types from the hippocampus, striatum, and amygdala that expressed the same markers (**Fig. 6A**), consistent with their shared ventral telencephalic origin (*26*, *52*). One possibility is that homologous, developmentally linked populations of interneurons in cortex and striatum show similar broad spatial distributions across structures. To examine this possibility, we imaged and quantified spatial distributions of the major cortical interneuron types in the same sagittal sections prepared for striatum (**Fig. 6B-G**).

In total, we quantified 247,596 neocortical interneurons by smFISH across 16 sagittal sections, some of which were also processed for the striatum. We expressed interneuron proportions as a percent of all cells (DAPI+) with probes for *SST, PVALB, CXCL14, VIP* and *LAMP5* (**Table S1**), which collectively account for all major cortical interneuron populations **(Fig. 6A**). *CXCL14* is a marker for caudal ganglionic eminence derived cortical interneurons. It is expressed in most *VIP*+ and *LAMP5*+ cortical neurons, as well as a smaller population of *VIP*-, *LAMP5*- types, some of which are *PAX6*+ (and which correspond to the *SNCG*+ population in humans (*2*, *3*)). *VIP*+ and *LAMP5*+ interneurons are the two other major CGE-derived populations present in primates and mice. As *LAMP5* is also expressed in subsets of excitatory neurons, we performed dual labeling smFISH with *LAMP5* and *GAD1* to avoid counting glutamatergic types.

Developmentally linked interneuron populations in striatum and neocortex (e.g *PVALB*+ types in both structures) displayed distinct spatial distributions. While all interneuron subtype distributions in the striatum followed a gradient along a primary axis (either medial-lateral or lateral-medial), the interneurons in the neocortex followed much more complex distributions that for the most part were not captured by simple gradients. For example, while striatal *PVALB*+ interneurons show a clear medial-lateral gradient (**Fig. 4D**), neocortical *PVALB*+ interneurons are typified by strong enrichments in the occipital lobe, particularly along the calcarine sulcus in the medial sections, as well as the occipital pole more laterally (**Fig. 6D; Fig. S13, Data S1, Data S2**). There is also an enrichment of *PVALB*+ neurons in lateral aspects of PFC (**Fig. 6D**). *SST*+ interneurons in the neocortex increase medio-laterally (**Fig. 6C**), as in striatum (**Fig. 4C**), but closer inspection reveals 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 prefrontal cortex.

CXCL14+ neurons are enriched along the calcarine sulcus medially, as well as in ventral aspects of the occipital cortex more laterally including V1, V2 and V3. There are also higher proportions dorsomedially in the parietal cortex (**Fig. 6E**). *VIP*+ neurons were enriched in medial PFC, including rostral/subgenual anterior cingulate and ventromedial/orbitomedial cortex at the midline, as well as somatosensory cortex, posterior parietal cortex and the most medial aspect of the calcarine sulcus. High proportional densities of *VIP*+ cells were observed laterally at or near somatosensory cortex and posterior parietal cortex, as well as medial occipital cortex. Medial frontal areas also had relatively high enrichment of *VIP*+ neurons, including area 32, 10, 14 and 25 (**Fig. 6F)**. *LAMP5+, GAD1+* interneurons were enriched in occipital regions and in the frontal pole. The upper layer bias typical of *LAMP5*+ interneurons is evident in lateral slices (*53*) but not in the most medial sections (**Fig. 6G)**.

To capture cortical interneuron morphology, we imaged neocortical GFP+ cells from the same marmosets that had received systemic mDlx-AAV-GFP injections for striatal interneuron characterization. We imaged 4,374 GFP+ neurons in neocortex, and, using combinations of 2-3 probes for marker genes of different types, were able to molecularly identify 235 based on their expression of the major neocortical subtype markers (*SST*, *PVALB*, *CXCL14*, *VIP*, *LAMP5*) with smFISH (**Fig. 6B**). A complete set of 216 telencephalic GABAergic neuron reconstructions (**Fig. S14**) using the NeuTube pipeline are available for download at https://www.brainimagelibrary.org/. As with the striatum, we used Imaris to reconstruct the most complete exemplars of neocortical interneurons (**Fig. S12**). From this subset, we did not observe

differences in soma diameter or other morphological parameters (**Table S2**), though larger sample sizes of fully reconstructed cells are likely needed to overcome interindividual and within-region variability.

Discussion

Single-nucleus RNA sequencing of over 2.4 million brain cells across 16 brain regions in the marmoset revealed that lineage, not neurotransmitter utilization, is the primary factor shaping adult transcriptomic identity in the primate brain. Quantitative smFISH revealed, for the first time in a primate, the spatial distributions of molecularly-resolved major GABAergic interneuron types. Different types have different spatial biodistributions within a brain structure, and the same subtype (e.g. *PVALB*+ GABAergic neurons) shows distinct spatial gradients in striatum and cortex. We used AAV reporters to reconstruct local morphologies of GABAergic interneuron subtypes, including in the recently discovered *TAC3*+ striatal type.

Within telencephalic structures, there is strong partitioning of glutamatergic and GABAergic neurons in mammals as well as in amphibians (*20*), suggesting an evolutionarily conserved distinction. In mouse, glutamatergic and GABAergic neurons from diencephalic and midbrain structures also partition almost perfectly by neurotransmitter usage (*7*). However, in marmoset, non-telencephalic glutamatergic and GABAergic types form highly intermixed clades (**Fig. 2**).

Our results suggest that transcriptomic distinctions between glutamatergic and GABAergic neurons do not hold for neurons in other brain structures. This has implications for generalizing transcriptomic associations to other phenotypes. For example, transcriptomic changes in glutamatergic or GABAergic neurons associate to diseases such as autism and schizophrenia (*54–56*). 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 (*57*).

The observation that telencephalic glutamatergic and GABAergic neurons are more similar to each other than to extra-telencephalic counterparts raises questions about their evolutionary and developmental origins. A previous analysis, based on shared patterns of a small number of key transcription factors, proposed that telencephalic GABAergic neurons are developmentally and evolutionarily related to diencephalic GABAergic neurons (*58*). Our results suggest that, in primates, a limited number of telencephalic GABAergic types are transcriptionally similar to diencephalic types, and may largely arise from cephalic boundary crossings or phenotypic convergence. Most neocortical, hippocampal, and some amygdalar and striatal GABAergic types are so distinct from diencephalic GABAergic types that they share more similarities with telencephalic glutamatergic types.

Recent advances in lineage tracing shed new light on relationships between transcriptionally defined cell types (*5*, *59*). Bandler et al. provide evidence that transcriptomic convergence, whereby adult cell types converge on similar transcriptomic identities despite a non-shared developmental origin, may be widespread. Delgado et al. show that human cortical progenitors have the capacity to generate both excitatory and GABAergic neurons. Though these human experiments were performed in culture, the remarkable multipotency of human (but apparently not mouse (*5*)) cortical progenitors suggests another potential route whereby some subtypes of glutamatergic and GABAergic telencephalic neurons could share details in their transcriptomic programs relative to extra-telencephalic types in adulthood.

Our data support known instances of cross-cephalic migrations in thalamus and amygdala, and also suggest several new ones. The similarity of the primate-specific *TAC3*+ striatal type to hypothalamic and basal forebrain *TAC3*+ types in particular is unexpected. Recent work (*12*) suggests that the striatal *TAC3*+ 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 transcription factor necessary for ventral telencephalic fate (*49*), in all three transcriptionally similar *TAC3*+ populations.

Homologous striatal interneurons in mouse and marmoset follow distinct distributions. In marmoset, each population is distributed in some form of gradient along the medial-lateral axis. The spatial distribution of striatal *TAC3*+ interneurons is similar to the distribution of *CHAT*+ and *TH*+ neurons, and is distinct from *SST*+ and *PVALB*+ interneuron populations. Marmoset neocortical interneuron populations each displayed distinct distributions that defied simple gradients. The calcarine sulcus, which predominantly consists of V1 (and parts of V2) functionally represents foveal vision and had relatively high densities of several interneuron subtypes, in keeping with extraordinarily high overall neuron densities in primates (*51*), as did some extra-striate visual areas, somatosensory cortex, medial prefrontal cortex and cingulate cortex.

Morphological characterization of the striatal and neocortical interneuron populations suggests variation in overall size and dendritic arborization amongst subtypes. Appreciating the significance of these morphological differences will require studying how these cells affect the function of circuits they reside in (i.e., cellular/subcellular targeting biases and functional properties (*60–62*).

The development of virally-based tools enables cell-type-specific access in nonhuman primates (*2*, *8*, *63*, *64*). To maximize translatability across species, previous approaches have nominated candidates using evolutionary conservation (usually between mice and humans) as a selection criterion. However, the evolutionarily conserved mDlx enhancer did not drive expression in the novel striatal *TAC3*+ type. We observed that the mDlx enhancer selectively under-ascertained several interneuron populations in marmoset in these experiments. 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 underlabeling could in part be due to the titer and systemic delivery approach we adopted, which was necessary in order to achieve sparse labeling for morphology. The overall low efficiency in our ability to molecularly characterize GFP positive cells suggests a need for further optimization for this species and application. The development and refinement of additional viruses constrained by cell type specific regulatory elements will further expand the toolbox for nonhuman primate research (*63*, *64*), and will elucidate the principles governing brain region, cell type, and species regulatory specificity.

METHODS

Animals used for study

Marmosets. Marmosets were pair-housed in spacious holding rooms with environmental control of temperature (23–28°C), humidity (40–72%), and 12 hr 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 MIT 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 (3–5 mice per cage) on a 12:12 hr 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

Marmoset specimens for snRNA-seq. Marmoset experiments were approved by and in accordance with Massachusetts Institute of Technology CAC protocol number 051705020. Adult marmosets (1.5–2.5 years old, 10 individuals) were deeply sedated by intramuscular injection of ketamine (20–40 mg kg−1) or alfaxalone (5–10 mg kg−1), followed by intravenous injection of sodium pentobarbital (10–30 mg kg−1). When the pedal with-drawal reflex was eliminated and/or the respiratory rate was diminished, animals were trans-cardially perfused with ice-cold sucrose-HEPES buffer (*3*, *6*). 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 (*3*, *6*). All regions were dissected using a marmoset atlas as reference (*65*), and were snap-frozen in liquid nitrogen or dry ice-cooled isopentane, and stored in individual microcentrifuge tubes at −80 °C.

Mouse specimens for snRNA-seq. Three adult (P80-90) male wild-type mice were deeply anesthetized with isoflurane and sacrificed by decapitation. Brains were quickly excised, washed in ice-cold sterile 0.1M phosphate buffer 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 S3**) 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 1.5ml micro-centrifuge tubes, snap-frozen on dry ice, and stored at -80 °C until used.

Marmoset specimens for smFISH. One 3-year-old female marmoset was deeply sedated by intramuscular injection of ketamine (20–40 mg kg−1) or alfaxalone (5–10 mg kg−1), followed by intravenous injection of sodium pentobarbital (10–30 mg kg−1). When the pedal with-drawal reflex was eliminated and/or the respiratory rate was diminished, animals were trans-cardially 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 16µm, adhered to SuperFrost Plus microscope slides (VWR, 48311-703), and stored at -80 °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 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 S3**) or the Advanced Cell Diagnostics RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (ACD, 323100, https://acdbio.com/sites/default/files/USM-323100%20Multiplex%20Fluorescent%20v2%20User%20Manual_10282019_0.pdf, **Table S3**) protocol.

Single nucleus assays, library preparation, sequencing

10x RNA-seq. Unsorted single-nucleus suspensions from frozen marmoset and mouse samples were generated as in (*66*). GEM generation and library preparation followed the manufacturer's protocol (10X Chromium single-cell 3′ v.3, protocol version #CG000183_ ChromiumSingleCell3′_v3_UG_Rev-A).

Raw sequencing reads were aligned to the 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 7,262 UMIs per cell.

RNA sequencing data processing, curation, clustering

Processing and alignment steps follow those outlined in: https://github.com/broadinstitute/Drop-seq/ (**Table S3)**. 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 TSO and polyA trimming. Reads were aligned using STAR, and tagged with their gene mapping (exonic, intronic, or UTR) and function (strand). The reads were then processed through GATK BQSR, and tabulated in a digital gene expression matrix (DGE) 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 (*67*), % 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 (*13*). Detection of cell-cell doublets was performed using a two-step process based on the R package DoubletFinder (*68*). DoubletFinder implements a nearest-neighbors approach to doublet detection. First, artificial doublets are simulated from input single-cell data and are co-clustered 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. Because doublet libraries are initially categorized as true libraries in the nearest-neighbor search, we find 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. Non-telencephalic neurons were clustered by brain structure. All clusterings were performed in two stages: first-round clustering was based on the top 60 independent components (ICs) and a default resolution (res) of 0.1, nearest neighbors (nn) = 25. Following the process outlined in (*6*), 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 between the final number of clusters and the number of included ICs. 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.

Clustering, annotation, and integration of macroglial nuclei

Macroglial nuclei were processed using custom Python scripts in scanpy (*69*). Following cell-selection, background removal using CellBender, and cluster annotation, putative macroglia were subjected to additional bioinformatic quality control measures were performed to eliminate low-quality nuclei, cellular debris, and doublets. All cells containing fewer than 300 genes and all genes detected in fewer than 10 cells were eliminated. Nuclei with high proportions of genes associated with unhealthy or dying cells, such as heatshock proteins and tubulins, were also filtered out. Doublets were automatically identified and removed using Scrublet (*70*). Nuclei called as doublets were visualized in a two-dimensional embedding to enable manual heuristic-based confirmation of reasonable doublet calling.

Macroglia (astrocytes, oligodendrocytes, and OPCs) were classified and subsetted from each marmoset cortical region separately through preliminary clustering using primarily default scanpy methods. Briefly, gene expression counts were normalized on a per-cell basis to counts per million and logarithmized (natural log + 1). Highly variable genes (*71*) were annotated and used for subsequent dimensionality reduction and clustering steps. Each gene in the digital expression matrix (DGE) was scaled to unit variance and zero mean, and Principal Components Analysis (PCA) was performed. Distances between cells were then computed from the PCA embedding with Uniform Manifold Approximation and Projection (UMAP), resulting in a weighted adjacency matrix of the neighborhood graph of observations (*72*). Cells were plotted and visualized with the 2-dimensional UMAP coordinates, and clustered into groups using the Leiden algorithm (*73*). Clusters were assigned to major cell classes (astrocyte, oligodendrocyte, microglia, polydendrocyte, neuron, vasculature) upon reference to canonical marker genes for neurons and glia, which have been described previously (*6*). Macroglia-specific DGE matrices were then subsetted on the basis of cell type assignment and used for subsequent cross-species and cross-region analyses. To account for donor batch effects, macroglial nuclei from each donor were integrated using Seurat scTransform (*74*), with donor ID as the integration variable and 3,000 features for integration.

Following Seurat integration, astrocytes and oligodendrocyte lineage cells were subsetted and re-clustered separately for annotation. UMAP coordinates and the neighborhood graph were re-computed using the PCA embedding from the scTransform integration. Nuclei were then clusters using the Leiden algorithm with the resolution parameters determined as follows. First, a high resolution parameter (0.5 or above) was chosen to intentionally over-cluster the nuclei. Marker (top differentially expressed) genes for resulting clusters were determined using scanpy's "rank_genes_groups" function using a Wilcoxon rank-sum test. For each cluster, we counted the number of marker genes with log fold-change (logFC) > 2. If 6 or more marker genes with logFC > 2 existed for all clusters, we then checked for uniqueness of the marker genes between clusters. If any cluster had fewer than 6 unique marker genes, the clustering resolution parameter was decreased by 0.05 and the clustering procedure was repeated until the log fold-change and uniqueness requirements were met. Resulting macroglial clusters were annotated according to their cell class (e.g. astrocyte, oligodendrocyte, OPC) and top marker gene.

Identification of cluster differentially expressed genes (DEGs)

To create the dotplots in **Fig. 1F,G**, clusters for each cortical cell type plotted (GLUT, GAD, AST, and OL) were randomly downsampled to 5,000 nuclei. DEGs were identified from logCPM expression values using scanpy's "rank_genes_groups" function using the Wilcoxon rank-sum method, grouping by cluster. DEGs were then filtered according to the following criteria: minimum in-group expression fraction of 0.25, minimum log base 2 fold-change (logFC) of 1.661, and maximum out-group expression fraction of 0.5. We then examined top 10-30 rDEGS, ranked according to scanpy output, and chose to plot the 1-2 genes passing the filter with the highest logFC. A few hand-picked genes, selected based on prior knowledge and field standards, were included. Dotplots were created in scanpy using downsampled expression matrices with logCPM values.

Identification of regionally differentially expressed genes (rDEGs)

To nominate regionally differentially expressed genes (rDEGs) between cortical areas, normalized metacells are log10 transformed and pairwise differences are examined. Genes with > 3 fold log10 difference between two regions are considered rDEGs. Dotplots for rDEG expression and associated dendrograms were created using "scanpy.pl.dotplot", grouped by region and organized by branch on the dendrogram. Each dot represents the mean expression within each region (visualized by color) and fraction of cells in each cluster expressing the gene of interest (visualized by the size of the dot). The dendrogram was calculated using scanpy's "tl.dendrogram" function using the log base 2 counts per million expression values of the top 1,000 most variably expressed genes, calculated using scanpy's "pp.highly_variable_genes" function.

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. This 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 MSN-like GABAergic projection neurons, the reconstructed parent node of striatal MSNs had higher expression of known markers of striatal projection neurons such as *DACH1*.

PEER factor analysis

We applied probabilistic estimation of expression residuals, (PEER, (*75*)), a Bayesian factor analysis method, to the set of 288 scaled, log-normalized neuronal metacells. PEER produces learned factor sets, their effects on each gene and a residual data set of the expression values after subtracting each factor contribution. We examined a range of factors (K = 25-40) which produced similar results; presented results (**Fig. S9**) use K = 40 factors. For each factor, we output a matrix of weights (gene loadings on factors) as well as the residual expression matrix. To visualize the latent factors on the original dendrogram depicted in Fig. 2A, we used fastAnc, treating each factor as a trait.

Spatial smFISH experiments

All probes are listed in **Table S1**. 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. A three year old, adult male marmoset was euthanized and perfused with saline. The brain was removed, embedded rapidly in OCT, and stored in the -80C freezer. Tissue was then cut to 16µm on a cryostat and stored in the -80C until needed. In situ hybridization was performed for genes of interest (see Table S1) with HCR or ACD antisense probes, incubated with TrueBlack Lipofuscin Autofluorescence Quencher (Biotium, 23007) for 10-30 seconds (16µm thick tissue) and 3-5 minutes (120µm thick tissue) 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×/0.8 NA air objective for ACD stains or a 40×/1.2 NA water-immersion objective for HCR stains.

Series images were segmented using StrataQuest, a software package from TissueGnostics, that 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 ROIs were manually drawn to avoid certain areas of white matter and large artifacts and autofluorescence before computing intensity and other statistical and morphological measurements (20 parameters) for each channel.

Segmented cells were further analyzed using in-house code (https://github.com/klevando/BICCN-StrataQuest-Script, **Table S3**). Specifically, 50 cells were hand-labeled as positive or negative for the markers of interest in order to identify the appropriate threshold for feature selection using the parameters that best discriminated this binary. Parameters included mean intensity, maximum intensity, standard deviation of intensity, and range of intensity, and equivalent diameter. These filters were then applied to the unlabeled data to identify positive cells. Spatial locations of the cells were visualized by plotting the x-y coordinates associated with each nuclei. These were then binned in histograms across the x or y axis (corresponding to the rostrocaudal plane or to the dorsoventral plane respectively) and normalized to DAPI. A size of 30 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 normalized to the number of detected nuclei (total DAPI+). Normalizations across slices correspond to the mediolateral plane. Density plots were generated using the *dscatter* function in Matlab with a bin size calculated for x and y axes and across slices based on a given number of bins (100) for the length of x of the first slice. DAPI and marker of interest counts are available in **Data S1-S3**.

Morphology and smFISH experiments

AAV2/9-hDlx5/6-GFP-fGFP virus was generated as in (*8*). Virus was systemically IV injected (400ul-700ul at 1.7- 2.4^{10} titer) in 5 marmosets. The virus was allowed to incubate for approximately 2 months.

After systemic IV injection with AAV2/9-hDlx5/6-GFP-fGFP, marmosets were euthanized and perfused with saline followed by 4% paraformaldehyde (PFA). Brains were removed and 120µm sections were cut on a vibratome into 70% ethanol for storage at -20C. 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 2-fold increase in concentration of probe to hybridization buffer (Molecular Instruments, https://files.molecularinstruments.com/MI-Protocol-RNAFISH-GenericSolution-Rev8.pdf), for markers of interest (**Table S1**) that corresponded with RNA-seq defined clusters. The sections were then stained with anti-GFP antibody (Rabbit anti-GFP, Invitrogen, A11122) and a secondary antibody (Goat anti-Rabbit AF488, Thermo-Fisher, A-11034) to amplify the endogenous GFP signal (https://www.protocols.io/view/marmoset-nhp-free-floating-anti-gfp-antibody-stain-3byl47nb2lo5/v1, **Table S3**). Sections were incubated in TrueBlack (Biotium, 23007) for 3-5 minutes in order 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 in order to select GFP+ cells for imaging at 40× and to highlight the exact location of the cell. GFP+ cells were imaged for stained markers of interest.

Morphological reconstruction and feature quantification - Imaris. Without pre-processing 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+ 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 (**Fig. 4C-G** and **Fig. S12**), corresponding to each surface rendered image (data not shown), were generated using the Filament Tracing wizard in Imaris. Total number of primary dendritic branches, total dendrite area, total length and the number of dendrite branch points were 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 1 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 reported in **Table S2**.

Morphological reconstruction and feature quantification - NeuTube. Automatic or semi-automatic tracing algorithms are challenged by some data, perhaps due to the low SNR of a given image. Manually reconstructed the sparse neurons via Neutube(*76*) tracing software. With the software, we 1) Create 3D volume rendering of the GFP-AAV marmoset neuron , 2) use the signal transfer function (e.g., histogram equalization) for overall intensity and opacity values to optimize the signal-to-noise ratio by manually examining the clearest visualization of the dendrites, 3) Build the neuron skeleton over the 3D volume by tracing the processes, 4) Scan through the 3D volume to make sure no parts of the neuron are missed, 5) Double check the raw 2D images to see if any of the branches were not presented well in 3D due to volume rendering artifacts, 6) Label axon, dendrites and soma parts of the skeleton model. Reconstructed neurons were saved as SWC format.

References

1. S. Ma, M. Skarica, Q. Li, C. Xu, R. D. Risgaard, A. T. N. Tebbenkamp, X. Mato-Blanco, R. Kovner, Ž. Krsnik, 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-Ludiye, J. J. Ely, E. W. Daadi, D. Mi, M. Daadi, O. Marín, 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*, eabo7257 (2022).

2. 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. Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow, 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. Macosko, 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).

3. 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).

4. J.-P. Lin, H. M. Kelly, Y. Song, R. Kawaguchi, D. H. Geschwind, S. Jacobson, D. S. Reich, Microenvironment Impacts the Molecular Architecture and Interactivity of Resident Cells in Marmoset Brain. *bioRxiv* (2021), p. 2021.01.25.426385.

5. R. C. Bandler, I. Vitali, R. N. Delgado, M. C. Ho, E. Dvoretskova, 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 (2022).

6. A. Saunders, E. Z. Macosko, 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).

7. 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).

8. 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).

9. E. Boldog, T. E. Bakken, R. D. Hodge, M. Novotny, B. D. Aevermann, J. Baka, S. Bordé, J. L. Close, F. Diez-Fuertes, S.-L. Ding, N. Faragó, Á. K. Kocsis, B. Kovács, Z. Maltzer, J. M. McCorrison, J. A. Miller, G. Molnár, G. Oláh, A. Ozsvár, M. Rózsa, S. I. Shehata, K. A. Smith, S. M. Sunkin, D. N. Tran, P. Venepally, A. Wall, L. G. Puskás, P. Barzó, F. J. Steemers, N. J. Schork, R. H. Scheuermann, R. S. Lasken, E. S. Lein, G. Tamás, Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018).

10. 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. Hollt, B. P. Levi, S. I. Shehata, B. Aevermann, A. Beller, D. Bertagnolli, 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**, 1–38 (2019).

11. A. M. M. Sousa, Y. Zhu, M. A. Raghanti, R. R. Kitchen, M. Onorati, A. T. N. Tebbenkamp, B. Stutz, K. A. Meyer, M. Li, Y. I. Kawasawa, F. Liu, R. G. Perez, M. Mele, T. Carvalho, M. Skarica, F. O. Gulden, M. Pletikos, A. Shibata, A. R. Stephenson, M. K. Edler, J. J. Ely, J. D. Elsworth, T. L. Horvath, P. R. Hof, T. M. Hyde, J. E. Kleinman, D. R. Weinberger, M. Reimers, R. P. Lifton, S. M. Mane, J. P. Noonan, State, Matthew W, E. S. Lein, J. A. Knowles, T. Marques-Bonet, C. C. Sherwood, M. B. Gerstein, N. Šestan, Molecular and cellular reorganization of neural circuits in the human lineage. *Science*. **358**, 1027–1032 (2017).

12. 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).

13. 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).

14. B. Tasic, Z. Yao, L. T. Graybuck, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M. N. Economo, 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**, 1–41 (2018).

15. 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).

16. M. Baraban, S. Mensch, D. A. Lyons, Adaptive myelination from fish to man. *Brain Res.* **1641**, 149–161 (2016).

17. L. A. Akay, A. H. Effenberger, L.-H. Tsai, Cell of all trades: oligodendrocyte precursor cells in synaptic, vascular, and immune function. *Genes Dev.* **35**, 180–198 (2021).

18. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. D. Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Vawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. N. Grant, A. R. Jones, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. **489**, 391–399 (2012).

19. F. M. Krienen, B. T. T. Yeo, T. Ge, R. L. Buckner, C. C. Sherwood, Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proceedings of the National Academy of Sciences*. **113**, E469–E478 (2016).

20. J. Woych, A. Ortega Gurrola, A. Deryckere, E. C. B. Jaeger, E. Gumnit, G. Merello, J. Gu, A. Joven Araus, 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).

21. O. A. Bayraktar, L. C. Fuentealba, A. Alvarez-Buylla, D. H. Rowitch, Astrocyte development and heterogeneity. *Cold Spring Harb. Perspect. Biol.* **7**, a020362 (2014).

22. 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).

23. M. Y. Batiuk, A. Martirosyan, J. Wahis, F. de Vin, C. Marneffe, C. Kusserow, J. Koeppen, 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).

24. A. Kriegstein, A. Alvarez-Buylla, The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).

25. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*. **3**, 217–223 (2012).

26. 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).

27. 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).

28. 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).

29. K. Letinic, P. Rakic, Telencephalic origin of human thalamic GABAergic neurons. *Nat. Neurosci.* **4**, 931–936 (2001).

30. 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).

31. 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** (2021), doi:10.7554/eLife.59272.

32. G. Meyer, Building a human cortex: the evolutionary differentiation of Cajal-Retzius cells and the cortical hem. *J. Anat.* **217**, 334–343 (2010).

33. L. W. Swanson, G. D. Petrovich, What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).

34. 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).

35. 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. *NSC*. **166**, 220–230 (2010).

36. A. J. Granger, M. L. Wallace, B. L. Sabatini, ScienceDirect Multi-transmitter neurons in the mammalian central nervous system. *Curr. Opin. Neurobiol.* **45**, 85–91 (2017).

37. N. X. Tritsch, A. J. Granger, B. L. Sabatini, Mechanisms and functions of GABA co-release. *Nat. Genet.* **17**, 139–145 (2016).

38. 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).

39. 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** (2021), doi:10.7554/eLife.61718.

40. 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** (2021), doi:10.1210/endocr/bqab012.

41. 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).

42. 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** (2021), doi:10.1073/pnas.1917623117.

43. 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** (2021), doi:10.1073/pnas.2020125118.

44. 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).

45. 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).

46. M. Matamales, 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).

47. 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).

48. 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).

49. 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).

50. 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**, 456–460.e22 (2017).

51. 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).

52. C. Mayer, X. H. Jaglin, L. V. Cobbs, R. C. Bandler, C. Streicher, C. L. Cepko, S. Hippenmeyer, G. Fishell, Clonally Related Forebrain Interneurons Disperse Broadly across Both Functional Areas and Structural Boundaries. *Neuron*. **87**, 989–998 (2015).

53. 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).

54. 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).

55. 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. *bioRxiv* (2022), , doi:10.1101/2022.08.31.22279406.

56. 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).

57. 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).

58. D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).

59. R. N. Delgado, D. E. Allen, M. G. Keefe, W. R. Mancia Leon, R. S. Ziffra, E. E. Crouch, A. Alvarez-Buylla, T. J. Nowakowski, Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature*. **601**, 397–403 (2022).

60. J. M. Tepper, F. Tecuapetla, T. Koós, O. Ibáñez-Sandoval, Heterogeneity and diversity of striatal GABAergic interneurons. *Front. Neuroanat.* **4**, 150 (2010).

61. 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. Kisvárday, Y. Kubota, D. A. Lewis, O. Marín, H. Markram, C. J. McBain, H. S. Meyer, H. Monyer, 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).

62. R. Tremblay, S. Lee, B. Rudy, GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*. **91**, 260–292 (2016).

63. 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. Franken, 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. Zaghloul, 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).

64. 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).

65. G. Paxinos, C. Watson, M. Petrides, M. Rosa, H. Tokuno, *The marmoset brain in stereotaxic coordinates* (2012).

66. 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).

67. 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. *bioRxiv* (2019), , doi:10.1101/791699.

68. 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).

69. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

70. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*. **8**, 281–291.e9 (2019).

71. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888–1902.e21 (2019).

72. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018), (available at http://arxiv.org/abs/1802.03426).

73. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

74. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

75. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).

76. L. Feng, T. Zhao, J. Kim, neuTube 1.0: A New Design for Efficient Neuron Reconstruction Software Based on the SWC Format. *eNeuro*. **2** (2015), doi:10.1523/ENEURO.0049-14.2014.

Acknowledgements:

We thank Tim Blosser for his early involvement in developing the spatial transcriptomics workflows. We thank Martin Wienisch, Cindy Chen, Andrew Harrahill and Eric Nyase for manuscript comments and assistance with AAV experiments. We thank the MIT veterinarian staff for animal husbandry and for their assistance with surgical procedures. We thank Monika Burns and Yuanyuan Hou for assistance with AAV IV injections and animal perfusions.

Funding:

National Institutes of Health grant U01MH114819 (GF, SAM, EB)

NSF GRFP # 1745302 (MES)

MathWorks Science Fellowship (MES)

Collamore-Rogers Fellowship at MIT (MES)

NSF GRFP # 1122374 (TWS)

Broad Institute's Stanley Center for Psychiatric Research (SAM, GF)

Dean's Innovation Award (Harvard Medical School) (SAM)

Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT (GF)

Poitras Center for Psychiatric Disorders Research at MIT (GF)

McGovern Institute for Brain Research at MIT (GF)

Author Contributions:
RNA/ATAC Data Generation: FMK, MG, AL
Spatial Data Generation: KML, HZ
Data Analysis: FMK, KML, HZ, RCHdR, MES, MG, AL, KXL, VFB-G, TWS, SAM
Data Interpretation: FMK, KML, HZ, RCHdR, MES, MG, KXL, SB, EB, SAM, GF
Tissue Samples/Tissue Processing: FMK, KML, HZ, MG, AL, QZ, GC, SB
AAV generation & experiments: QZ, JS, SJV, JD
Morphology Data generation/analysis: KML, HZ, VFB-G, TWS, AM, EB
Software/Data management: FMK, RCHdR, MG, AW, JN, SK
Writing: FMK, KML, HZ, MES, VFB-G, SAM, GF

Competing interests: Authors declare that they have no competing interests

Data and materials availability:
Raw sequence data were produced as part of the BRAIN Initiative Cell Census Network are available for download from the Neuroscience Multi-omics Archive (https://data.nemoarchive.org/biccn/grant/u01_feng/mccarroll/transcriptome/sncell/) and the Brain Cell Data Center (https://biccn.org/data). Morphological reconstructions and single molecule FISH of interneuron types are available for download through the Brain Image Library (https://submit.brainimagelibrary.org/search?grant_number=1-U01-MH114819-01).

Supplementary Materials:
Figs. S1-S13
Table S1-S3
Data S1-S3

Figure 1

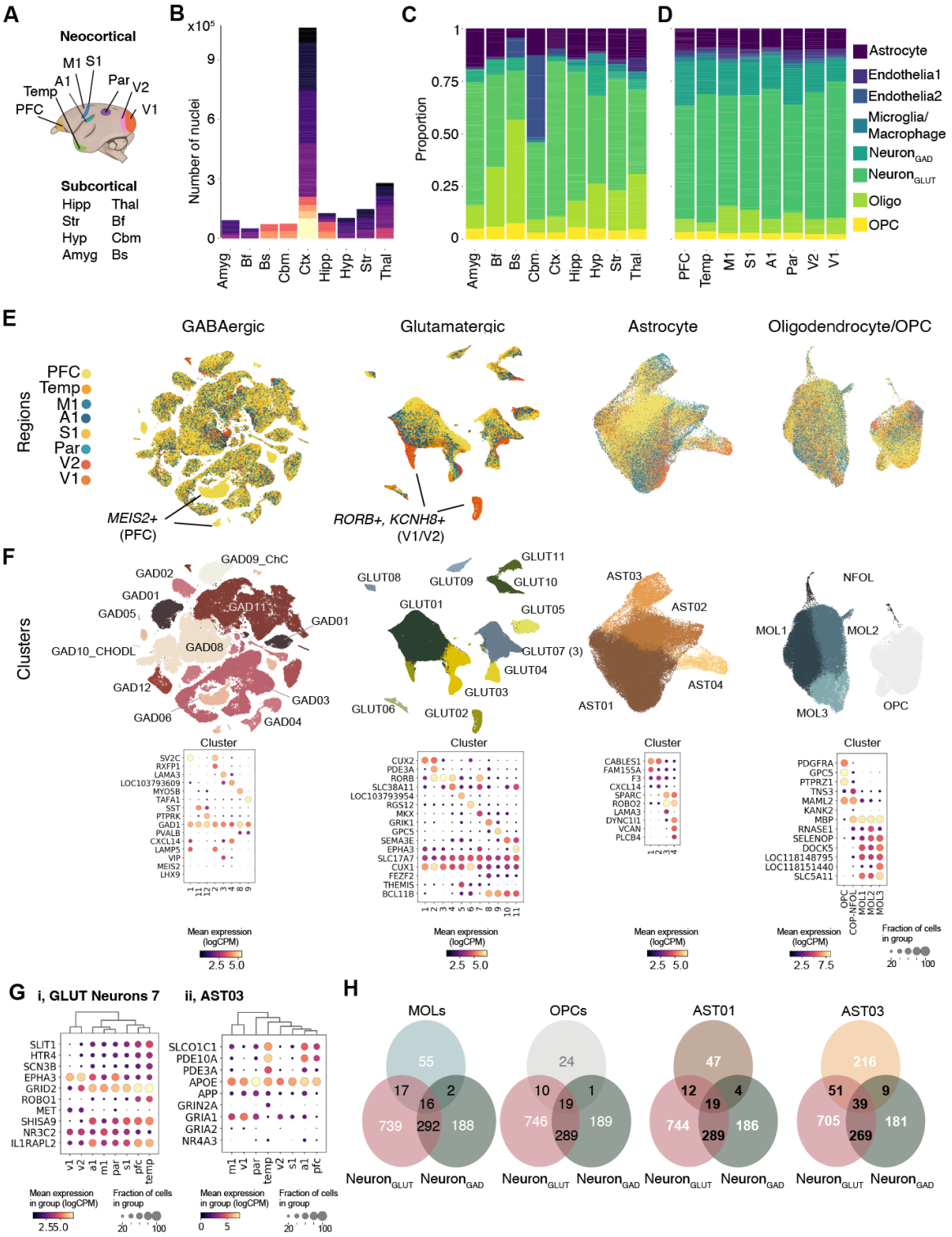
A
Neocortical
M1 S1
A1
Par
V2
Temp
V1
PFC
Subcortical
Hipp Thal
Str Bf
Hyp Cbm
Amyg Bs
B
x10^5
Number of nuclei
Amyg Bf Bs Cbm Ctx Hipp Hyp Str Thal
C
Proportion
Amyg Bf Bs Cbm Ctx Hipp Hyp Str Thal
D
PFC Temp M1 S1 A1 Par V2 V1
Astrocyte
Endothelia1
Endothelia2
Microglia/ Macrophage
Neuron GAD
Neuron GLUT
Oligo
OPC
E
GABAergic
Glutamatergic
Astrocyte
Oligodendrocyte/OPC
Regions
PFC
Temp
M1
A1
S1
Par
V2
V1
MEIS2+ (PFC)
RORB+, KCNH8+ (V1/V2)
F
Clusters
GAD09_ChC
GAD02
GAD01
GAD05
GAD11
GAD01
GAD10_CHODL
GAD08
GAD12
GAD03
GAD06
GAD04
GLUT08
GLUT09
GLUT11
GLUT10
GLUT01
GLUT05
GLUT07 (3)
GLUT04
GLUT03
GLUT06
GLUT02
AST03
AST02
AST04
AST01
NFOL
MOL1
MOL2
MOL3
OPC
Cluster
Mean expression (logCPM)
Fraction of cells in group
G
i, GLUT Neurons 7
ii, AST03
SLIT1
HTR4
SCN3B
EPHA3
GRID2
ROBO1
MET
SHISA9
NR3C2
IL1RAPL2
SLCO1C1
PDE10A
PDE3A
APOE
APP
GRIN2A
GRIA1
GRIA2
NR4A3
Mean expression in group (logCPM)
Fraction of cells in group
H
MOLs
OPCs
AST01
AST03
Neuron GLUT
Neuron GAD

Figure 1. **Single nucleus RNA sequencing of marmoset brain. (A)** Brain regions in snRNA-seq dataset. **(B)** Number of nuclei per region. Colors indicate biological replicates. **(C)** Proportions of major cell classes across major brain structures sampled. **(D)** Proportions of major cell classes per cortical region sampled. **(E)** UMAP plots of cortical GABAergic neurons, glutamatergic neurons, astrocytes and oligodendrocyte lineage cells, with nuclei colored by cortical location. Arrows identify two clusters that are regionally-specific: a GABAergic *MEIS2*+ type in medial PFC, and a glutamatergic *KCNH8*+, *RORB*+ type in V1 and V2. **(F)** Same as ***(E)*** but colored by main cluster. Dot plots: expression of type markers and differentially expressed genes in log base 2 counts per million (logCPM, color) and fraction of cells with nonzero expression in the region (dot size) **(G)** Expression of selected regionally differentially expressed genes (rDEGs) for two cell types: GLUT07 (i) and AST03 (ii), in logCPM and fraction of cells with nonzero expression in the region (dot size). Dendrograms represent hierarchical clustering of cortical regions based on top 1000 most variable genes. **(H)** Venn diagrams showing number of genes that are shared rDEGs between neuronal types and two astrocyte and oligodendrocyte types. Hipp = hippocampus, Str = striatum, Hyp = hypothalamus, Amy = amygdala, Th = thalamus, Bf = basal forebrain, Cb = cerebellum, Bs = brainstem.

Figure 2

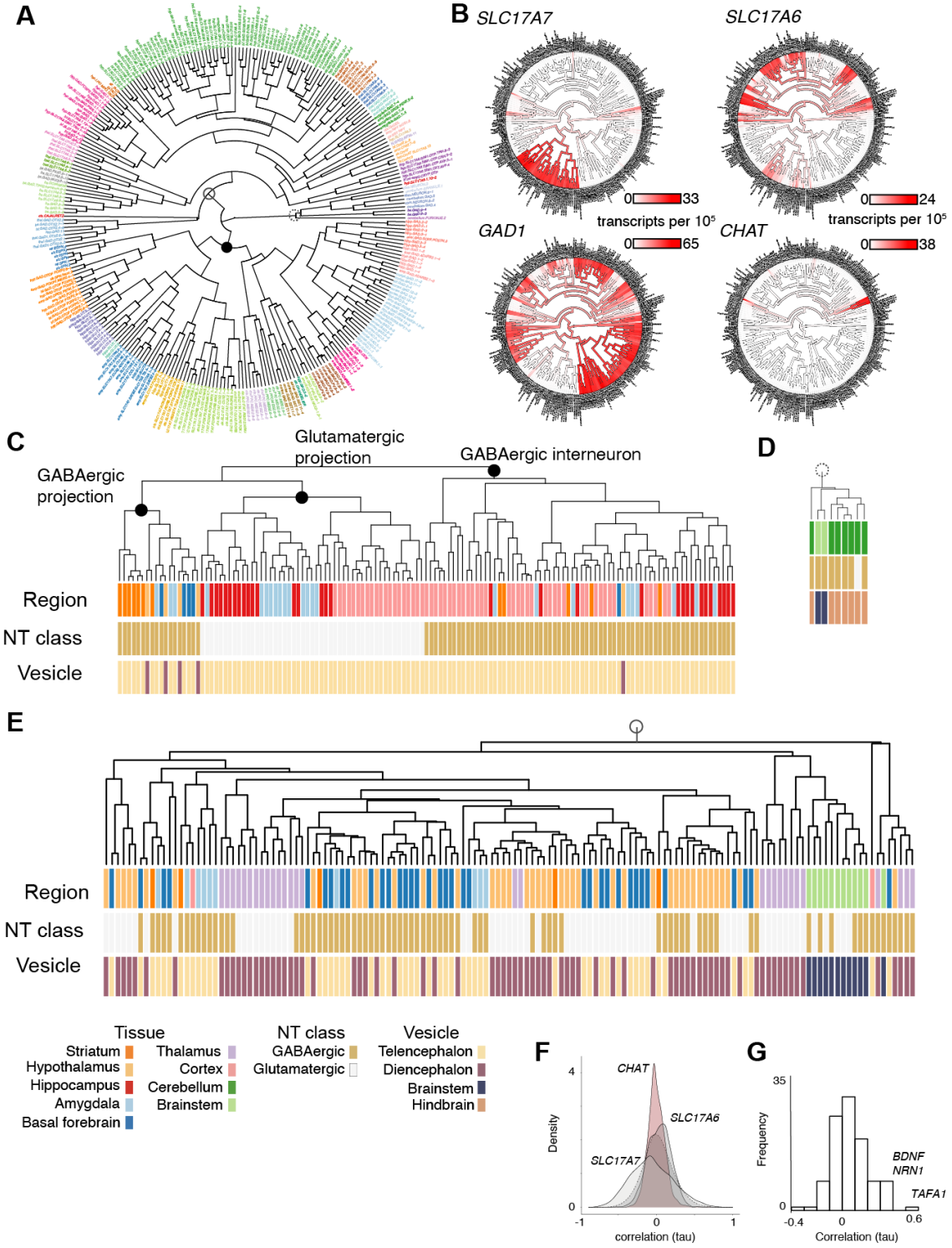
A
B
SLC17A7
SLC17A6
0
33
0
24
transcripts per 10^5
transcripts per 10^5
GAD1
CHAT
0
65
0
38
C
Glutamatergic projection
GABAergic interneuron
GABAergic projection
D
Region
NT class
Vesicle
E
Region
NT class
Vesicle
Tissue
Striatum
Hypothalamus
Hippocampus
Amygdala
Basal forebrain
Thalamus
Cortex
Cerebellum
Brainstem
NT class
GABAergic
Glutamatergic
Vesicle
Telencephalon
Diencephalon
Brainstem
Hindbrain
F
CHAT
SLC17A6
SLC17A7
Density
4
0
-1
0
1
correlation (tau)
G
35
Frequency
BDNF
NRN1
TAFA1
0
-0.4
0
0.6
Correlation (tau)

Figure 2. Neuronal taxonomy and neurotransmitter utilization. (A) Hierarchical clustering of 288 neuronal clusters across 19 brain regions. Clusters are first resolved within each brain region dataset, then arranged in the dendrogram using a correlation-based distance metric. Leaf colors distinguish 40 terminal clades for better visibility of clade boundaries. **(B)** Expression of four neurotransmitters (*SLC17A7, SLC17A6, GAD1, CHAT*) plotted on the dendrogram using normalized values for each leaf node (cluster). Red gradient represents normalized, scaled expression (number of transcripts per 100,000) (C-E) Linearized partial dendrogram for the three main clades in *(A)*. Dot arrays represent metadata associated with each cluster: brain region (Region), primary neurotransmitter usage (NT), and cephalic partition (Vesicle). **(C)** The parent node of telencephalic neurons and of GABAergic projection neurons. **(D)** Same as in *(C)* but for a midbrain and hindbrain clade. **(E)** Same as *(C)* but for a mixed telencephalic/diencephalic clade. **(F)** Gene-gene correlations (Spearman's tau) between all 5,904 meaningfully expressed genes (dotted line), or between all genes and *SLC17A7* or *CHAT*. **(G)** Histogram of correlations between *SLC17A7* and the top 116 genes correlated (tau > 0.4) with *SLC17A6*.

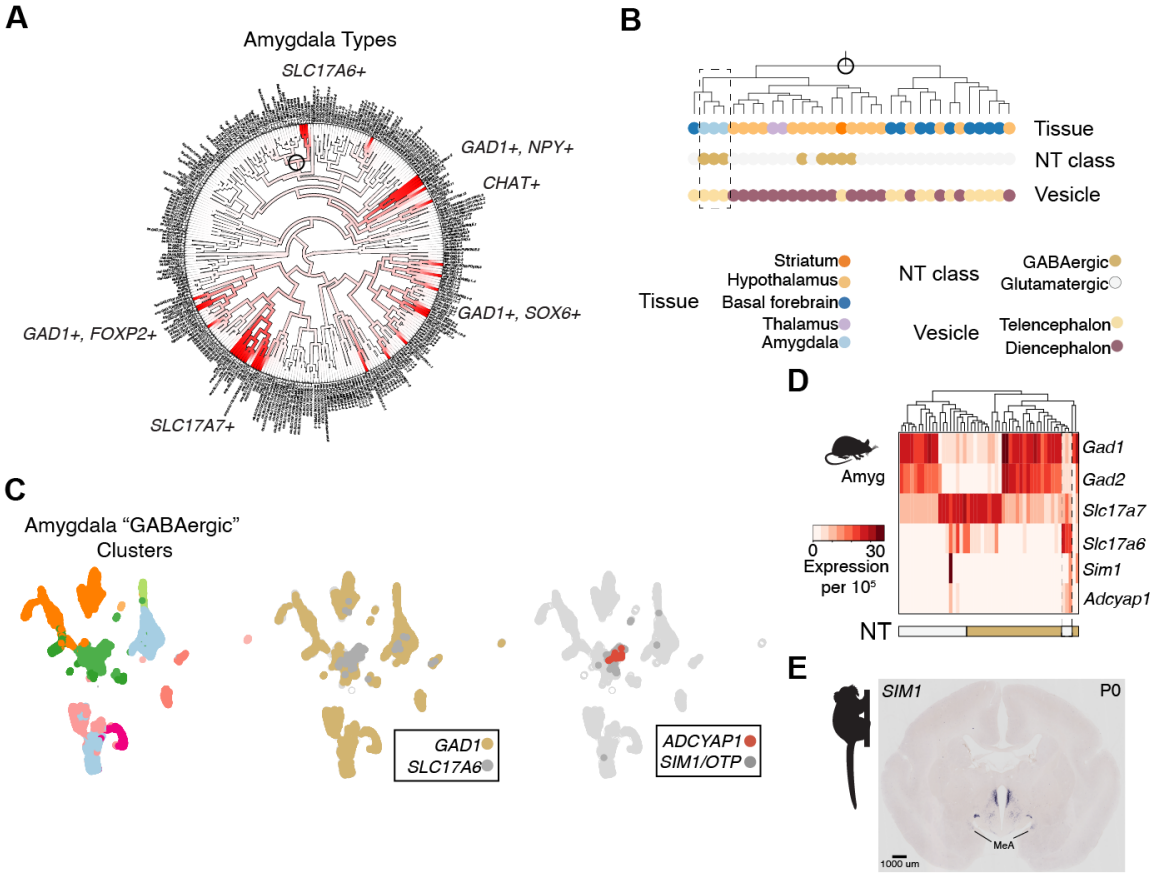
A
Amygdala Types
SLC17A6+
GAD1+, NPY+
CHAT+
GAD1+, SOX6+
GAD1+, FOXP2+
SLC17A7+
B
Tissue
NT class
Vesicle
Tissue
Striatum
Hypothalamus
Basal forebrain
Thalamus
Amygdala
NT class
GABAergic
Glutamatergic
Vesicle
Telencephalon
Diencephalon
C
Amygdala "GABAergic" Clusters
GAD1
SLC17A6
ADCYAP1
SIM1/OTP
D
Amyg
Gad1
Gad2
Slc17a7
Slc17a6
Sim1
Adcyap1
0
30
Expression per 10^5
NT
E
SIM1
P0
MeA
1000 um

Figure 3. A glutamatergic diencephalic type in amygdala with GABAergic-like properties. (A) Locations of amygdala clusters in the dendrogram from **Fig. 2A**. **(B)** Position of three amygdala GABAergic-like *SLC17A6*+ subtypes within a clade dominated by hypothalamic and basal forebrain types. **(C)** UMAP of clustering analysis that includes only marmoset amygdala neurons (n = 44,165 neurons) shows that the *SLC17A6*+ subtypes cluster with amygdala GABAergic types despite themselves lacking *GAD1* or *GAD2* expression. They selectively express *ADCYAP1* (PACAP) and hypothalamic-associated transcription factors *SIM1* and *OTP.* **(D)** Normalized, scaled expression of neurotransmitters in neurons from snRNA-seq of adult mouse amygdala (n = 25,930 nuclei, 3 replicates pooled). Dashed lines indicate the GABAergic-like *Slc17a6*+ populations. **(E)** *SIM1* ISH in P0 marmoset (https://gene-atlas.brainminds.jp) shows labeled cells in medial amygdala (MeA) (arrows).

Figure 4

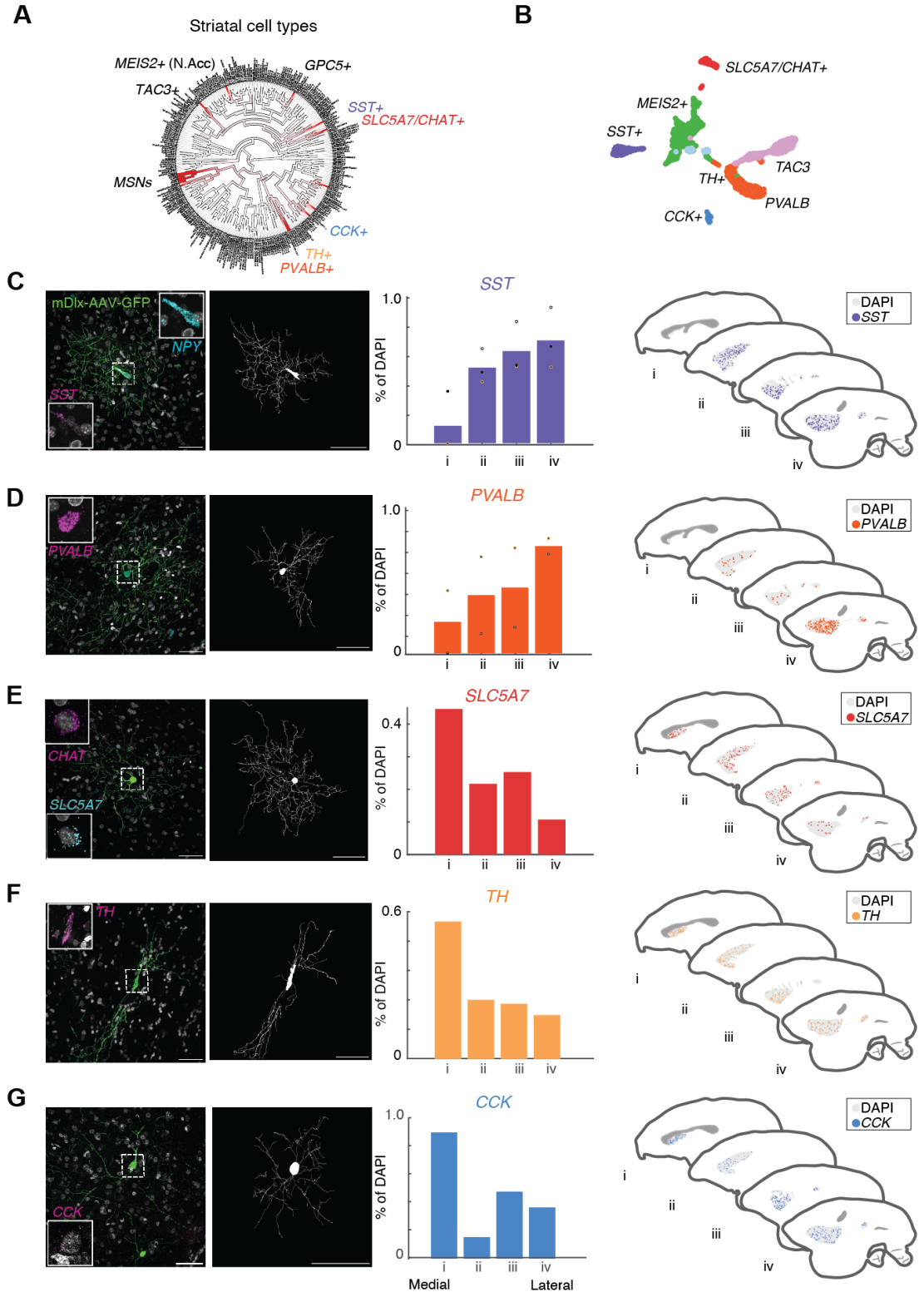

Figure 4. Quantitative single-molecule FISH and morphology of conserved striatal interneurons. **(A)** Locations of striatal clusters in the dendrogram from **Fig. 2A**. **(B)** UMAP of striatal GABAergic

interneurons. **(C-G)** Densities and exemplar morphological reconstructions of the major conserved striatal interneuron populations. *(left*) Striatal interneurons labeled with mDlx5/6-AAV-GFP. Insets show smFISH of markers confirming molecular identity. Second panel shows morphological reconstruction of GFP+ cell (**Table S2**). (*middle*) Barplots quantify positive cells as proportion of all (DAPI+) cells from medial to lateral sections of smFISH performed on thin (16 µm) sagittal sections of marmoset striatum. Dots = individual replicates. (*right*) Scatter plots on brain anatomy cartoons show locations of positive cells summarized in barplot (**Data S2**).

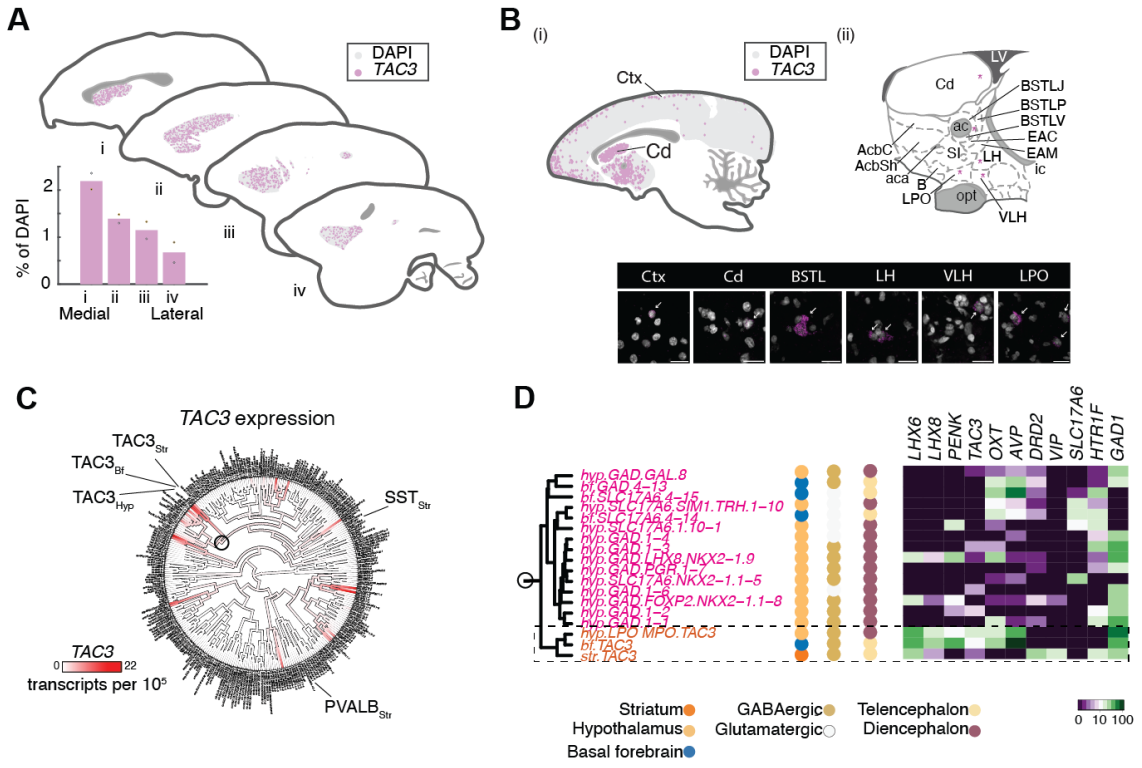
A
DAPI
TAC3
i
ii
iii
iv
% of DAPI
2
1
0
Medial
Lateral
B
(i)
(ii)
Ctx
Cd
LV
BSTLJ
BSTLP
BSTLV
EAC
EAM
ic
ac
SI
LH
AcbC
AcbSh
aca
B
LPO
opt
VLH
Ctx
Cd
BSTL
LH
VLH
LPO
C
TAC3 expression
TAC3Str
TAC3Bf
TAC3Hyp
SSTStr
PVALBStr
TAC3
0
22
transcripts per 10^5
D
LHX6
LHX8
PENK
TAC3
OXT
AVP
DRD2
VIP
SLC17A6
HTR1F
GAD1
hyp.GAD.GAL.8
bf.GAD.4–13
bf.SLC17A6.4–15
hyp.SLC17A6.SIM1.TRH.1–10
bf.SLC17A6.4–14
hyp.SLC17A6.1.10–1
hyp.GAD.1–4
hyp.GAD.1–3
hyp.GAD.LHX8.NKX2–1.9
hyp.GAD.PGR.1–2
hyp.SLC17A6.NKX2–1.1–5
hyp.GAD.1–6
hyp.GAD.FOXP2.NKX2–1.1–8
hyp.GAD.1–2
hyp.GAD.1–5
hyp.LPO.MPO.TAC3
bf.TAC3
str.TAC3
Striatum
Hypothalamus
Basal forebrain
GABAergic
Glutamatergic
Telencephalon
Diencephalon
0 10 100

Figure 5. Biodistribution of primate-specific *TAC3+* interneurons. (A) Location of *TAC3+* interneurons in marmoset striatum. Barplot (inset) quantifies positive cells as proportion of all (DAPI+) cells from medial to lateral sections. Dots = individual replicates. **(B)** smFISH reveals anatomical locations and expression levels *TAC3+* types in different brain regions. *(i)* Schematic of TAC3+ cells imaged across cortex, dorsal striatum. Ctx = neocortex, Cd = Caudate. *(ii)* Cartoon close-up of nuclei in striatum, basal forebrain and hypothalamus. Magenta stars = locations of *TAC3+* cells in lower image panel. Cd = Caudate, AcbC = nucleus accumbens core, AcbSh = nucleus accumbens shell, SI = Substantia innominata, B = basal nucleus of Meynert, EAM = extended amygdala, medial, EAC = extended amygdala, central, ac = anterior commissure, BSTLP = bed nuc st, lateral posterior, BSTLJ = bed nuc st, juxtacap, BSTLV = bed nuc st, lateral ventral, LH = lateral hypothalamus, VLH = ventrolateral hypothalamus, LPO = lateral preoptic area. **(C)** *TAC3* is expressed in over 20 neuron types (mostly GABAergic). Red gradient represents normalized, scaled expression (number of transcripts per 100,000). **(D)** Hierarchical clustering of the clade which contains the *TAC3+* striatal type (black circle in (*C)* indicates stem location of clade on full dendrogram). The *TAC3+* striatal type is transcriptionally most similar to other *TAC3+* neurons in basal forebrain and hypothalamus (indicated by dashed outline); the neighborhood largely contains hypothalamic types (GABAergic and glutamatergic). Heatmap shows differentially expressed genes. Colors represent normalized, scaled expression (transcripts per 100,000).

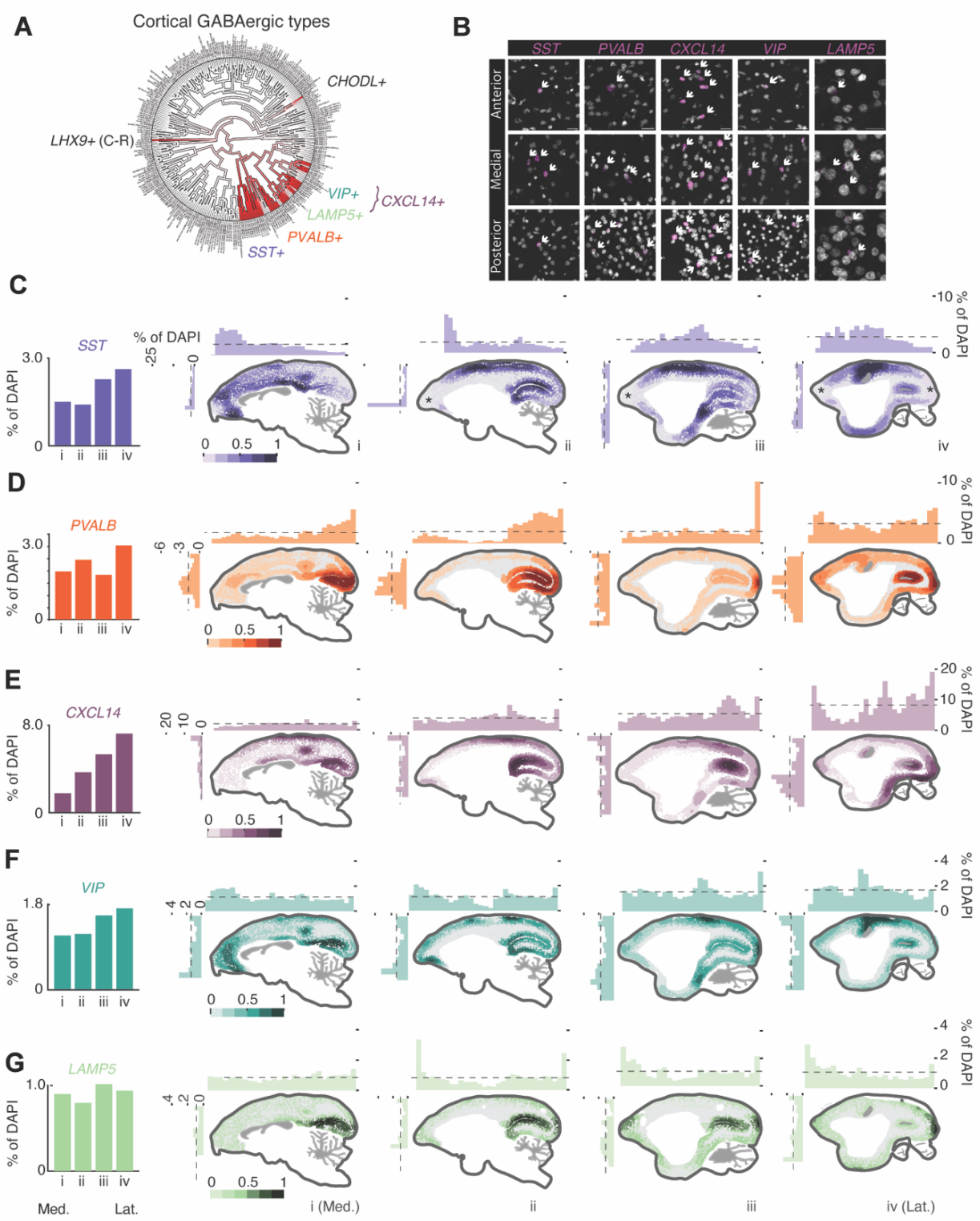
A
Cortical GABAergic types
CHODL+
LHX9+ (C-R)
VIP+
LAMP5+
CXCL14+
PVALB+
SST+
B
SST
PVALB
CXCL14
VIP
LAMP5
Anterior
Medial
Posterior
C
SST
% of DAPI
D
PVALB
E
CXCL14
F
VIP
G
LAMP5
i
ii
iii
iv
Med.
Lat.
i (Med.)
iv (Lat.)

Figure 6. Quantitative single-molecule FISH of cortical interneurons. (A) Locations of major neocortical interneuron types in the dendrogram from **Fig. 2A**. **(B)** Examples of AAV-mDlx-GFP and smFISH labeled cells along with reconstructed morphology. **(C-G)** Example densities of the major neocortical interneuron populations. Barplots quantify positive cells as proportion of all (DAPI+) cells from medial to lateral sections (**Data S2**). Dorsal-ventral and anterior-posterior density plots are arranged around each section. Color maps on brain sections show scaled density of labeled cells, normalized across all sections for each gene separately. Asterisks denote regions of tissue artifact (identified by lower numbers of DAPI+ nuclei relative to similar sections) which may have inaccurate counts. Ant. = Anterior, Post. = Posterior, Dors. = Dorsal, Vent. = Ventral, Med. = Medial, Lat. = Lateral.