1   **A marmoset brain cell census reveals influence of developmental origin and functional class on**
2   **neuronal identity**

3   Fenna M. Krienen[1,2]*⌂, Kirsten M. Levandowski[2,3,4], Heather Zaniewski[2,3,4], Ricardo C.H. del Rosario[2],

4   Margaret E. Schroeder[2,3,4], Melissa Goldman[2], Martin Wienisch[2,3,4], Alyssa Lutservitz[1], Victoria F. Beja-

5   Glasser[2,3,4], Cindy Chen[2,3,4], Qiangge Zhang[2,3,4], Ken Y. Chan[2], Katelyn X. Li[3,4], Jitendra Sharma[3,4], Dana

6   McCormack[2,3,4], Tay Won Shin[3,4,5], Andrew Harrahill[3,4], Eric Nyase[3,4], Gagandeep Mudhar[6], Abigail

7   Mauermann[3,4,5], Alec Wysoker[2], James Nemesh[2], Seva Kashin[2], Josselyn Vergara[2], Gabriele Chelini[7],

8   Jordane Dimidschstein[2], Sabina Berretta[8,9], Benjamin E. Deverman[2], Ed Boyden[3,4,5], Steven A.

9   McCarroll[1,2#]*, Guoping Feng[2,3#]*

10

11

12   [1] Department of Genetics, Harvard Medical School; Boston, Massachusetts 02115, USA.

13   [2] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard; Cambridge, Massachusetts

14   02142, USA.

15   [3] McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts

16   Institute of Technology; Cambridge, Massachusetts 02142, USA.

17   [4] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; Cambridge,

18   Massachusetts 02139, USA.

19   [5] HHMI, Cambridge, MA, USA.

20   [6] Princeton Neuroscience Institute, Princeton University. Princeton, NJ 08544, USA.

21   [7] Center for Mind/Brain Sciences, University of Trento. Piazza della Manifattura n.1, Rovereto (TN).

22   38068, IT.

23   [8] Basic Neuroscience Division, McLean Hospital, Belmont MA USA.

24   [9] Dept. of Psychiatry, Harvard Medical School, Boston, MA USA.

25

26   #These authors contributed equally to this work

27   *Corresponding authors. Email: fkrienen@princeton.edu, mccarroll@genetics.med.harvard.edu,

28   fengg@mit.edu

29   ⌂Present affiliation: Princeton Neuroscience Institute. Princeton NJ USA

30   **Abstract:**

31   The mammalian brain is composed of many brain structures, each with its own ontogenetic and

32   developmental history. Transcriptionally-based cell type taxonomies reveal cell type composition and

33   similarity relationships within and across brain structures. We sampled over 2.4 million brain cells across

34   18 locations in the common marmoset, a New World monkey primed for genetic engineering, and used

35   single-nucleus RNA sequencing to examine global gene expression patterns of cell types within and

36   across brain structures. Our results indicate that there is generally a high degree of transcriptional

37   similarity between GABAergic and glutamatergic neurons found in the same brain structure, and there are

38   generally few shared molecular features between neurons that utilize the same neurotransmitter but reside

39   in different brain structures. We also show that in many cases the transcriptional identities of cells are

40   intrinsically retained from their birthplaces, even when they migrate beyond their cephalic compartments.

41   Thus, the adult transcriptomic identity of most neuronal types appears to be shaped much more by their

42   developmental identity than by their primary neurotransmitter signaling repertoire. Using quantitative

43   mapping of single molecule FISH (smFISH) for markers for GABAergic interneurons, we found that the
44   similar types (e.g. *PVALB+* interneurons) have distinct biodistributions in the striatum and neocortex.
45   Interneuron types follow medio-lateral gradients in striatum but form complex distributions across the
46   neocortex that are not described by simple gradients. Lateral prefrontal areas in marmoset are
47   distinguished by high relative proportions of *VIP+* neurons. We further used cell-type-specific enhancer
48   driven AAV-GFP to visualize the morphology of molecularly-resolved interneuron classes in neocortex
49   and striatum, including the previously discovered novel primate-specific *TAC3+* striatal interneurons. Our
50   comprehensive analyses highlight how lineage and functional class contribute to the transcriptional
51   identity and biodistribution of primate brain cell types.

52

53   **One-Sentence Summary:** Adult primate neurons are imprinted by their region of origin, more so than by
54   their functional identity.

55

56   **Main Text:**
57   The mammalian brain's complex functional diversity stems from its vast cellular and molecular
58   repertoire. To provide a more complete understanding of the cell types across major cortical and
59   subcortical structures in a non-human primate brain, we conducted a census of cell types of the adult
60   marmoset brain based on their transcriptional profiles. Previous single cell sequencing studies of the
61   marmoset brain focused on single brain regions (*1*, *2*), or on specific cell classes across regions (*3*, *4*).
62   However, inclusion of both closely and distantly related brain structures and cell types can yield insights
63   into the developmental and ontological relationships between them (*5*). Comprehensive transcriptomic
64   cell type atlases have been produced for the mouse (*6–9*). Complementing these, recent transcriptomic
65   datasets that sample many regions across humans and nonhuman primates brains (*10–12*) offer powerful
66   resources for comparative analysis of brain cell type features. We generated single-nucleus RNA
67   sequencing (snRNA-seq; 10x Genomics 3' v3.1) data from 2.4 million unsorted brain cell nuclei across 8
68   neocortical and 10 subcortical locations from 10 young adult marmosets (4 M, 6 F), and resolved clusters
69   from all major neuronal and non-neuronal cell classes. snRNA-seq data were generated as part of the
70   Brain Initiative Cell Census Network (BICCN, RRID:SCR_015820) and are available on the BICCN
71   Data Center (RRID:SCR_022815; https://biccn.org/data) as well as the NeMO archive
72   (RRID:SCR_016152; https://doi.org/10.1101/2022.10.18.512442).

73   All neuron-containing brain structures in the central nervous system possess both excitatory and
74   inhibitory neuronal populations, though the proportions and degree of developmental relatedness between
75   these two populations varies by structure. In the neocortex and other telencephalic structures of
76   vertebrates, distinct populations of neurons are typically categorized based on their neurotransmitter
77   status as either inhibitory (GABAergic) or excitatory (glutamatergic). In other brain structures, primary
78   neurotransmission appears less essential to a cell type's identity. We found that the transcriptional
79   identities of excitatory and inhibitory neurons within telencephalic brain structures segregate strongly,
80   consistent with previous studies in other mammalian species (*2*, *6*, *7*, *13*). In contrast, there is much
81   greater transcriptional similarity between GABAergic and glutamatergic neuronal types in non-
82   telencephalic compartments. Moreover, few gene expression distinctions present in telencephalic
83   glutamatergic neurons are shared with glutamatergic neurons in non-telencephalic brain regions. While
84   primary neurotransmission did not drive transcriptional similarity between neurons, their brain structure

85 of origin did; the adult transcriptomic identity of most neuronal types appear to be shaped much more by
86 their developmental identity than by their neurotransmitter signaling repertoire.

87 In the mouse, overall interneuron/excitatory neuron ratios are thought to be largely invariant across the
88 neocortex (*14*). Even so, relative proportions of interneuron subtypes do vary and underlie principles of
89 functional organization in the mouse neocortex (*15*). The small size of the marmoset brain and its near-
90 lissencephalic neocortex enables quantitative, cell-type-resolved mapping in a primate. We used smFISH
91 to spatially profile major interneuron types across marmoset striatum and neocortex. In the striatum,
92 major interneuron types are distributed as medial-lateral gradients. In contrast, the marmoset neocortex
93 has a much more complex topography of interneuron concentrations that is not explained by a single
94 spatial axis. Lateral prefrontal areas in particular, which have undergone expansion in primate evolution
95 (*16*) are typified by higher proportions of *VIP+* and *LAMP5/ID2+* interneurons and lower proportions of
96 *SST+* interneurons relative to all other neocortical areas.

97 The development of viral tools to effectively deliver transgenes to specific primate brain cell types has
98 many powerful applications (*17–20*). We used this approach to sparsely label and reconstruct the
99 morphologies of selected types. We reconstructed a set of molecularly-characterized neurons that had
100 received systemic delivery of an AAV carrying a reporter (AAV9-hDlx5/6-GFP-fGFP) under an
101 interneuron-selective regulatory element(*19*). Reconstructions are available for download from the Brain
102 Image Library (BIL; https://doi.org/10.35077/g.609). As AAV9-hDlx5/6-GFP-fGFP did not label a
103 recently discovered, primate-specific striatal interneuron type (*3*) w(*3*)This enabled visualization(*3*).
104 Together, our census of major transcriptomically-defined brain cell types and quantitative mapping of
105 interneuron biodistributions provides a key resource for the primate neuroscience community and for
106 comparative studies of cell type evolution.

107 *A transcriptomic census of marmoset brain cell types*
108 We acquired snRNA-seq data from 18 brain regions collected using 10x Genomics 3' 3.1 chemistry
109 across 10 young adult marmosets (ages 1.5-4; 6 F) as well as a small dataset from PFC of two aged
110 animals (**Fig. 1, Fig. S1, Table S1**) . The number of donors per brain structure varied (min = 2, max = 10;
111 **Fig. S1**), as did the cell sampling rate per brain structure (**Fig. 1A; Fig. S1A, Table S2**). Neocortex was
112 the most comprehensively sampled in terms of total numbers, donors, and regional dissections. We
113 acquired data from: cerebellum, brainstem, hypothalamus, thalamus, amygdala, striatum (separate
114 dissections for caudate, putamen and nucleus accumbens), hippocampus, basal forebrain and neocortex.
115 Within the neocortex, we separately sampled 8 neocortical locations (prefrontal, temporal pole, S1, M1,
116 A1, V2, V1, lateral parietal association), and within PFC, from four prefrontal subdivisions (**Fig. S1B-C**).

117 Nuclei from each brain structure were pooled across donors and analyzed to identify major cell types and
118 their proportions by brain structure (**Fig. 1B-D**). Using linear discriminant analysis (scPred;(*21*)) trained
119 on a supervised set of cell class labels, we identified and discarded low quality cells and doublets, and
120 assigned each nucleus to its probable major type – astrocyte, endothelia, ependyma,
121 microglia/macrophage, neuron, oligodendrocyte, or oligodendrocyte precursor cell (OPC).

122 After major cell type assignment, nuclei were clustered by brain structure to reveal subtype diversity
123 within each major class. We used a previously described clustering pipeline (*6*) based on independent
124 components analysis (fastICA). Each clustering analysis involved additional curation of doublets and
125 outlier cells, followed by a second round of sub-clustering of major clusters. At each of these curation

126   stages, independent components that loaded on individual donors or batches were excluded and
127   reclustering was performed to attenuate donor and batch effects on clustering results. For each cluster, a
128   "metacell" was generated by summing transcript counts of individual cells of each type together, followed
129   by scaling and normalizing to 100,000 transcripts/nucleus. These regional and cell-type resolved
130   metacells were the starting point for most cross-region analyses. The final snRNA-seq dataset size after
131   curation contained 2.01 million cells (**Table S2)**.

132   *Hierarchical clustering of neurons*
133   Cell types are identified based on factors including their function, developmental origin, lineage, and
134   regional context (*22–24*). We studied how transcriptional profiles related to each other across 288
135   neuronal cell types (metacells) sampled across all 18 marmoset brain structures. Individual replicates and
136   other variables (age, sex) were generally proportionately represented across resolved cell types with
137   several exceptions that were likely mainly to differences in dissection across donors (**Fig. S1B-E, Fig S2**).
138   We used hierarchical clustering to situate the neuronal types on a dendrogram computed using distances
139   calculated from the top 100 PC scores across neuron types (**Fig. 1C**). Most telencephalic types
140   (neocortex, amygdala, hippocampus, striatum) clustered distinctly from diencephalic and hindbrain types,
141   indicating that developmental lineage continues to shape the transcriptional identity of adult primate brain
142   cell types. However, of the 7 major clades, 4 contained mixtures of cell types at the terminal (leaf level)
143   originating from distinct cephalic compartments. For example, basal forebrain neuron types intermingled
144   with hypothalamic types, suggesting closer transcriptional similarity of two structures that occupy distinct
145   cephalic compartments. The overall dendrogram configuration was broadly conserved when recomputed
146   using other distance functions (**Fig. S2A**).

147

148   Transcription factors are master regulators that determine cell type identity in development through
149   temporal patterning, suggesting that they may be a key class of genes that determine transcriptional
150   identity in adults (*8*). Supporting this view, some transcription factors are associated with specific brain
151   structures or cephalic compartments, such as *FOXG1* in the telencephalon (**Fig. 1D**) and *OTX2* in non-
152   telencephalic structures (**Fig. 1D**), show expression restricted to specific clades in the dendrogram.
153   Hierarchical clustering based only on transcription factor genes was highly similar to the original tree
154   computed over all expressed genes. However, the tree ordering generated by transcription factors alone
155   did not produce lower tree distances to the original tree than similarly sized sets of randomly selected
156   expressed genes (**Fig. S2D**). This suggests that though transcription factors undoubtedly play central roles
157   in determining cell identity in development, they do not determine global transcriptomic similarities
158   across neuron types in adulthood.

159

160   To determine whether a broadly similar transcriptomically-defined dendrogram is conserved across
161   species, we repeated the analysis using a prior single cell RNA sequencing dataset of regions sampled
162   across the mouse brain (*6*) (**Fig. S2C,D**). We found a conserved tree structure in the mouse data,
163   suggesting broad conservation of the features that drive transcriptomic similarities in neurons across adult
164   primates and rodents. The relative ordering of the major clades was strikingly similar across species: for
165   example, in both species, telencephalic glutamatergic neurons are most similar to a clade of GABAergic
166   neurons that includes striatal spiny projection neurons (SPNs) as well as transcriptionally similar types in
167   amygdala and hypothalamus (**Fig. S2C**). In both species, all telencephalic GABAergic interneurons

168   formed a large single clade, and cerebellar types were most similar to a mixed clade containing thalamic
169   and brainstem types.
170
171   Borrowing from ancestral state reconstruction methods typically used to estimate evolutionary
172   divergences between genetic sequences or species (*25*), we applied a maximum likelihood-based
173   approach (fastAnc) to the expression of all genes at each leaf (observed cell type), along with the branch
174   lengths between adjacent leaves, in order to reconstruct the transcriptomic state (inferred expression
175   values of genes) at internal nodes of the marmoset dendrogram. We then compared pairwise "ancestral"
176   expression values of all genes in the parent nodes for each of the 7 major clades depicted in **Fig. 1C, Fig.**
177   **S2E-F**). Transcription factors (TFs) were overrepresented in comparisons of internal nodes that contained
178   cell types stemming from developmentally related brain regions, but were not overrepresented when the
179   leaf cell types stemmed from multiple regions (**Fig. S2E-F**). While transcription factor expression reflects
180   the developmental origins of cell types, and while their expression alone recapitulates the tree structure
181   seen by including all genes (**Fig. S2D** and (*8*)), some cell types are transcriptionally similar despite having
182   distinct developmental origins. This may reflect convergence in adult transcriptional profiles (*5, 26*).

183   *Many adult neuronal types are imprinted by their developmental origin*
184   We next inspected RNA expression patterns along with details about each cell type's dissection region-of-
185   origin to assess which brain structures tended to contain highly similar cell types and which had more
186   dispersed transcriptional profiles. Some tissues, such as the neocortex, gave rise to cell types that
187   exclusively clustered within their cephalic domain (**Fig. 1C**), a result unlikely to be driven by ambient
188   RNA contamination since regionally variable genes were not over-enriched in ambient RNA estimates
189   (**Fig. S5A**). However, within a given cephalic domain, cell types from distinct brain structures were often
190   more similar to types sampled from other brain structures. For example, while hippocampal cell types
191   were all found in telencephalic clades in the dendrogram, many individual hippocampal types were more
192   similar to types in the amygdala or neocortex than they were to other hippocampal types (**Fig. 1C**).
193
194   Out of 62 neocortical neuron types, only two types joined clades outside of the major GABAergic and
195   glutamatergic telencephalic branches: (1) the *MEIS2*+ prefrontal GABAergic types (**Fig. S1C; Fig. 1D;**
196   **Fig. S3**), which formed a clade most similar to diencephalic and midbrain *OTX2*+ types, and (2)
197   Neocortical Cajal-Retzius (C-R) neurons, which were more similar to a clade of *LHX9*+ thalamic neurons
198   despite indications that they predominantly originate from the cortical hem in primates (*27*) (**Fig. 1D**). As
199   with hippocampal neurons, neocortical GABAergic neurons neighbored other GABAergic neurons from
200   the striatum, amygdala and hippocampus that tended to express the same marker. For example, all
201   *PVALB*+ types sampled across these telencephalic structures grouped together (**Fig. S3**). Unlike
202   neocortical GABAergic interneurons, glutamatergic neocortical neurons were broadly most similar to
203   other telencephalic glutamatergic types, but showed almost no mixing at the terminal leaf level (**Fig. 1C**).
204   Cerebellar neuron types were entirely restricted to a single, unmixed clade in both mouse and marmoset
205   (**Fig. S2C**), though this relative isolation could be due to the lack of other hindbrain structures in our
206   datasets (**Fig. 1C**).
207
208   Consistent with previous work suggesting that mammalian thalamus contains both midbrain-derived and
209   forebrain-derived GABAergic interneurons (*28*), we observed distinct clades of thalamic neurons that
210   were most similar to diencephalic or midbrain populations. Interestingly, thalamic GABAergic neurons

211 that were *OTX2+* were distinct from other thalamic populations (**Fig. 1D, Fig. S3**), and formed a clade
212 with other *OTX2+* neurons sampled from brainstem, hypothalamus, and basal forebrain. These
213 populations were *SOX14+*, while thalamic *SOX14-* populations joined mixed diencephalic-telencephalic
214 clades (**Fig. 1D; Fig. S3**). The dispersion of thalamic neurons to distinct diencephalic and midbrain-
215 dominated clades supports recent work suggesting multiple developmental origins for primate thalamic
216 GABAergic neurons (*28*). We did not find thalamic GABAergic populations that were transcriptionally
217 similar to telencephalic types (*29*), though we note that the next most proximal clade to the thalamic
218 *OTX2+* types were *MEIS2+* GABAergic neocortical neurons (**Fig. 1D**).
219
220 Amygdala neuron types were distributed to 5 of the 7 major clades (**Fig. 1C**). Consistent with this widely
221 dispersed cellular profile, the amygdala is composed of loosely associated nuclei with diverse
222 phylogenetic and developmental origins, as documented by previous studies (*30, 31*). Specifically, the
223 basolateral amygdala, which has a high proportion of excitatory neurons, shares properties with cortical
224 and claustrum neurons, whereas the intercalated nuclei of the amygdala contain inhibitory *FOXP2+*
225 projection neurons that share developmental origins with some striatal GABAergic populations such as
226 striatal spiny projection neurons, SPNs (*5, 32, 33*). In our dataset, glutamatergic amygdalar neurons were
227 most similar to neocortical and hippocampal glutamatergic neurons, while *FOXP2+* GABAergic
228 amygdalar neurons clustered with SPN types (**Fig. S3**), in line with recent lineage tracing studies in mice
229 (*5*) and with analysis of fetal macaque types (*33*). Amygdala GABAergic interneuron types had highly
230 divergent transcriptomic profiles; while some were most similar to telencephalic GABAergic neurons,
231 others clustered with hypothalamic GABAergic neurons. These results underscore the diversity and
232 developmental complexity of cell types comprising the amygdala nuclei.

233 *Neurotransmitter usage is not strongly associated with transcriptomic identity*
234 The neurotransmitters used by neurons are central to their function, and neurons are often classified by
235 their primary neurotransmitter. However, synthesis of the major neurotransmitters – glutamate (in
236 mammals, excitatory) and GABA or glycine (inhibitory) – requires only a modest number of genes, and
237 GABA itself is synthesized from glutamate (*34*). It is unclear whether many or few other genes are
238 correlated with GABAergic (or glutamatergic) identity. For instance, whether neurotransmitter utilization
239 is strongly associated to the identity of neuron types may differ by brain structure: transcriptomically-
240 defined neuron types in the neocortex and other telencephalic structures are hierarchically grouped into
241 GABAergic and glutamatergic types (*2, 7, 26*), which reflects both their distinct developmental origins
242 and their distinct neurotransmitter repertoire. In other brain structures such as the hypothalamus, the
243 distinction between neurons utilizing GABA or glutamate is much less clear at the transcriptional level (*8,
244 35*) (**Fig. 1C,E,F, Fig. S3**).
245
246 To determine whether neurotransmitter identity was associated with a general transcriptional identity
247 across all neurons expressing the same neurotransmitter, we examined expression of genes encoding the
248 most prevalent vesicular glutamate transporters (*SLC17A6, SLC17A7*) and GABA synthesis enzymes
249 (*GAD1, GAD2*) (**Fig. 1C-F; Fig. S3**). If primary neurotransmitter usage was strongly associated with the
250 expression of many other genes, we would expect that neurons belonging to the same neurotransmitter set
251 (GABAergic or glutamatergic) would preferentially group together regardless of other factors (such as
252 developmental origin). However, we did not find evidence for strong global transcriptomic identities of
253 GABAergic and glutamatergic neurons. Neuronal types from each set were distributed across the tree,

254 suggesting divergent transcriptomic identities of cell types that share a common neurotransmitter **(Fig.**
255 **1C).**

256

257 In telencephalic structures such as neocortex, hippocampus, and amygdala, glutamatergic neurons all
258 express *SLC17A7* and segregate from GABAergic telencephalic neurons (**Fig 1C,D**). Telencephalic
259 GABAergic neurons separated into distinct clades: one for GABAergic interneurons, and another for
260 GABAergic projection neurons such as SPNs of the striatum (**Fig 1C, Fig S2**). In non-telencephalic
261 structures such as the hypothalamus, basal forebrain, thalamus, brainstem, and cerebellum, GABAergic
262 and glutamatergic types were highly intermixed. This pattern held in a mouse scRNA-seq dataset, and
263 when using different approaches for hierarchical clustering **(Fig. S2)**. Even within the telencephalon,
264 neurotransmitter identity did not drive global transcriptional similarity between major clades. For
265 instance, GABAergic projection neurons, such as SPNs of the striatum, were transcriptionally closer to
266 glutamatergic neurons in the neocortex, amygdala, and hippocampus than they were to telencephalic
267 GABAergic interneurons **(Fig. 1C; Fig. S3)**.

268

269 Although glutamatergic neurons from distinct cephalic origins do not cluster together, maintaining
270 glutamatergic neurotransmission or associated function could require a common, core set of genes. To
271 assess how neurotransmitter utilization relates to genome-wide RNA expression patterns, we examined
272 distributions of gene-gene correlations across cell types (**Fig. 1E**). Surprisingly few genes are strongly
273 positively correlated with both *SLC17A6* (VGLUT2) and *SLC17A7* (VGLUT1) expression, even those
274 associated with glutamate synthesis and packaging **(Fig. S3)**. 116 genes had correlated expression to
275 *SLC17A7* (Spearman's tau > 0.5). The median correlation of *SLC17A6* to those 116 genes was centered at
276 tau = 0.05 (**Fig. 1F**). Only a few genes correlated above 0.5 to both *SLC17A6* and *SLC17A7,* including
277 *BDNF, NRN1*, and *TAFA1*. Moreover, only 10 genes (*ARPP21, BDNF, CACNA2D1, CHN1, CHST8,*
278 *CPNE4, LDB2, NRN1, PTPRK, TAFA1*) are differentially expressed (> 2.5 fold change) in both
279 *SLC17A6*+ glutamatergic neurons and *SLC17A7*+ glutamatergic neurons relative to *GAD1*+ neurons. We
280 examined the principal component loadings to see if any were associated with primary neurotransmission.
281 While individual PCs loaded strongly on specific parts of the tree (**Fig. S2B**), none of the top 20 PCs
282 (accounting for 92% variance) distinguished cell types by their neurotransmission. The bulk of gene
283 expression that distinguishes neuronal types from each other appears incidental to neurotransmission.

284

285 *Neocortical expression fingerprints differ between neurons and glia*
286 The mammalian neocortex is partitioned into functionally, connectionally, and cytoarchitectonically
287 distinct regions, called areas. We examined regional distinctions in proportions and gene expression of
288 cell types from 8 cortical locations (**Fig. 2A**). Consistent with prior reports (*3, 6, 13, 36*), cell subtypes
289 identified in one cortical region were generally present in all other cortical regions, though in different
290 proportions (**Fig. 2B**). There were two main exceptions in neurons. GABAergic *MEIS2*+ cells
291 (GABAergic cluster 6, **Fig. 2B**) were far overrepresented in PFC samples (specifically in dissections of
292 medial and medio-orbital PFC, **Fig. S1C**), a compositional distinction not observed in mouse (*6, 36*). The
293 second exception was a cluster of *RORB+, KCNH8*+ glutamatergic neurons in V1 (and to a lesser extent
294 V2) that diverged from *RORB*+ populations found in the other cortical regions (Glutamatergic cluster 2,
295 **Fig. 2B-D**). The expansion and divergence of *RORB*+ populations in visual cortex is consistent with the
296 elaboration and sub-specialization of primary V1 layer IV in primates (*37*). One astrocyte type, marked

297 by high expression of *VCAN*, was highly enriched in V1 and V2 (Astrocyte cluster 3, **Fig. 2D**). Using
298 smFISH, we validated higher co-localization of *VCAN* and *GFAP* in V1-adjacent white matter compared
299 with PFC and V1 gray matter (**Fig. S4I**). *VCAN* is also expressed in OPCs. We did not observe higher
300 *VCAN* expression in V1 OPCs, suggesting that the regional variation in *VCAN* expression is specific to
301 astrocytes (**Fig. 2E**).

302

303 Glutamatergic neurons sampled from different neocortical locations show regionally distinct gene
304 expression (*36*). In primates, this is also true of neocortical GABAergic neurons (*3*). Astrocytes in mouse
305 and primate brains are heterogeneous across major subdivisions (**Fig. S4D-H**), but the extent to which
306 they are locally customized in distinct regions of neocortex is less well understood (*2, 7, 26, 38*)(*4*).
307 Studies in mice have revealed layer-specific astrocyte subpopulations in the cortex (*39*) and variation
308 between neocortical and hippocampal astrocytes (*40*).

309

310 To address whether cortical regional variation in gene expression is shared across cell types, we
311 performed pairwise comparisons between major clusters of cortical excitatory neurons, inhibitory
312 neurons, astrocytes and oligodendrocyte lineage types across all eight neocortical locations (**Fig. 2F,**
313 **Table S3**). Each of these cell classes displayed regionally differentially expressed genes (rDEGs) across
314 neocortical regions (**Fig. 2F**), an effect that could not be attributed to ambient RNA contamination (**Fig.**
315 **S5A**). Similar to what has been described in the mouse cortex (*39, 40*), astrocytes within the marmoset
316 cortex exhibited regional transcriptional variation, but overall neurons had more rDEGs than macroglia
317 (**Fig. 2F; Fig. S5B**). 62% of rDEGs in interneurons were also rDEGs in excitatory neurons. Interestingly,
318 though cortical astrocytes and oligodendrocytes arise from a common lineage with cortical excitatory
319 neurons (*41*), they shared a lower percentage of rDEGs in common with excitatory neurons as compared
320 with interneurons (25% astrocytes, 25% oligodendrocyte lineage). Regionally differentially expressed
321 genes within a cell class (glutamatergic, GABAergic, astrocyte, oligodendrocyte lineage) tend to be
322 biased in the same regions across subtypes within that class. However, certain subtypes and regions
323 accumulated more rDEGs than others. For example, across all cell types and particularly within neurons,
324 higher-order temporal association cortex and prefrontal cortex tended to be most distinct from V1 and V2
325 (**Fig. S5B**). To determine the extent to which rDEGs are private to individual donors or represent shared
326 features of variability within conserved cell types, we analyzed the consistency of neocortical rDEGs
327 across the three donors that had been sampled at all eight neocortical locations (**Table S3**). Our findings
328 revealed that 55% and 39% of glutamatergic and GABAergic rDEGs, respectively, were consistent
329 among donors, while only 24% and 14% of astrocyte and oligodendrocyte rDEGs, respectively, were
330 shared between donors. These results indicate lower inter-individual consistency in regional gene
331 expression signatures in non-neuronal cells.

332 *A hypothalamic-origin neuron type in medial amygdala*
333 In mammalian CNS, some cell types migrate long distances from proliferative zones to their mature
334 destinations (*42, 43*). However, neurons generally tend to respect cephalic boundaries and remain within
335 the same subdivision as their progenitors (*44*). This tight control over migration potential makes it
336 difficult to disentangle the persistent influence of developmental origin from potential later influences
337 arising from shared tissue context that might affect all neurons similarly in a given brain structure.
338 However, cephalic boundary crossings do exist (*28, 29, 45*). Though rare, such boundary crossing events

339   can reveal whether cell types that embarked on cross-cephalic migration retain transcriptomic profiles
340   more in common with their tissues of origin, or more in common with their final destinations.

341

342   We found a striking example of cross-cephalic migration in the amygdala. First, we observed an
343   unexpected clustering pattern in our analysis, wherein three amygdala neuron types joined a clade not
344   with other telencephalic neurons, but instead with *SLC17A6+* hypothalamic and basal forebrain types
345   (**Fig. 3A**). Moreover, despite expressing *SLC17A6* and lacking expression of *GAD1* and *GAD2,* these
346   cells showed a closer association with GABAergic rather than glutamatergic types (**Fig. 3B**). They did not
347   express other genes required for GABAergic transmission such as *SLC32A1* (VGAT) and also lacked the
348   molecular machinery for non-canonical GABA reuptake or release observed in other cell populations
349   (**Fig. S3**)(*46*, *47*). These findings suggest that these particular amygdalar neurons exhibit a "cryptic"
350   transcriptomic identity, in which they are glutamatergic but have transcriptomic profiles that are much
351   more similar to GABAergic types, relative to other telencephalic neurons.

352

353   The "cryptic" amygdalar subtypes display additional atypical gene expression features compared to other
354   amygdala neuronal types, including expression of *OTP* and *SIM1* (**Fig. 3B**). These transcription factors
355   are typically expressed in neuronal lineages originating from proliferative zones around the third
356   ventricle. In mice, there is a migratory stream of diencephalic neurons into the telencephalon (*45*, *48*) into
357   the medial amygdala, and this migration is dependent on *Otp* expression (*45*) (**Fig. 3D**). We generated
358   mouse amygdala snRNA-seq data from 53,745 nuclei and verified the presence of a *Sim1+* neuron type in
359   mice that also clusters with GABAergic neurons and expresses *Slc17a6* but not *Gad1* or *Gad2* (**Fig. 3C**).
360   In mice, these neurons constitute the majority population in the amygdala that express *Adcyap1* (encoding
361   the protein PACAP) (**Fig. 3C**), a neuropeptide that is extensively (but not exclusively) expressed in
362   hypothalamic populations (*49*) and associated with energy homeostasis (*50*), stress, anxiety (*51*), and
363   immune responses (*52*). (In marmosets, *ADCYAP1* is additionally expressed in subsets of *SLC17A7+*
364   neurons that do not share the "cryptic" phenotype, indicating that *ADCYAP1* has a distinct distribution of
365   expression in amygdala neurons between mice and marmosets.)

366

367   We further confirmed the expression of *Sim1+* cells in the medial amygdala in mice using the Allen
368   Institute ISH atlas (**Fig. 3E**). To precisely locate the specific amygdalar nuclei housing the cryptic
369   population, we investigated the expression of *SIM1* in neonatal marmosets (*53*, *54*). *SIM1* expression was
370   highly enriched in the neonatal marmoset medial amygdala (MeA) (https://gene-atlas.brainminds.jp; **Fig.**
371   **3F**), which mirrors the migration of *Sim1+* neurons from the diencephalon in mice (28). We verified
372   *SIM1* and *SLC17A6* but not *SIM1* and *GAD1* colocalization in adult marmoset medial amygdala using
373   single-molecule fluorescence *in situ* hybridization (smFISH) (**Fig. 3G**). These results suggest that the
374   cryptic amygdala neurons are a conserved population in both mice and primates that likely have
375   diencephalic origins. Consistent with a "birthplace imprinting" effect, they retain transcriptional identities
376   more similar to diencephalic types than to the telencephalic types with which they ultimately reside (**Fig.**
377   **3A**).

378   *Primate-specific striatal TAC3+ interneurons are similar to specific TAC3+ diencephalic types*
379   Previously, we discovered a *TAC3+, LHX6+* interneuron subtype in the striatum of humans, macaques,
380   and marmosets that was absent in mice and ferrets (*3*). Compared with other striatal types, they are
381   transcriptionally most similar to *PVALB+* interneurons and, because they expressed transcription factor

382 *LHX6*, we surmised that they likely also arose from the medial ganglion eminences (MGE) (*3*). Consistent
383 with this inference, a recent study of fetal macaque single nucleus RNA-seq data from ventral
384 telencephalic progenitor domains found that *TAC3*+ striatal interneurons likely arise from progenitors in
385 the MGE, and diverge from an ancestral progenitor class that also gives rise to conserved *MAF*+
386 progenitors that produce *PVALB*+ and *TH*+ striatal interneurons (*33*).

388 The broader census of brain structures in the current dataset allowed us to compare the transcriptional
389 identity of the striatal *TAC3*+ type to cell types outside of the striatum. Beyond expression in the striatal
390 type, *TAC3* is expressed in 20 different neuron types in our dataset, including expression in cortical
391 GABAergic neurons as well as several amygdala, basal forebrain, thalamic, and hippocampal types (**Fig.
392 3H; Fig. S3, Fig 6A,B**). *TAC3*+ types did not form a single clade in the dendrogram, and their
393 transcriptional resemblance to other neuron types largely reflected their tissue or cephalic origin. For
394 example, thalamic *TAC3*+ subtypes were found in a thalamic-only clade, while hippocampal *TAC3*+
395 neurons were most similar to other hippocampal and amygdala types. The thalamic types were *SLC17A6*+
396 while all other *TAC3*+ types were GABAergic.

398 Unexpectedly, the striatal *TAC3*+ type was most similar not to other striatal interneuron types as we
399 previously concluded (*3*), but rather to two other *TAC3*+ GABAergic types found in basal forebrain and
400 hypothalamus (**Fig. 3H**, tissue validation in **Fig. S6A**, **Data S2**). The broader clade containing these three
401 types (depicted by green arrowhead in **Fig. 3H**) consisted entirely of basal forebrain and hypothalamic
402 types, with the exception of the striatal *TAC3*+ type. Each of the similar *TAC3*+ populations had distinct
403 gene expression (such as high expression of OXT and AVP in the hypothalamic type, and DRD2
404 expression exclusively in the striatal type), ruling out dissection contamination (**Fig. 3I**). The 3 types
405 remained direct neighbors when the dendrogram was recomputed using other distance metrics (e.g.
406 correlation-based or PCA scores using transcription factor expression, see **Fig. S2**). As this clade
407 assignment was unexpected, we omitted the basal forebrain and hypothalamic *TAC3*+ types and
408 recomputed the dendrogram (retaining 286 of the original 288 neuronal types) to determine whether the
409 striatal *TAC3*+ was broadly more similar to hypothalamic types than to telencephalic types (**Fig. 6C**). As
410 striatal interneurons are generally divergent from hypothalamic/basal forebrain types, omitting these two
411 cell types enabled us to test whether the *TAC3*+ striatal type is broadly more similar to hypothalamic
412 neurons or to striatal neurons. The global structure of the dendrogram was essentially the same as the
413 original, except that the striatal *TAC3*+ type now joined the major telencephalic GABAergic clade with
414 other striatal interneurons (neighboring the striatal *PVALB*+ subtype) (**Fig. 6C**), recapitulating our
415 original similarity assignment (*3*).

417 Considering their unexpected transcriptional similarity to both a telencephalic (basal forebrain) and a
418 diencephalic (hypothalamus) type, the *TAC3*+ striatal type may either arise from a telencephalic
419 progenitor (*33*) that also gives rise to sister diencephalic (hypothalamic) types, or else shows striking
420 transcriptional convergence with diencephalic types that have distinct developmental origins. Favoring a
421 telencephalic origin as suggested by previous report (*33*), the hypothalamic, basal forebrain and striatal
422 *TAC3*+ types are all *FOXG1*+, a transcription factor associated with telencephalic origin (**Fig 3I**). They
423 also express *LHX6*+ and *NKX2-1*+ (**Fig S3**), consistent with a medial ganglionic eminence origin,
424 however we note that in mice some hypothalamic types also express *Nkx2-1* and *Lhx6 (9, 55, 56)*.

425 Ultimately, lineage tracing of the striatal *TAC3+* type in a primate would resolve whether a shared
426 progenitor gives rise to both telencephalic and diencephalic types.

427 *Quantitative maps of interneuron distribution across primate neocortex and striatum*
428 Our analysis of RNA expression in the single nucleus dataset indicates that developmentally-linked
429 telencephalic GABAergic populations (*42*, *43*) retain globally similar identities in adulthood. For
430 example, *SST+* striatal interneurons are more similar to *SST+* hippocampal and neocortical interneurons
431 than they are to other striatal types (**Fig. S3**). Within each brain structure, developmentally linked
432 populations become functionally specialized and follow distinct spatial rules for their allocation. In the
433 mouse, quantitative imaging of molecularly identified interneuron types has revealed that densities of
434 *Sst+*, *Vip+*, and *Pvalb+* cortical interneurons vary across the cortical mantle; their relative proportions
435 relate to unique functional and microcircuit properties of different cortical areas (*15*). In primates, overall
436 neuron densities are more variable by cortical location than in mice: they vary by as much as 5 fold across
437 the cortical sheet, with highest neuron proportions and densities found in occipital cortex and particularly
438 in V1 (*57*). However, quantitative mapping of molecularly-defined subtypes of neurons have not been
439 performed in a primate, and it is not known if they have conserved or distinct spatial distributions
440 compared to mice.
441
442 We used smFISH to image the major cortical and striatal interneuron types across sagittal sections in
443 marmoset (**Fig. 4** and **Fig. S7-9**). We quantified proportions and densities of each type relative to all cells
444 (DAPI; **Fig. S7**). In the neocortex, we binned these using an areal parcellation of marmoset neocortex
445 (https://doi.org/10.24475/bma.4520) to determine whether interneuron proportions varied systematically
446 by brain area. In total, we quantified 377,554 neocortical interneurons across 30 sections by smFISH. In
447 striatum, we quantified 6,848 interneurons across 32 sections. Each series sampled sagittal sections ~160
448 μm apart, beginning 1,184-1,584 μm from the midline up to 6,384 μm laterally, including the majority of
449 the striatum with the exception of the most-lateral portion of the putamen. We used a cell segmentation
450 algorithm to count positive cells across sagittal sections and expressed interneuron proportions as a
451 percent of all cells (DAPI+).
452
453 Neocortical types were identified with probes for *SST, PVALB, CXCL14, VIP,* and *LAMP5* (**Table S4**),
454 which collectively account for all major cortical interneuron populations (**Fig. 4A**). *CXCL14* is a marker
455 for caudal ganglionic eminence derived cortical interneurons. It is expressed in most *VIP+* and *LAMP5+*
456 cortical neurons, as well as a smaller population of *VIP-, LAMP5-* types, some of which are *PAX6+* (and
457 which correspond to the *SNCG+* population in humans and mice (*2, 3*))). *VIP+* and *LAMP5+*
458 interneurons are the two other major CGE-derived populations present in primates and mice. As *LAMP5*
459 is also expressed in subsets of excitatory neurons, we performed dual labeling smFISH with *GAD1* to
460 avoid counting glutamatergic types. Major striatal interneuron types were identified with probes for *SST,*
461 *PVALB, CHAT/SLC5A7, TH, CCK,* and *TAC3* (**Table S4**). These together account for most major
462 populations of non-SPN neurons in the striatum, with the exception of a population of *MEIS2+*
463 GABAergic striatal neurons that cluster together with non-SPN GABAergic neurons (**Fig. 1C, Fig. S3**)
464 but are difficult to distinguish uniquely as several other markers are also expressed at variable levels in
465 other striatal cell types.

466 *Neocortical interneuron types have highly focal biodistributions*

467   Quantitative analysis of smFISH revealed highly focal and variable distributions of interneuron subtypes
468   across marmoset neocortex (**Fig. 4B-E, Fig. S8A-D**). In both absolute numbers and relative proportions,
469   *PVALB+* interneurons were strongly enriched in the occipital lobe, particularly along the calcarine sulcus
470   in the medial sections, as well as the occipital pole more laterally (**Fig. 4B-C; Fig. S8A, Data S1**). *SST+*
471   interneurons in the neocortex increase medio-laterally (**Fig. 4B-C, Fig. S8A, Data S1**), but closer
472   inspection revealed this is driven not by a spatial gradient so much as by highly focal enrichments around
473   primary motor area (M1) and primary somatosensory cortex (S3, S1/2), the cingulate cortex, entorhinal
474   cortex and medial prefrontal cortex (**Fig. 4C**). *CXCL14+* neurons are enriched along the calcarine sulcus
475   medially, as well as in ventral aspects of the occipital cortex more laterally. There are higher proportions
476   dorsomedially in the parietal cortex (**Fig. 4C**). *VIP+* neurons were enriched in PFC and also increased
477   laterally at or near somatosensory cortex and posterior parietal cortex (**Fig. 4C**). *LAMP5+* interneurons
478   showed a bias to the top of the cortical layers, consistent with dominant composition of this class as
479   neurogliaform Layer 1 types (*3, 58, 59*), though this class also contains the *LAMP5/LHX6* type that is
480   found in deeper layers (*3, 13*) (**Fig. 4C**).

482   To better appreciate how these distributions relate to neocortical areas, we used a histologically-based
483   marmoset neocortical parcellation (https://doi.org/10.24475/bma.4520; **Fig. S8B-C**) to bin smFISH
484   interneuron proportions by cortical area (**Fig. 4D-E; Fig. S8B**). Small adjacent areas were merged,
485   resulting in 26 areal groupings (**Fig. S8C**). Overall interneuron proportions relative to all cells varied by
486   4.5-fold across areal groupings (**Fig. 4D-E, Fig. S8B-C**). As a fraction of all cells, A31 (dorsal posterior
487   cingulate) had the lowest overall proportion of interneurons (3%), while insular cortex (In) had the
488   highest (17%) (**Fig. 4D**).

490   In mice, quantitative mapping of interneuron densities showed higher *Sst+* densities and lower *Pvalb+*
491   densities in neocortical areas involved in higher cognitive functions, such as medial frontal and lateral
492   association areas (*15*). This local circuit feature follows cortico-cortical connectivity network topography
493   in mouse (*60*). In contrast, lower *Sst/Pvalb* ratios were associated with mouse primary motor and sensory
494   areas, which are associated with less distributed (and more local) cortical connectivity (*15*). To determine
495   whether primate neocortex followed similar rules as mouse neocortex of interneuron allocation, we
496   examined normalized proportions of the four largely mutually exclusive interneuron classes (*PVALB+,*
497   *SST+, LAMP5+, VIP+*) within our marmoset areal groupings (**Fig. 4D-E; Fig. S8B**). Lateral temporal
498   cortex, including A36, had the highest *SST/PVALB* ratio. Other areas with high *SST/PVALB* ratios
499   included piriform cortex (Pir), M1 and several medial/orbitofrontal areas (A25, A13, A32). This suggests
500   that some marmoset association areas, notably medial frontal area and lateral temporal cortex, have high
501   *SST/PVALB* ratio composition consistent with high *Sst/Pvalb* ratios in higher order association network in
502   mice (*15*) (**Fig S8B**).

504   Strikingly, polar and lateral prefrontal areas (including A8, A46, A10, A47, A45), which are thought to be
505   the most divergent relative to prefrontal areas in rodents (*61*), and which in primates are characterized by
506   long-range cortico-cortical connectivity to other association areas (*62–64*), do not exhibit high
507   *SST/PVALB* ratios (**Fig. 4D-E; Fig. S8B**). Of all cortical areal groupings we measured, the lowest
508   *SST/PVALB* ratio was found in A45, a higher-order lateral prefrontal area with extensive projections to all
509   lobes of the neocortex (*64*); http://analysis.marmosetbrain.org/). Thus, unlike the high *Sst/Pvalb* ratios
510   observed in frontal areas in mouse (*15*), primate lateral prefrontal areas are characterized not by

511   exceptional *SST/PVALB* ratios but rather as having some of the highest *VIP+* interneuron proportions
512   (**Fig. 4D-E; Fig. S8B**). Lateral prefrontal areas also have the highest total fraction of *VIP+* and
513   *LAMP5/GAD1+* interneurons (all above 50% of all interneurons); both of these populations
514   predominately arise from the caudal ganglionic eminence, a progenitor zone that has expanded in primate
515   evolution (*65*). These results suggest that primate lateral prefrontal areas do not follow the same local
516   cortical circuit organizing principles that typify frontal areas in the mouse (*15*), or medial frontal areas in
517   marmoset (**Fig. 4D-E; Fig. S8B**).

518   *Interneuron proportions in marmoset striatum follow medio-lateral gradients*
519   The striatum, a crucial brain region involved in motor control, reward, and decision-making, displays
520   complex functional topography and connectivity. In humans, anterio-medial portions containing nucleus
521   accumbens/ventral striatum are functionally coupled to limbic and higher order cognitive networks, while
522   lateral subdivisions are more coupled to sensory and motor networks (*66*). Prior work relating bulk gene
523   expression measurements in human and macaque striatum to cortico-striatal network organization found a
524   relationship between functional domain (e.g. somato/motor vs limbic network) and gene expression,
525   including enrichment of *PVALB* in lateral portions of the striatum that are coupled to somatosensory and
526   motor cortex (*67*). These results suggest that differences in local cell type composition across striatum
527   may underlie aspects of its functional topography.
528
529   In mice, striatal interneuron subtypes have different spatial distributions: cholinergic (*Chat+*) neuron
530   proportions increase dorsally and anteriorly (*68*), *Pvalb+* interneurons are more abundant in dorsolateral
531   striatum than in dorsomedial striatum, and *Sst+* interneurons are spatially homogeneous (*69*) (**Fig. S9A**).
532   Primates retain the major populations of striatal interneurons found in mice (*6, 70*), and additionally have
533   gained the novel type distinguished by *TAC3* expression (*3, 33*) described in previous sections (**Fig. 3H,**
534   **4F**). A systematic quantification of striatal interneuron types has not been performed comprehensively in
535   a primate, and it is unknown if they follow similar or distinct spatial distributions observed in mice. We
536   used single-molecule FISH (smFISH) to investigate distributions of the major types of conserved striatal
537   interneurons (*SST+, PVALB+, SLC5A7+/CHAT+, TH+, CCK+, TAC3+*, probes in **Table S4**) in serial
538   sagittal sections of marmoset striatum (**Fig. 4G-H**).
539
540   Each striatal interneuron population exhibited a non-uniform distribution across the marmoset striatum,
541   particularly in the medial-lateral axis. Similar to mice, the proportion of striatal *PVALB+* interneurons
542   increases in lateral sections, from ~0% to 0.8% of all cells (**Fig. 4G, Data S1**). Unlike mice, marmoset
543   *SST+* interneuron distribution is non-uniform, appearing sparse near the midline and increasing in
544   proportion (0.1-0.7%; **Fig. 4G, Data S1**). Cholinergic neurons (*CHAT+*) show the opposite medial-lateral
545   gradient (0.45%-0.1%; **Fig. 4G, Data S1**). Similar to *CHAT+* neurons, *TH+* striatal interneurons, which
546   are transcriptionally similar to the *PVALB+* type, exhibited a decreasing medial-lateral gradient (**Fig. 4G,**
547   **Data S1**). *CCK+* striatal interneurons, which are a minority population in marmoset (*3*) and mouse (*70*)
548   are enriched close to the midline and become much sparser laterally (**Fig. 4G, Data S1**). The *TAC3+*
549   interneurons showed an increasing medial-lateral gradient, similar to *CHAT+* neurons (2%-0.5%; **Fig.**
550   **4G, Data S1**). No striatal population exhibited anterior-posterior or dorsal-ventral gradients with the
551   exception of *PVALB* interneurons, which showed a modest dorsal-ventral gradient. Unlike striatal
552   interneurons, MSNs proportions were largely uniform across the major axes: *DRD1* and *DRD2*, which
553   distinguish direct and indirect MSNs, respectively, were slightly enriched medially but otherwise

554   exhibited largely uniform distributions across the striatum (**Fig. S9B**). Overall, these findings suggest that
555   the spatial distribution of interneuron subtypes within the primate striatum may play a role in shaping its
556   functional topography and connectivity. While mice and primates share similar striatal interneuron
557   populations, there are also differences in their spatial distributions.

558   *Morphology of conserved neocortical and striatal interneuron types*
559   The morphology of interneuron types relates essentially to their function and contributions to neural
560   circuits. While methods such as biocytin filling and Golgi staining are the gold standard for
561   morphological reconstructions, the administration of low-titer AAVs carrying membrane-bound
562   fluorescent reporters can be used to sparsely label cells for morphological reconstruction. We utilized a
563   reporter AAV under the control of the forebrain interneuron-specific mDlx enhancer (*19*) to label
564   neocortical and striatal interneurons, and then performed smFISH (probes in **Table S4**) on thick sections
565   (120 µm) to confirm the molecular identity of GFP+ cells. Reconstructions using viral labeling are more
566   challenging than with single cell filling methods, because GFP expression from neighboring cells and
567   passing fibers have to be distinguished from signal attributable to the target cell. In some cases, the GFP
568   signal appears punctate, making it challenging to follow discontinuous processes. For these reasons, our
569   reconstructions are conservative: as we aimed to avoid reconstructing false positive fibers (fibers
570   originating from other cells), we may in some cases under-ascertain the full dendritic arborization of
571   target cells.
572
573   We imaged 1,203 GFP+ neurons in the striatum and 4,374 GFP+ neurons in the neocortex. We used
574   NeuTube to reconstruct the top 216 telencephalic neurons that had the best GFP signal, did not have other
575   cells labeled in the field of view, and were positive for at least one smFISH probe **(Table S5; examples**
576   **in Fig. S10)**. Raw image stacks and NeuTube reconstructions are available at
577   https://doi.org/10.35077/g.609. Using combinations of 1-2 probes for marker genes of different types, we
578   identified GFP+ neocortical and striatal interneurons, respectively, with smFISH based on their type
579   markers (neocortex: *SST*, *PVALB*, *CXCL14*, *VIP*, *LAMP5*; striatum: *SST*, *PVALB*, *SLC5A7*, *TH*, *CCK*,
580   **Table S4**) (**Fig. 5A-B,** *top rows*). We used a second tracing and reconstruction method (Imaris) on the 41
581   most complete GFP+ cells. Of these, we discarded 5 due to incomplete soma or highly discontinuous
582   processes in the image stack, retaining 36 cells (**Fig. 5A-B,** *bottom rows***; Table S6**). Morphological
583   parameters were measured using the Surface function, which detects surface area and volume based on
584   the fluorescence of the mDlx-AAV-GFP expression, and the Filament Tracer function, which traces
585   structural features starting from the soma to the terminal processes based on the diameter of the soma and
586   the thinnest projection of the cell **(Table S6)**. To assess whether there are region-dependent
587   morphological differences within molecularly similar types, we compared several parameters between a
588   collection of neocortical and striatal *PVALB*+ GFP+ cells that were reconstructed from the same
589   marmoset (Cj 17-154; **Fig. 5C-D**) to avoid variable tissue shrinkage arising from different tissue storage
590   conditions across the animals (see *Methods*). Striatal *PVALB*+ GFP+ cells were larger than cortical
591   *PVALB*+ GFP+ cells in terms of length, area, but not soma diameter or the volume of GFP+ fluorescence
592   (**Fig. 5D**). While the number of branches issued by their somas also did not differ, the striatal *PVALB*+
593   GFP+ interneurons exhibited more dendritic branch points than their cortical counterparts (**Fig. 5D**),
594   suggesting that striatal *PVALB*+ interneurons have greater arborization than *PVALB*+ cortical
595   interneurons.

596 *Development of a novel TAC3-rAAV-GFP reporter*

597 Given that the *TAC3+* type comprises almost 30% of striatal interneurons in marmoset, we expected a
598 sizable proportion of striatal GFP+ cells labeled by AAV9-hDlx5/6-GFP-fGFP (**Fig. 5**) to be *TAC3+*.
599 However, while 55 GFP+ striatal cells were imaged across all smFISH *TAC3*-probe treated sections, we
600 failed to find any colocalization with *TAC3* transcripts. In these experiments, which were double-labeled
601 with *TAC3* and *PVALB* probes, 9/55 were *PVALB+* (16% *PVALB+*, vs expected 21% of interneurons
602 expected from snRNA-seq proportions), while 0/55 were *TAC3+* (0% vs 28% expected from snRNA-
603 seq).

604

605 The mDlx enhancer is a regulatory element specific to forebrain interneurons (*71*). Though the regulatory
606 element, and the forebrain-interneuron expressed genes flanking it, *Dlx5* and *Dlx6*, are highly conserved
607 in vertebrates, we wondered whether the lack of accessibility of the mDlx locus in the *TAC3* interneurons
608 could explain our inability to find colocalization of *TAC3* expression and GFP. To assess this, we
609 generated single nucleus ATAC-seq (snATAC-seq) data (69,808 nuclei) from fresh marmoset striatum
610 (**Fig. 6A**). We used Signac (*72*) to integrate our previously annotated striatal snRNA-seq data and identify
611 major striatal types. We then examined accessibility of the marmoset sequence homologous to the mDlx
612 locus across interneuron types. While other striatal interneuron types (particularly the *SST+* and *PVALB+*
613 types) showed accessibility (ATAC-seq peaks, reflecting chromatin accessibility) at the mDlx locus, the
614 *TAC3+* type did not (**Fig. 6B**).

615

616 To develop a viral tool that could transduce the striatal *TAC3+* cell type, we next nominated new
617 candidate regulatory elements specific to the *TAC3+* interneuron type. We used Signac to identify
618 differentially accessible peaks in the *TAC3+* cluster relative to all others, and filtered the set by fold
619 change, percent accessibility across the target cell type population, and peak size (**Fig. 6C**). To maximize
620 the likelihood of obtaining a functional reporter while minimizing the number of marmosets used, and
621 because several of our top candidates were very short, we selected four top regulatory element candidates
622 for the *TAC3+* type for tandem packaging (**see *Methods***). The four candidates were on four different
623 chromosomes and spanned between 94-215 bp. One site was in an exon of *CDH13*, one was in the first
624 exon of *TAC3*, one was in an exon of *LOC108592287*, and one was intergenic (closest gene 200kb
625 distance). These four elements were packaged in tandem in an AAV9 vector (AAV9-tandemE-TAC3-
626 EGFP) containing a cytoplasmic GFP reporter (**Fig. 6D**). We delivered the virus via MRI-guided local
627 injection into the anterior striatum of two marmosets (Cj 19-207 and 17-B111, **Fig. 6E and Fig. S11A,**
628 **Table S1**), and imaged coronal sections of striatum for GFP positive cells after 6 & 10 weeks' incubation
629 time (**Fig. 6F and S11C**). smFISH confirmed colocalization of *TAC3* transcripts in GFP+ neurons in
630 striatum (**Fig. 6G and Fig. S11B,D**). In one animal the virus diffused beyond the boundary of the
631 striatum. In this animal we also detected strongly labeled GFP+ cells in a border zone between striatum,
632 BNST, and the globus pallidus where the viral injection had diffused beyond striatum (**Fig. S11B-E**), and
633 smFISH confirmed that these too were *TAC3+*.

634

635 To determine the broader biodistribution of cells transduced by the tandemE-TAC3 enhancer, we next
636 packaged the same enhancers in a novel AAV capsid (BI103) capable of efficient transduction of brain
637 cell types after systemic IV delivery in marmosets (Chan et al, in prep). We delivered this AAV (AAV-
638 BI103-tandemE-TAC3-EGFP) to one marmoset (Cj 20-214, **Table S1**). After a 4 week incubation, the
639 brain was perfused, sliced and stained with anti-GFP antibody to amplify GFP signal, and smFISH probes

640   against *TAC3+* to confirm colocalization. We found GFP+/*TAC3+* colocalized cells in striatum as well as
641   several extra-striatal locations including hypothalamus, substantia nigra, superior colliculus, brainstem,
642   and neocortex (**Fig. 6I**). To assess the morphology of *TAC3+* striatal interneurons produced by viral
643   labeling, we reconstructed several of the most complete cells from both the local injections (AAV9-
644   tandemE-TAC3-EGFP) and from the systemic injection (AAV-BI103-tandemE-TAC3-EGFP) using
645   Imaris (**Figs. 6H,J,K** and **Table S6**). The cells tended to have 2-3 thick branches that extended from the
646   soma, and which bifurcated close to the soma and became thinner thereafter. Reconstructed cells had
647   median 13.1 µm (+/- 5.1 s.t.d) soma diameter, 2221.3 volume (µm^3; +/- 1192.8 s.t.d) and 16.5 (+/- 6.11)
648   total dendritic branch points. These experiments show that whereas systemic injections of AAV9 under
649   the mDlx enhancer could not transduce the striatal *TAC3+* type, presumably due to loss of accessibility at
650   the mDlx locus (**Fig. 6B**), local injections of AAV9 as well as systemic injections with alternative capsid
651   with the cell type targeted enhancers were successful. Broadly, the use of cell type specific regulatory
652   elements, coupled with viral engineering, enables a new horizon for the study of primate brain cell types.

653   Discussion

654   In this study, we generated a molecular and cellular census of the adult marmoset brain using single-
655   nucleus RNA sequencing, quantitative smFISH, and cell type specific AAV labeling. Our study reveals
656   the complex repertoire of cell types in the marmoset brain. Our snRNA-seq dataset of over 2.4 million
657   brain cells across 18 brain regions in the marmoset indicates that lineage is an important factor shaping
658   adult transcriptomic identity of neuronal types, apparently more so than neurotransmitter utilization.
659   Using quantitative smFISH we revealed, for the first time in a primate, the spatial distributions of
660   molecularly-resolved GABAergic interneuron types. Using GFP delivered by interneuron-specific AAVs,
661   we generated morphological reconstructions of all major interneuron types in the neocortex and striatum.
662   We generated a novel viral genetic tool under the control of cell type specific regulatory elements to
663   transduce a previously described putative primate-specific striatal interneuron type. These datasets,
664   generated as part of the BRAIN Inititive Cell Census Network, complement other recent and extensive
665   cellular profiling studies in other species (*8–11*, *73*) and will enable insights in cell type innovations and
666   modifications in future comparative studies.

667   Telencephalic glutamatergic and GABAergic neurons strongly segregate in mammals as well as in
668   homologous structures in amphibians (*26*), suggesting an evolutionarily conserved distinction. An initial
669   atlas in mouse indicated that glutamatergic and GABAergic neurons from diencephalic and midbrain
670   structures also partition almost perfectly by neurotransmitter usage (*7*), suggesting this could be a general
671   rule. However, more recent transcriptomic censuses in mouse (*8*, *9*) and human (*11*) indicate that across
672   mammalian species, many non-telencephalic glutamatergic and GABAergic types tend to form highly
673   intermixed clades that do not separate clearly by neurotransmitter identity. Our results in the adult
674   marmoset (and mouse, **Fig S2**) concord with the notion that gene expression distinctions between
675   telencephalic glutamatergic and GABAergic neurons do not hold for neurons in other brain structures
676   (**Fig. 1; Fig. S2**). This has implications for generalizing transcriptomic associations to other phenotypes.
677   For example, transcriptomic changes in glutamatergic or GABAergic neurons have been associated to
678   diseases such as autism and schizophrenia (*74–76*). Such associations may not generalize to GABAergic
679   or glutamatergic types outside of the sampled brain region (usually neocortex), consistent with
680   observations that "global" changes to glutamatergic or GABAergic neurons in relation to disease actually
681   often only surface in a few brain regions (*77*).

682   A previous analysis, based on shared patterns of several key transcription factors, proposed that
683   telencephalic GABAergic neurons are developmentally and evolutionarily related to diencephalic
684   GABAergic neurons (*78*). Our results indicate that when profiled at adulthood, only a limited number of
685   telencephalic GABAergic types are transcriptionally similar to diencephalic types, some of which may
686   arise from cephalic boundary crossings. Most neocortical, hippocampal, and some amygdalar and striatal
687   GABAergic types are so distinct from diencephalic GABAergic types that they share more gene
688   expression in common with telencephalic glutamatergic types (**Fig. 1C**).

689   Developmental origin or shared lineage plays a strong role in shaping the adult transcriptomic identity of
690   neurons, but phenotypic convergence, whereby adult cell types converge on similar transcriptomic
691   identities despite a non-shared developmental origin, may also drive apparent similarities amongst cell
692   types. Recent advances in lineage tracing coupled with single cell RNA sequencing demonstrates that
693   phenotypic convergence is surprisingly common between transcriptomically defined types (*5, 79*). As
694   such, it is difficult to disambiguate between these possibilities in absence of data that confirm lineage
695   directly (*5, 26*). For example, in our data the clade of GABAergic projection neurons that contains striatal
696   SPNs and amygdala and basal forebrain *GAD1+, FOXP2+* neurons also contained several subtypes of
697   hypothalamic *GAD1+, FOXP2+* neurons (**Fig. 1C; Fig. S2E, Fig. S3**). Either these hypothalamic
698   *FOXP2+* subtypes have a convergent expression identity with long-range GABAergic projection neurons
699   of the telencephalon, or else they arise from a common lineage.

700   Our data support known instances of cross-cephalic migrations in thalamus and amygdala, and also
701   suggest new ones (**Fig. 3**). The similarity of the primate-specific *TAC3+* striatal type to hypothalamic and
702   basal forebrain *TAC3+* types in particular is unexpected (**Fig. 3H**) and suggests a potential example of
703   cross-cephalic vesicle migration. Recent work (*33*) proposes that the striatal *TAC3+* type has a ventral
704   ventricular telencephalic origin, similar to most other GABAergic interneuron types destined for striatum.
705   While phenotypic convergence remains possible, another possibility is that a ventral telencephalic
706   progenitor gives rise to both telencephalic and diencephalic types. This is supported by the expression of
707   *FOXG1*, a transcription factor necessary for ventral telencephalic fate (*80*), in all three transcriptionally
708   similar *TAC3+* populations.

709   Developmentally linked interneuron populations in striatum and neocortex (e.g *PVALB+* types in both
710   structures) displayed distinct spatial distributions measured by cell counting of smFISH (**Fig. 4**). While
711   all interneuron subtype distributions in the striatum followed a gradient along a medial-lateral axis, the
712   interneurons in the neocortex followed much more complex distributions that for the most part were not
713   captured by simple gradients. While in the mouse, *Sst+/Pvalb+* ratios are a hallmark of higher order
714   cortex (*15*), we found a different pattern of relative interneuron proportions across much of the higher-
715   order association cortex in marmosets. Areas in the lateral prefrontal cortex were unique in their relatively
716   high proportions of *VIP+* neurons (and had unexceptional or even low ratios of *SST+/PVALB+* neurons).
717   These results suggest that primate lateral prefrontal areas may not follow the same local cortical circuit
718   organizing principles of association cortex in the mouse.

719   Morphological characterization of the striatal and neocortical interneuron populations suggests variation
720   in overall size and dendritic arborization amongst subtypes. To avoid interindividual and technical
721   variability arising from different sample storage conditions (see Methods), we analyzed several
722   morphological parameters in a subset of cortical and striatal *PVALB+* interneurons reconstructed from a
723   single marmoset (**Figs. 5C-D**). Striatal *PVALB+* cells were found to be larger in length, volume, and
724   surface area and consisted of more dendritic branch points than the cortical cells. Our data altogether

725 suggest that the striatal *PVALB+* interneurons exhibit higher dendritic complexity compared to their
726 respective counterparts in the neocortex. These morphological differences could result in differing
727 electrophysiological properties. For example, it is possible that *PVALB+* cells in the striatum have higher
728 capacitance and receive more synaptic inputs onto them, thereby impacting local signal integration and
729 computation, than cortical *PVALB+* interneurons. The functional significance of morphological
730 differences across the interneuron subtypes identified in our study will ultimately require studying how
731 these cells affect the circuits in which they reside (i.e., cellular/subcellular targeting biases and functional
732 properties)(*81–83*).

733 The development of virally-based tools enables cell-type-specific access in nonhuman primates (*2*, *18–*
734 *20*). To maximize translatability across species, most approaches nominate candidates using evolutionary
735 conservation (usually between mice and humans) as a selection criterion. However, the evolutionarily
736 conserved mDlx enhancer did not drive expression in the novel striatal *TAC3+* type (**Fig. 6B**). In general,
737 we observed that the mDlx enhancer selectively under-ascertained several interneuron populations in
738 marmoset in these experiments. For example, it systematically under-labeled *VIP+* and *SST+* types in the
739 neocortex (expected 22% and 26% of interneurons, obtained 2% and 3%, respectively), as well as *SST+*
740 interneurons in the striatum (14%, obtained 1.8%). This underlabeling could in part be due to the titer and
741 systemic delivery approach we adopted, which was necessary in order to achieve sparse labeling for
742 morphology. The overall low efficiency in our ability to molecularly characterize GFP positive cells
743 suggests a need for further optimization for this species and application.

744 The lack of transduction of *TAC3+* striatal interneurons using the AAV-mDlx virus prompted us to
745 develop a novel AAV under the control of *TAC3+* interneuron striatal regulatory elements (**Fig. 6D-H**).
746 The novel reporter virus (AAV9-tandemE-TAC3-EGFP) that we developed to study the striatal *TAC3+*
747 type was driven by striatal *TAC3+* enhancer elements. When delivered systemically, the virus also labeled
748 *TAC3+* neurons elsewhere in the brain, including in the hypothalamus, neocortex, substantia nigra. As
749 neocortical *TAC3+* cells bear little resemblance to *TAC3+* striatal interneurons in terms of their global
750 gene expression profiles (**Fig. 3H**), the widespread transduction of *TAC3+* neurons suggests that the
751 regulatory elements that we identified in the *TAC3+* striatal type (defined only relative to other striatal
752 cell types) may endogenously regulate the *TAC3* gene itself, or else a gene whose expression is highly
753 correlated to *TAC3*.

754 Our study reveals the complex landscape of transcriptionally-defined cell types in the marmoset brain.
755 Compared to mouse or other outgroup species, few primate lineage-gained cell types have emerged;
756 likely more common is the redirection, repurposing, or elaboration of conserved types (*2*, *13*, *33*, *84*, *85*).
757 The persistent fingerprint of developmental origin present in neuronal gene expression underscores
758 important roles for novel neurogenic niches, developmental patterning and altered mechanisms guiding
759 proliferation and cell migration in primate brain evolution that could be the subject of future study. For
760 example, *in vivo* lineage tracing in primates could reveal the developmental origin of the striatal TAC3+
761 type and its relationship to other TAC3+ populations in other brain structures. We report the largest
762 transcriptomic cell census of the marmoset brain and uncover the complex biodistribution of telencephalic
763 interneurons by quantitative smFISH mapping. We present a compendium of morphological
764 reconstructions of marmoset telencephalic interneurons, and describe new viral reagents for targeting and
765 visualizing a primate-specific striatal interneuron type. Together, these resources will enable further
766 comparative studies of the evolution and development of brain cell types.

767

768 METHODS

769 **Animals used for study**

770 *Marmosets.* Marmosets were pair-housed in spacious holding rooms with environmental control of
771 temperature (23–28°C), humidity (40–72%), and 12 hr light/dark cycle. Their cages were equipped with a
772 variety of perches and enrichment devices, and they received regular health checks and behavioral
773 assessment from MIT DCM veterinary staff and researchers. All animal procedures were conducted with
774 prior approval by the MIT Committee for Animal Care (CAC) and following veterinary guidelines.

775 *Mice.* Experimental mice were purchased from The Jackson Laboratory company and housed at the
776 Mclean hospital animal facility (3–5 mice per cage) on a 12:12 hr light-dark cycle in a temperature-
777 controlled colony room with unrestricted access to food and water. All procedures were conducted in
778 accordance with policy guidelines set by the National Institutes of Health and were approved by the
779 McLean Institutional Animal Care and Use Committee (IACUC).

780 **Tissue processing for single nucleus sequencing and smFISH**
781 *Marmoset specimens for snRNA-seq.* Marmoset experiments were approved by and in accordance with
782 Massachusetts Institute of Technology CAC protocol number 051705020. Adult marmosets (1.5–14.5
783 years old, 12 individuals; **Table S1**) were deeply sedated by intramuscular injection of ketamine (20–40
784 mg kg−1) or alfaxalone (5–10 mg kg−1), followed by intravenous injection of sodium pentobarbital (10–
785 30 mg kg−1). When the pedal with-drawal reflex was eliminated and/or the respiratory rate was
786 diminished, animals were trans-cardially perfused with ice-cold sucrose-HEPES buffer (*3, 6*). Whole
787 brains were rapidly extracted into fresh buffer on ice. Sixteen 2-mm coronal blocking cuts were rapidly
788 made using a custom-designed marmoset brain matrix. Slabs were transferred to a dish with ice-cold
789 dissection buffer (*3, 6*). All regions were dissected using a marmoset atlas as reference(*86*), and were
790 snap-frozen in liquid nitrogen or dry ice-cooled isopentane, and stored in individual microcentrifuge tubes
791 at –80 °C.
792
793 *Marmoset specimens for snATAC-seq.* Tissue from 1 marmoset (female, 1.5 y.o., **Table S1**) was used for
794 both snRNA-seq and snATAC-seq (**Table S1**). Fresh tissue was dissected from anterior striatum
795 (including caudate and putamen) and used immediately for snATAC-seq.
796
797 *Mouse specimens for snRNA-seq.* Three adult (P80-90) male wild-type mice were deeply anesthetized
798 with isoflurane and sacrificed by decapitation. Brains were quickly excised, washed in ice-cold sterile
799 0.1M phosphate buffer saline (PBS) and dissected onto an ice-cold glass surface. Amygdala nuclei were
800 identified and isolated using "The Allen mouse brain atlas" (https://mouse.brain-map.org/static/atlas,
801 **Table S7**) as a reference for anatomical landmarks. The basolateral amygdaloid nucleus was exposed by
802 performing two coronal cuts using the borders of *primary somatosensory cortex* and *primary visual*
803 *cortex* as landmarks. Dissected specimens were collected in 1.5ml micro-centrifuge tubes, snap-frozen on
804 dry ice, and stored at -80 °C until used.
805
806 *Marmoset specimens for smFISH.* Two marmosets were deeply sedated by intramuscular injection of
807 alfaxalone (5–10 mg kg−1) (**Table S1**), followed by intravenous overdose of sodium pentobarbital (10–
808 30 mg kg−1). When the pedal with-drawal reflex was eliminated and/or the respiratory rate was

809 diminished, animals were trans-cardially perfused with ice-cold saline. The brain was immediately
810 removed, embedded in Optimal Cutting Temperature (OCT) freezing medium, and flash frozen in an
811 isopropyl ethanol-dry ice bath. Samples were cut on a cryostat (Leica CM 1850) at a thickness of 16μm,
812 adhered to SuperFrost Plus microscope slides (VWR, 48311-703), and stored at -80 °C until use. Portions
813 of the brain that were not cut were recoated in OCT and stored again for future use. Samples were
814 immediately fixed in 4% paraformaldehyde and stained on the slide according to the Molecular
815 Instruments HCR generic sample in solution RNA-FISH protocol (Molecular Instruments,
816 https://files.molecularinstruments.com/MI-Protocol-RNAFISH-GenericSolution-Rev7.pdf) or the
817 Advanced Cell Diagnostics RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (ACD, 323100,
818 https://acdbio.com/sites/default/files/USM-
819 323100%20Multiplex%20Fluorescent%20v2%20User%20Manual_10282019_0.pdf) protocol (**Table
820 S7**).
821
822 **Single nucleus RNA-seq, library preparation, sequencing**
823 *10x RNA-seq.* Unsorted single-nucleus suspensions from frozen marmoset and mouse samples were
824 generated as in (*87*). GEM generation and library preparation followed the manufacturer's protocol (10X
825 Chromium single-cell 3′ v.3, protocol version #CG000183_ ChromiumSingleCell3′_v3_UG_Rev-A).
826 Raw sequencing reads were aligned to the NCBI CJ1700 reference (marmoset) or GRCm38 (mouse).
827 Reads that mapped to exons or introns were assigned to annotated genes. Libraries were sequenced to a
828 median read depth of 8 reads per Unique Molecular Identifier (UMI, or transcript), obtaining a median
829 7,262 UMIs per cell.
830
831 **RNA sequencing data processing, curation, clustering**
832 Processing and alignment steps follow those outlined in: https://github.com/broadinstitute/Drop-seq/
833 (**Table S7**). Raw BCL files were processed using IlluminaBasecallsToSam, and reads with a barcode
834 quality score below Q10 were discarded. Cell barcodes (CBCs) were filtered using the 10X CBC
835 whitelist, followed by TSO and polyA trimming. Reads were aligned using STAR, and tagged with their
836 gene mapping (exonic, intronic, or UTR) and function (strand). The reads were then processed through
837 GATK BQSR, and tabulated in a digital gene expression matrix (DGE) containing all CBCs with at least
838 20 transcripts aligning to coding, UTR, or intronic regions. Cell selection was performed based on
839 CellBender remove-background non-empties (*88*), % intronic (% of a CBC's reads that are intronic), and
840 number of UMIs for a CBC. A new filtered DGE containing only these selected CBCs was then
841 generated. Finally, a gene-metagene DGE was created by merging the selected CBCs DGE with a
842 metagene DGE (made by identifying reads from the selected CBCs that have a secondary alignment
843 mapping to a different gene than its primary alignment).
844
845 Cell type classification models were trained using our annotations and scPred R package version 1.9.2
846 (*21*). Detection of cell-cell doublets was performed using a two-step process based on the R package
847 DoubletFinder (*89*). DoubletFinder implements a nearest-neighbors approach to doublet detection. First,
848 artificial doublets are simulated from input single-cell data and are co-clustered with true libraries. True
849 doublet libraries are identified by their relative fraction of artificial doublet nearest-neighbors in gene
850 expression space. In our workflow this process is run twice, once with high stringency to identify and
851 remove clear doublets, and again with a lower threshold to identify remaining, subtler doublet libraries.
852 We used the following parameters for round 1: PN = 0.4, PK = 10 ^ seq(-4, -1, length.out=50),

853   NUM_PCS = 5. For the second round: PN = 0.45, NUM_PCS = 10, PK = 10 ^ seq(-4, -1, length.out=50).
854   Because doublet libraries are initially categorized as true libraries in the nearest-neighbor search, we find
855   this two-step process improves the sensitivity and accuracy of doublet detection.
856
857   Clustering was performed using independent component analysis (ICA; R package fastICA)
858   dimensionality reduction followed by Louvain. Cells assigned to one of the major glial types
859   (oligodendrocyte lineage, astrocytes, vascular/endothelia, microglia/macrophages) by scPred were
860   collected across all brain regions and clustered together. Neurons from most telencephalic structures
861   (neocortex, hippocampus, striatum, amygdala) confidently assigned to the categories "GABAergic" and
862   "glutamatergic" and so were clustered separately by neurotransmitter usage for each brain structure.
863   Striatal neuron categories were "SPN" and "GABAergic interneuron". Neurons in non-telencephalic brain
864   structures were clustered separately by brain structure. All clusterings were performed in two stages: first-
865   round clustering was based on the top 60 independent components (ICs) and a default resolution (res) of
866   0.1, nearest neighbors (nn) = 25. Following the process outlined in(6), each was manually reviewed for
867   skew and kurtosis of gene loadings on factors and cells to identify ICs that loaded on outliers, doublets, or
868   artifactual signals. These were discarded and reclustering was performed on the remaining ICs. Each
869   resulting cluster was then subjected to second-round clustering, during which ICs were again curated.
870   Second-round clustering explored a range of parameters: nn=10,20,30;
871   res=0.01,0.05,0.1,0.2,0.3,0.4,0.5,1.0. Final parameter values were chosen to optimize concordance, when
872   possible, between the final number of clusters and the number of included ICs, such that each cluster was
873   defined by one primary IC. Metacells for each cluster are generated by summing transcript counts for
874   each cell across all cells in the cluster, normalizing by total number of transcripts, and scaling to counts
875   per 100,000.
876

877   **Identification of regionally differentially expressed genes (rDEGs)**

878   We computed rDEGs for neocortical gluatmatergic neurons, GABAergic neurons, astrocytes, and
879   oligodendrocyte lineage types. We used individual cell level cluster assignments from the initial
880   neocortical clustering (which contained all regions together) to create per-region metacells for each
881   cluster. Normalized metacells were log10 transformed and pairwise differences across regions within the
882   same cluster (cell type) were examined. Genes with > 3 fold difference in the same cluster between two
883   regions were considered rDEGs. For this analysis we omitted from comparisons any region-cluster
884   metacell generated from fewer than 50 cells, but retain comparisons between regions that had > 50 cells
885   in that cluster. Genes that were consistently rDEGs in at least 3 individuals for each cluster-region pair are
886   reported in **Table S3**.
887

888   **Ancestral State Reconstruction**
889   Ancestral state reconstruction (ASR) is a method to infer hidden ancestral traits from extant observations.
890   For example, given a phylogenetic tree of species and genomic sequences thought to be homologous
891   across those species, the reconstruction takes into account branch lengths to reconstruct the most likely
892   ancestral sequence. We applied a maximum likelihood-based ASR approach (R package: fastAnc) to the
893   scaled, normalized metacells of cell types and the dendrogram of their similarity to produce estimates of
894   expression of each gene at the internal nodes of the tree. This enabled comparisons of leaf nodes to
895   internal nodes as well as internal nodes to each other. For example, compared to the parent node of

896 amygdala, basal forebrain, and hypothalamic SPN-like GABAergic projection neurons, the reconstructed
897 parent node of striatal SPNs had higher expression of known markers of striatal projection neurons such
898 as *DACH1*. We used these reconstructions of internal node gene expression to compare major clades of
899 the tree, and used a Chi-square test to determine whether transcription factors were overrepresented
900 among differentially expressed genes (threshold: 3-fold difference) between pairs of internal nodes.
901
902 **Spatial smFISH experiments**
903 All probes are listed in **Table S4**. All smFISH validation experiments were carried out on distinct
904 biological replicates from those used for snRNA-seq or single-cell ATAC-seq experiments.
905
906 *smFISH tissue processing and quantification.* Two marmosets (Cj 18-134 and 19-212) were euthanized
907 and perfused with PBS (**Table S1**). The brain was removed, embedded rapidly in OCT, and stored in the -
908 80C freezer. Tissue was then cut to 16µm on a cryostat and stored in the -80C until needed. *In situ*
909 hybridization was performed for genes of interest (see **Table S4**) with HCR or ACD antisense probes,
910 incubated with TrueBlack Lipofuscin Autofluorescence Quencher (Biotium, 23007) for 10 seconds at
911 room temperature to eliminate confounding lipofuscin autofluorescence present in the tissue. Samples
912 were then coverslipped with ProLong Diamond Antifade Mountant (Invitrogen, P36970). Z-stack serial
913 images were taken through the whole depth across striatum, hypothalamus and basal forebrain regions,
914 and several regions of neocortex, on the TissueGnostic TissueFAXS SL slide-scanning, spinning disk
915 confocal microscope (Hamamatsu Orca Flash 4.0 v3) using a 20×/0.8 NA air objective for ACD stains or
916 a 40×/1.2 NA water-immersion objective for HCR stains.
917
918 Series images were segmented using StrataQuest, a software package from TissueGnostics, which enables
919 the quantification of signals within segmented images (similar to CellProfiler). Nuclei objects were
920 generated using the DAPI channel, and artifacts were removed based on size and intensity. Exclusion
921 ROIs were manually drawn to avoid areas of white matter, large artifacts, and autofluorescence before
922 computing intensity and other statistical and morphological measurements (20 parameters) for each
923 channel. Specifically, 50 cells were hand-labeled as positive or negative for the markers of interest in
924 order to identify the appropriate threshold for feature selection using the parameters that best
925 discriminated this binary. Parameters included mean intensity, maximum intensity, standard deviation of
926 intensity, and range of intensity, and equivalent diameter. These filters were then applied to the unlabeled
927 data to identify positive cells.
928
929 Segmented cells were further analyzed using in-house code (https://github.com/klevando/BICCN-
930 StrataQuest-Script, **Table S7**). Spatial locations of the cells were visualized by plotting the x-y
931 coordinates associated with each nuclei. These were then binned into 2-D histograms across the x and y
932 axis (corresponding to the rostrocaudal plane and to the dorsoventral plane respectively). A size of 100
933 bins was chosen for the first (medial-most) slice of a series across the x axis and the calculated bin size
934 was then used across the y axis of the first slice and across other slices in the series. Positive events for a
935 gene in a given bin were either normalized to the number of detected nuclei in that bin (DAPI) and plotted
936 in 2-D as a relative heatmap, or simply plotted as a density heatmap without being normalized to DAPI.
937 Whole slice normalizations across the mediolateral axis (all positive events for a gene in a given slice
938 relative to DAPI) were plotted as bar graphs. DAPI and marker of interest counts are available in **Data**
939 **S1-S2**.

940

Neocortical Areal Proportions

In addition to the mediolateral subdivisions of marmoset neocortex, we further parcellated the neocortex into subregions referencing the Brain/MINDS 3D Digital Marmoset Atlas (https://doi.org/10.24475/bma.4520). Slice selection was based upon visual recognition of prominent landmarks (white matter, striatal boundary as well as DAPI nuclei staining) within our tissue and matched with the nearest atlas slice. ROIs for each parcellated region were created using the StrataQuest software in the anterior-posterior and dorsal-ventral axes (**Fig. S8C**). The corresponding feature selection that was previously set within the full neocortical sections was carried across gene and slice. The parameters used were mean intensity, maximum intensity, standard deviation of intensity, and range of intensity, and equivalent diameter. These filters were then applied to the unlabeled data to identify positive cells within each individual ROI. The parcellated neocortical areas were then analyzed using in-house code (https://github.com/klevando/BICCN-StrataQuest-Script, **Table S7**). As above, positive events for a gene in a given ROI were normalized to DAPI and plotted in a (stacked) bar graph. Relative percentages between genes were calculated and plotted stacked.

Morphology and smFISH experiments

AAV9-hDlx5/6-GFP-fGFP virus was generated as in (*19*). 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 AAV9-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 PBS-Azide and stored at 4C or moved into 70% ethanol for storage at -20C. 70% ethanol storage prevents RNA degradation at this temperature without significant tissue shrinkage for short storage times. Due to lab shutdowns during the pandemic, sections from two marmosets were stored in 70% ethanol for approximately 4 months. These samples exhibited significant shrinkage, measured by DAPI-stained nuclei diameters (**Table S1**), therefore we only compared morphology parameters within-donor. Sections were taken as needed and *in situ* hybridization was performed with HCR antisense probes, following the generic sample in solution HCR protocol with a 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 S4**) that corresponded with RNA-seq defined clusters. The sections were then stained with anti-GFP antibody (**Table S4**) and a secondary antibody (**Table S4**) to amplify the endogenous GFP signal (https://www.protocols.io/view/marmoset-nhp-free-floating-anti-gfp-antibody-stain-3byl47nb2lo5/v1, **Table S4**). 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. Selected sections were imaged on an upright confocal laser scanning microscope

983   (Olympus Fluoview FV3000) using a 40x/0.95 NA air objective or a 60x/1.50 NA oil-immersion
984   objective and cooled GaAsP PMTs.

985   *Morphological reconstruction and feature quantification - Imaris.* Without pre-processing the confocal
986   images, three-dimensional (3D) reconstruction and surface rendering of striatal and neocortical
987   interneurons were performed using Imarisx64 9.9 software (Oxford Instruments) based on GFP+ signal.
988   Surface-rendered images were used to determine the soma diameter, total volume, and total surface area
989   for each z-stack image. 3D-skeleton diagrams (**Figs. 5A-C** and **6H**), corresponding to each surface-
990   rendered image (data not shown), were generated using the Filament Tracing wizard in Imaris and then
991   pseudo-colored in Adobe Illustrator. The total number of primary dendritic branches, dendritic branch
992   points, area, volume, and length of the 3D-skeleton diagrams were calculated using the AutoPath (no
993   loops) algorithm in the filament tracing wizard in Imaris. The total number of primary dendritic branches
994   for each cell is defined by the number of dendrite branches in the filament trace 1 distance value away
995   from the soma. The distance value is calculated automatically by the AutoPath (no loops) algorithm based
996   on the diameter of the soma and the diameter of the thinnest cellular process. All data were exported to
997   CSV, and data collected from exemplar cells (**Fig. 5A-B** and **6G)** were reported in **Table S6**.

998   A separate dataset containing GFP+ interneurons from one animal (Cj 17-154) was first blinded using a
999   custom Python script, then reconstructed with Imaris (**Figs. 5C-D**). All results are presented as mean ±
1000   SEM. Comparisons between soma diameter, surface area, surface volume, and the number of primary
1001   branches were carried out using unpaired *t*-tests. Unpaired Mann-Whitney *U* tests were used to
1002   statistically compare the total length, area, volume, and total number of dendritic branch points
1003   measurements. Statistically significant analyses were denoted as follows: *$*p < 0.05$; $**p < 0.01$.

1004

1005   *Morphological reconstruction and feature quantification - NeuTube.* Automatic or semi-automatic tracing
1006   algorithms are challenged by some data, perhaps due to the low SNR of a given image. To overcome this,
1007   we manually reconstructed the sparse neurons via Neutube (*90*) tracing software. With the software, we
1008   1) Create 3D volume rendering of the GFP-AAV marmoset neuron, 2) use the signal transfer function
1009   (e.g., histogram equalization) for overall intensity and opacity values to optimize the signal-to-noise ratio
1010   by manually examining the clearest visualization of the dendrites, 3) Build the neuron skeleton over the
1011   3D volume by tracing the processes, 4) Scan through the 3D volume to make sure no parts of the neuron
1012   are missed, 5) Double check the raw 2D images to see if any of the branches were not presented well in
1013   3D due to volume rendering artifacts, 6) Label axon, dendrites and soma parts of the skeleton model.
1014   Reconstructed neurons were saved as SWC format.

1015

1016   **Single nucleus ATAC-seq, library preparation, sequencing**
1017   Single nucleus ATAC-seq (snATAC-seq; 10x Genomics Single Cell ATAC v1) was conducted on fresh
1018   marmoset tissue (1 female, 1.5 y.o.) dissected from anterior striatum. Nuclei suspensions were generated
1019   following the 10x recommended protocol (CG000212 Rev B). Library preparation followed 10x
1020   Genomics Single Cell ATAC v1 Guide (CG000168 Rev D). Library was sequenced on an Illumina
1021   NovaSeq (RRID:SCR_016387) using 100 bp paired-end reads to a median per-cell fragment depth of
1022   25,472. Alignment and fragment counting was conducted using Cell Ranger (RRID:SCR_01734), aligned
1023   to marmoset genome cj1700 (https://www.ncbi.nlm.nih.gov/assembly/GCF_009663435.1/;
1024   GCA_009663445.2). snATAC-seq data were integrated with marmoset striatal snRNA-seq data from two
1025   independent animasl (bi005, bi006) using Signac (RRID:SCR_021158) with the following parameters:
1026   integration method: cca; weight.reduction = lsi. Differentially accessible peaks for each cluster were

1027 calculated with a 3-fold change cutoff and the minimum fraction of expressed cells (in target cluster) =
1028 0.2. Enhancer locations were as follows (CJ1700 coordinates): Eh14 = chr1-122696345-122696560 (215
1029 bp); Eh15 = chr17-32919427-32919564 (137 bp); Eh16 = chr20-38770509-38770619 (110 bp); Eh17 =
1030 chr9-60258694-60258788 (94 bp).
1031
1032 **TAC3-AAV viral design and production.**
1033 AAVs were produced in accordance with the protocol (*91*). 5.7 μg/150 mm dish of construct DNA was
1034 transfected with 22.8 μg/150 mm dish of pAAV9 capsid plasmid and 11.45 μg/150 mm dish of
1035 pAdDeltaF6 helper plasmid using polyethylenimine 25K MW (Polysciences, 23966-1). Collection of
1036 cells and media for AAV harvesting began 72 hours following transfection before iodixanol gradient
1037 ultracentrifugation purification using a Type 70 Ti Fixed-Angle Titanium Rotor (Beckman-Coulter,
1038 337922). Titer was calculated using digital droplet PCR (ddPCR) according to the protocol "ddPCR
1039 Titration of AAV Vectors" (https://www.addgene.org/protocols/aav-ddpcr-titration/) using the QX200
1040 AutoDG Droplet Digital PCR System (BioRad, 1864100). gBlock fragments containing the tandem *TAC3*
1041 enhancers (E14E15E16E17) were synthesized by Integrated DNA Technologies (IDT). The enhancers
1042 were cloned into AAV plasmid vector backbones containing a minimal CMV promoter, the reporter and a
1043 barcode unique to each enhancer sequence using the NEBuilder® HiFi DNA Assembly Cloning Kit
1044 (NEB-E5520S), following standard protocol. 2 uL of Gibson assembly product was used to transform 50
1045 uL of home-made Stbl3 cells following a standard transformation protocol. Mini-preparation of plasmids
1046 was carried out using the NucleoSpin mini kit (Macherey-Nagel). Positive clones were identified by
1047 restriction enzyme digestion and sequencing. The positive clones were grown in 300 mL LB cultures and
1048 the plasmids were extracted using the NucleoBond Xtra Midi EF kit (Macherey-Nagel). The plasmids
1049 were sent out for sequencing again before being packaged into AAVs.

1050 **TAC3-AAV local injection procedure**

1051 *Structural MRI Scanning*. In preparation for MR imaging under anesthesia, an animal (2 marmosets, Cj
1052 19-207 and 17-B111, **Table S1**) checked for robust health was fasted overnight with *ad libitum* access to
1053 water. Prior to scanning, the animal was sedated with Alphaxalone (5-10 mg/kg) or a combination of
1054 Alphaxalone (4-8 mg/kg) and Ketamine (5-10 mg/Kg) given intramuscularly. It was then transferred to a
1055 custom designed MRI cradle equipped with a 3D printed nose cone for delivering anesthetic gases. The
1056 animal's head was securely held to minimize motion artifacts and stereotaxically align the rostral-caudal
1057 axis with the MR scanner bore using contrast filled MR compatible ear bars, an adjustable palette bar and
1058 eye bars. During scanning, anesthesia was maintained at 1-2% isoflurane in a mixture of oxygen. A
1059 temperature controlled water circulating heat pad was used for thermal support. Heart rate and blood
1060 oxygen saturation levels were continuously monitored and recorded every 10 min with an MR compatible
1061 monitoring system (Nonin, MN). MR scans consisted of high-resolution 3D T1 mapping using a
1062 Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) sequence (Liu et al., 2011) with repetition
1063 time (TR) = 6000 ms, TE (echo time) = 4.6 ms; flip angle (FA) =12 degrees; FOV = 80, matrix size =
1064 256x256. A high-resolution T2-weighted anatomical scan was also obtained using a RARE pulse
1065 sequence with effective TE = 35.5 ms, TR = 2500 ms, RARE factor = 8, spatial resolution = 156 μm ×
1066 156 μm × 0.5 mm, and matrix size = 256 × 256, RARE factor = 4, 80 mm FOV.
1067
1068 *Virus injection surgery*. Following sedation with Alphaxalone (5-10 mg/kg), the animal was intubated and
1069 maintained at 1.5-2% isoflurane in a mixture of oxygen throughout the rest of the procedure. An

1070   intravenous (IV) catheter was placed in the saphenous vein for infusion of fluids and drugs during the
1071   surgery and the recovery period. Heart rate, SpO2, ECG, end-tidal CO_2 and rectal temperature were
1072   continuously monitored and logged every 5 minutes throughout. Once the animal acquired a stable plane
1073   of anesthesia, it was placed in a stereotaxic apparatus (Narishige, Japan). A thin layer of sterile eye
1074   lubricant was applied to protect against corneal drying. Animal was also provided with a single bolus
1075   dose of intravenous dexamethasone (0.4 mg/kg) to guard against brain swelling. The scalp and fascia
1076   were removed in layers via blunt dissection to expose the injection sites. Using MRI atlas guidance,
1077   bilateral craniotomies were performed to expose the medial anterior caudate and ventral posterior
1078   putamen region. The tandem enhancer virus AAV9-tandemE-TAC3-EGFP was injected in left Caudate
1079   (high titer, 10^13 low volume, 0.5 ul) and left Putamen (low titer, 10^11, high volume 1.5 ul). Each
1080   injection was delivered with a controlled syringe pump (100 nL/minute). After each injection the needle
1081   was kept in place for 10 minutes then slowly retracted over 1-2 minutes to minimize injection backflow.
1082   Following withdrawal, the cranial openings were covered with bonewax, fascia and scalp were sutured in
1083   layer wise manner with 3-0 vicryl sutures. The suture site was cleaned with warm sterile saline and
1084   covered with hypafix. For recovery the animal was transferred in an incubator and monitored closely till it
1085   was able to move effortlessly and accepted treats before returning to its home cage. Post-op medications
1086   were provided under veterinary advice for pain and inflammation control until full recovery. The animals
1087   were euthanized 6-10 weeks after viral injection and were perfused transcardially with 4% PFA. The
1088   brain was extracted for histology, smFISH, and confocal microscopy.
1089

1090   **TAC3-AAV-BI103 systemic injection procedure**

1091   One 2-year old female marmoset (**Table S1**) was sedated with Alphaxalone (5-10 mg/kg) given
1092   intramuscularly. A catheter was placed into the tail vein and 200 μl of purified viral particles were
1093   injected at a dose of 7.85×10^{13} vg/kg, followed by injection of 1.5 ml saline to flush the line. The animal
1094   was euthanized 4 weeks after viral injection and was perfused transcardially with 4% PFA. The brain was
1095   extracted for histology, smFISH, and confocal microscopy.

1096   **TAC3-AAV histology**

1097   For AAV9-tandemE-TAC3-EGFP experiments, coronal sections around the injection sites were cut at
1098   100 μm on a vibratome (Leica, VT1000S) and then stored in PBS-NaN3 at 4 °C until use. For AAV-
1099   BI103-tandemE-TAC3-EGFP experiments, coronal and sagittal sections were processed. Sections were
1100   stained for *TAC3* (Molecular Instruments, PRC843, **Table S4**) using the HCR v.3.0 protocol (Molecular
1101   Instruments) described above and a rabbit anti-GFP antibody (Invitrogen, A11122, **Table S4**) with
1102   subsequent secondary goat anti-rabbit conjugated with AF488 (ThermoFIsher, A-11008). Stained sections
1103   were then incubated 1-3 minutes in TrueBlack to eliminate confounding lipofuscin autofluorescence. The
1104   striatum was imaged for each section using the TissueFAXS SL with a 40×/1.2 NA water-immersion lens
1105   through the whole thickness of the tissue.
1106

1107   **References**

1108   1.   S. Ma, M. Skarica, Q. Li, C. Xu, R. D. Risgaard, A. T. N. Tebbenkamp, X. Mato-Blanco, R. Kovner,

1109   Ž. Krsnik, X. de Martin, V. Luria, X. Martí-Pérez, D. Liang, A. Karger, D. K. Schmidt, Z. Gomez-
1110   Sanchez, C. Qi, K. T. Gobeske, S. Pochareddy, A. Debnath, C. J. Hottman, J. Spurrier, L. Teo, A. G.
1111   Boghdadi, J. Homman-Ludiye, J. J. Ely, E. W. Daadi, D. Mi, M. Daadi, O. Marín, P. R. Hof, M.-R.
1112   Rasin, J. Bourne, C. C. Sherwood, G. Santpere, M. J. Girgenti, S. M. Strittmatter, A. M. M. Sousa,
1113   N. Sestan, Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science*,
1114   eabo7257 (2022).

1115   2.   T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F.
1116   M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D.
1117   Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow,
1118   D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R.
1119   Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D.
1120   Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu,
1121   J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed,
1122   C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc,
1123   M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M.
1124   Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz,
1125   C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte,
1126   S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M.
1127   Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R.
1128   Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset
1129   and mouse. *Nature*. **598**, 111–119 (2021).

1130   3.   F. M. Krienen, M. Goldman, Q. Zhang, R. C. H. del Rosario, M. Florio, R. Machold, A. Saunders,
1131   K. Levandowski, H. Zaniewski, B. Schuman, C. Wu, A. Lutservitz, C. D. Mullally, N. Reed, E.
1132   Bien, L. Bortolin, M. Fernandez-Otero, J. D. Lin, A. Wysoker, J. Nemesh, D. Kulp, M. Burns, V.
1133   Tkachev, R. Smith, C. A. Walsh, J. Dimidschstein, B. Rudy, L. S. Kean, S. Berretta, G. Fishell, G.
1134   Feng, S. A. McCarroll, Innovations present in the primate interneuron repertoire. *Nature*. **586**, 262–
1135   269 (2020).

1136   4.   J.-P. Lin, H. M. Kelly, Y. Song, R. Kawaguchi, D. H. Geschwind, S. Jacobson, D. S. Reich,
1137   Microenvironment Impacts the Molecular Architecture and Interactivity of Resident Cells in
1138   Marmoset Brain. *bioRxiv* (2021), p. 2021.01.25.426385.

1139   5.   R. C. Bandler, I. Vitali, R. N. Delgado, M. C. Ho, E. Dvoretskova, J. S. Ibarra Molinas, P. W. Frazel,
1140   M. Mohammadkhani, R. Machold, S. Maedler, S. A. Liddelow, T. J. Nowakowski, G. Fishell, C.
1141   Mayer, Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature*. **601**, 404–
1142   409 (2022).

1143   6.   A. Saunders, E. Z. Macosko, A. Wysoker, M. Goldman, F. M. Krienen, H. de Rivera, E. Bien, M.
1144   Baum, L. Bortolin, S. Wang, A. Goeva, J. Nemesh, N. Kamitaki, S. Brumbaugh, D. Kulp, S. A.
1145   McCarroll, Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell*.
1146   **174**, 1015–1030.e16 (2018).

1147   7.   A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E.
1148   Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J.
1149   Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, Molecular Architecture of the
1150   Mouse Nervous System. *Cell*. **174**, 999–1014.e22 (2018).

1151   8.   Z. Yao, C. T. J. van Velthoven, M. Kunst, M. Zhang, D. McMillen, C. Lee, W. Jung, J. Goldy, A.
1152   Abdelhak, P. Baker, E. Barkan, D. Bertagnolli, J. Campos, D. Carey, T. Casper, A. B. Chakka, R.
1153   Chakrabarty, S. Chavan, M. Chen, M. Clark, J. Close, K. Crichton, S. Daniel, T. Dolbeare, L.

Ellingwood, J. Gee, A. Glandon, J. Gloe, J. Gould, J. Gray, N. Guilford, J. Guzman, D. Hirschstein, W. Ho, K. Jin, M. Kroll, K. Lathia, A. Leon, B. Long, Z. Maltzer, N. Martin, R. McCue, E. Meyerdierks, T. N. Nguyen, T. Pham, C. Rimorin, A. Ruiz, N. Shapovalova, C. Slaughterbeck, J. Sulc, M. Tieu, A. Torkelson, H. Tung, N. V. Cuevas, K. Wadhwani, K. Ward, B. Levi, C. Farrell, C. L. Thompson, S. Mufti, C. M. Pagan, L. Kruse, N. Dee, S. M. Sunkin, L. Esposito, M. J. Hawrylycz, J. Waters, L. Ng, K. A. Smith, B. Tasic, X. Zhuang, H. Zeng, A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *bioRxiv* (2023), p. 2023.03.06.531121.

9. J. Langlieb, N. Sachdev, K. Balderrama, N. Nadaf, M. Raj, E. Murray, J. Webber, C. Vanderburg, V. Gazestani, D. Tward, C. Mezias, X. Li, D. Cable, T. Norton, P. P. Mitra, F. Chen, E. Macosko, The cell type composition of the adult mouse brain revealed by single cell and spatial genomics. *bioRxiv* (2023), doi:10.1101/2023.03.06.531307.

10. K. L. Chiou, X. Huang, M. O. Bohlen, S. Tremblay, D. R. O'Day, C. H. Spurrell, A. A. Gogate, T. M. Zintel, Cayo Biobank Research Unit, M. G. Andrews, M. I. Martinez, L. M. Starita, M. J. Montague, M. L. Platt, J. Shendure, N. Snyder-Mackler, A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *bioRxiv* (2022), p. 2022.09.30.510346.

11. K. Siletti, R. Hodge, A. M. Albiach, L. Hu, K. W. Lee, P. Lönnerberg, T. Bakken, S.-L. Ding, M. Clark, T. Casper, N. Dee, J. Gloe, C. Dirk Keene, J. Nyhus, H. Tung, A. M. Yanny, E. Arenas, E. S. Lein, S. Linnarsson, Transcriptomic diversity of cell types across the adult human brain. *bioRxiv* (2022), p. 2022.10.12.511898.

12. J.-P. Lin, H. M. Kelly, Y. Song, R. Kawaguchi, D. H. Geschwind, S. Jacobson, D. S. Reich, Transcriptomic architecture of nuclei in the marmoset CNS. *Nat. Commun.* **13**, 5531 (2022).

13. 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).

14. S. Sahara, Y. Yanagawa, D. D. M. O'Leary, C. F. Stevens, The Fraction of Cortical GABAergic Neurons Is Constant from Near the Start of Cortical Neurogenesis to Adulthood. *Journal of Neuroscience*. **32**, 4755–4761 (2012).

15. 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).

16. J. Hill, T. Inder, J. Neil, D. Dierker, J. Harwell, D. Van Essen, Similar patterns of cortical expansion during human development and evolution. *Proceedings of the National Academy of Sciences*. **107**, 13135–13140 (2010).

17. D. Goertsen, N. C. Flytzanis, N. Goeden, M. R. Chuapoco, A. Cummins, Y. Chen, Y. Fan, Q. Zhang, J. Sharma, Y. Duan, L. Wang, G. Feng, Y. Chen, N. Y. Ip, J. Pickel, V. Gradinaru, AAV capsid

1197   variants with brain-wide transgene expression and decreased liver targeting after intravenous
1198   delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).

1199 18. J. K. Mich, L. T. Graybuck, E. E. Hess, J. T. Mahoney, Y. Kojima, Y. Ding, S. Somasundaram, J. A.
1200   Miller, B. E. Kalmbach, C. Radaelli, B. B. Gore, N. Weed, V. Omstead, Y. Bishaw, N. V.
1201   Shapovalova, R. A. Martinez, O. Fong, S. Yao, M. Mortrud, P. Chong, L. Loftus, D. Bertagnolli, J.
1202   Goldy, T. Casper, N. Dee, X. Opitz-Araya, A. Cetin, K. A. Smith, R. P. Gwinn, C. Cobbs, A. L. Ko,
1203   J. G. Ojemann, C. D. Keene, D. L. Silbergeld, S. M. Sunkin, V. Gradinaru, G. D. Horwitz, H. Zeng,
1204   B. Tasic, E. S. Lein, J. T. Ting, B. P. Levi, Functional enhancer elements drive subclass- selective
1205   expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).

1206 19. J. Dimidschstein, Q. Chen, R. Tremblay, S. L. Rogers, G.-A. Saldi, L. Guo, Q. Xu, R. Liu, C. Lu, J.
1207   Chu, J. S. Grimley, A.-R. Krostag, A. Kaykas, M. C. Avery, M. S. Rashid, M. Baek, A. L. Jacob, G.
1208   B. Smith, D. E. Wilson, G. Kosche, I. Kruglikov, T. Rusielewicz, V. C. Kotak, T. M. Mowery, S. A.
1209   Anderson, E. M. Callaway, J. S. Dasen, D. Fitzpatrick, V. Fossati, M. A. Long, S. Noggle, J. H.
1210   Reynolds, D. H. Sanes, B. Rudy, G. Feng, G. Fishell, A viral strategy for targeting and manipulating
1211   interneurons across vertebrate species. *Nat. Neurosci.* **19**, 1743–1749 (2016).

1212 20. D. Vormstein-Schneider, J. D. Lin, K. A. Pelkey, R. Chittajallu, B. Guo, M. A. Arias-Garcia, K.
1213   Allaway, S. Sakopoulos, G. Schneider, O. Stevenson, J. Vergara, J. Sharma, Q. Zhang, T. P.
1214   Franken, J. Smith, L. A. Ibrahim, K. J. M. Astro, E. Sabri, S. Huang, E. Favuzzi, T. Burbridge, Q.
1215   Xu, L. Guo, I. Vogel, V. Sanchez, G. A. Saldi, B. L. Gorissen, X. Yuan, K. A. Zaghloul, O.
1216   Devinsky, B. L. Sabatini, R. Batista-Brito, J. Reynolds, G. Feng, Z. Fu, C. J. McBain, G. Fishell, J.
1217   Dimidschstein, Viral manipulation of functionally distinct interneurons in mice, non-human primates
1218   and humans. *Nat. Neurosci.* **23**, 1–21 (2020).

1219 21. J. Alquicira-Hernandez, A. Sathe, H. P. Ji, Q. Nguyen, J. E. Powell, scPred: accurate supervised
1220   method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).

1221 22. H. Zeng, J. R. Sanes, Neuronal cell-type classification: challenges, opportunities and the path
1222   forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).

1223 23. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G.
1224   Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nat.
1225   Rev. Genet.* **17**, 744–757 (2016).

1226 24. H. Zeng, What is a cell type and how to define it? *Cell* (2022) (available at
1227   https://www.sciencedirect.com/science/article/pii/S0092867422007838).

1228 25. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things).
1229   *Methods in Ecology and Evolution*. **3**, 217–223 (2012).

1230 26. J. Woych, A. Ortega Gurrola, A. Deryckere, E. C. B. Jaeger, E. Gumnit, G. Merello, J. Gu, A. Joven
1231   Araus, N. D. Leigh, M. Yun, A. Simon, M. A. Tosches, Cell-type profiling in salamanders identifies
1232   innovations in vertebrate forebrain evolution. *Science*. **377**, eabp9186 (2022).

1233 27. G. Meyer, Building a human cortex: the evolutionary differentiation of Cajal-Retzius cells and the
1234   cortical hem. *J. Anat.* **217**, 334–343 (2010).

1235 28. P. Jager, G. Moore, P. Calpin, X. Durmishi, I. Salgarella, L. Menage, Y. Kita, Y. Wang, D. W. Kim,
1236   S. Blackshaw, S. R. Schultz, S. Brickley, T. Shimogori, A. Delogu, Dual midbrain and forebrain
1237   origins of thalamic inhibitory interneurons. *Elife*. **10** (2021), doi:10.7554/eLife.59272.

1238 29. K. Letinic, P. Rakic, Telencephalic origin of human thalamic GABAergic neurons. *Nat. Neurosci.* **4**,
1239   931–936 (2001).

1240 30. L. W. Swanson, G. D. Petrovich, What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).

1241 31. T. Hirata, P. Li, G. M. Lanuza, L. A. Cocas, M. M. Huntsman, J. G. Corbin, Identification of distinct
1242   telencephalic progenitor pools for neuronal diversity in the amygdala. *Nat. Neurosci.* **12**, 141–149
1243   (2009).

1244 32. T. Kaoru, F. C. Liu, M. Ishida, T. Oishi, M. Hayashi, M. Kitagawa, K. Shimoda, H. Takahashi,
1245   Molecular characterization of the intercalated cell masses of the amygdala: implications for the
1246   relationship with the striatum. *NSC*. **166**, 220–230 (2010).

1247 33. M. T. Schmitz, K. Sandoval, C. P. Chen, M. A. Mostajo-Radji, W. W. Seeley, T. J. Nowakowski, C.
1248   J. Ye, M. F. Paredes, A. A. Pollen, The development and evolution of inhibitory neurons in primate
1249   cerebrum. *Nature*. **603**, 871–877 (2022).

1250 34. D. L. Kaufman, C. R. Houser, A. J. Tobin, Two forms of the gamma-aminobutyric acid synthetic
1251   enzyme glutamate decarboxylase have distinct intraneuronal distributions and cofactor interactions.
1252   *J. Neurochem.* **56**, 720–723 (1991).

1253 35. J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D.
1254   Rubinstein, J. Hao, A. Regev, C. Dulac, X. Zhuang, Molecular, spatial, and functional single-cell
1255   profiling of the hypothalamic preoptic region. *Science*. **362**, eaau5324–80 (2018).

1256 36. B. Tasic, Z. Yao, L. T. Graybuck, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren,
1257   M. N. Economo, S. Viswanathan, O. Penn, T. Bakken, V. Menon, J. Miller, O. Fong, K. E.
1258   Hirokawa, K. Lathia, C. Rimorin, M. Tieu, R. Larsen, T. Casper, E. Barkan, M. Kroll, S. Parry, N.
1259   V. Shapovalova, D. Hirschstein, J. Pendergraft, H. A. Sullivan, T. K. Kim, A. Szafer, N. Dee, P.
1260   Groblewski, I. Wickersham, A. Cetin, J. A. Harris, B. P. Levi, S. M. Sunkin, L. Madisen, T. L.
1261   Daigle, L. Looger, A. Bernard, J. Phillips, E. Lein, M. Hawrylycz, K. Svoboda, A. R. Jones, C.
1262   Koch, H. Zeng, Shared and distinct transcriptomic cell types across neocortical areas. *Nature*. **563**,
1263   1–41 (2018).

1264 37. P. Balaram, J. H. Kaas, Towards a unified scheme of cortical lamination for primary visual cortex
1265   across primates: insights from NeuN and VGLUT2 immunoreactivity. *Front. Neuroanat.* **8**, 81
1266   (2014).

1267 38. O. A. Bayraktar, L. C. Fuentealba, A. Alvarez-Buylla, D. H. Rowitch, Astrocyte development and
1268   heterogeneity. *Cold Spring Harb. Perspect. Biol.* **7**, a020362 (2014).

1269 39. O. A. Bayraktar, T. Bartels, S. Holmqvist, V. Kleshchevnikov, A. Martirosyan, D. Polioudakis, L.
1270   Ben Haim, A. M. H. Young, M. Y. Batiuk, K. Prakash, A. Brown, K. Roberts, M. F. Paredes, R.
1271   Kawaguchi, J. H. Stockley, K. Sabeur, S. M. Chang, E. Huang, P. Hutchinson, E. M. Ullian, M.
1272   Hemberg, G. Coppola, M. G. Holt, D. H. Geschwind, D. H. Rowitch, Astrocyte layers in the
1273   mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat. Neurosci.* **23**,
1274   500–509 (2020).

1275 40. M. Y. Batiuk, A. Martirosyan, J. Wahis, F. de Vin, C. Marneffe, C. Kusserow, J. Koeppen, J. F.
1276   Viana, J. F. Oliveira, T. Voet, C. P. Ponting, T. G. Belgard, M. G. Holt, Identification of region-
1277   specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220 (2020).

1278   41. A. Kriegstein, A. Alvarez-Buylla, The glial nature of embryonic and adult neural stem cells. *Annu.*
1279   *Rev. Neurosci.* **32**, 149–184 (2009).

1280   42. C. C. Harwell, L. C. Fuentealba, A. Gonzalez-Cerrillo, P. R. L. Parker, C. C. Gertz, E. Mazzola, M.
1281   T. Garcia, A. Alvarez-Buylla, C. L. Cepko, A. R. Kriegstein, Wide Dispersion and Diversity of
1282   Clonally Related Inhibitory Interneurons. *Neuron*. **87**, 999–1007 (2015).

1283   43. C. Mayer, C. Hafemeister, R. C. Bandler, R. Machold, R. B. Brito, X. Jaglin, K. Allaway, A. Butler,
1284   G. Fishell, R. Satija, Developmental diversification of cortical inhibitory interneurons. *Nature*. **555**,
1285   457–462 (2018).

1286   44. C. Neyt, M. Welch, A. Langston, J. Kohtz, G. Fishell, A short-range signal restricts cell movement
1287   between telencephalic proliferative zones. *J. Neurosci.* **17**, 9194–9203 (1997).

1288   45. F. García-Moreno, M. Pedraza, L. G. Di Giovannantonio, M. Di Salvio, L. López-Mascaraque, A.
1289   Simeone, J. A. De Carlos, A neuronal migratory pathway crossing from diencephalon to
1290   telencephalon populates amygdala nuclei. *Nat. Neurosci.* **13**, 680–689 (2010).

1291   46. A. J. Granger, M. L. Wallace, B. L. Sabatini, ScienceDirect Multi-transmitter neurons in the
1292   mammalian central nervous system. *Curr. Opin. Neurobiol.* **45**, 85–91 (2017).

1293   47. N. X. Tritsch, A. J. Granger, B. L. Sabatini, Mechanisms and functions of GABA co-release. *Nat.*
1294   *Genet.* **17**, 139–145 (2016).

1295   48. M. Bupesh, I. Legaz, A. Abellán, L. Medina, Multiple telencephalic and extratelencephalic
1296   embryonic domains contribute neurons to the medial extended amygdala. *J. Comp. Neurol.* **519**,
1297   1505–1525 (2011).

1298   49. L. Zhang, V. S. Hernandez, C. R. Gerfen, S. Z. Jiang, L. Zavala, R. A. Barrio, L. E. Eiden,
1299   Behavioral role of PACAP signaling reflects its selective distribution in glutamatergic and
1300   GABAergic neuronal subpopulations. *Elife*. **10** (2021), doi:10.7554/eLife.61718.

1301   50. N. Bozadjieva-Kramer, R. A. Ross, D. Q. Johnson, H. Fenselau, D. L. Haggerty, B. Atwood, B.
1302   Lowell, J. N. Flak, The Role of Mediobasal Hypothalamic PACAP in the Control of Body Weight
1303   and Metabolism. *Endocrinology*. **162** (2021), doi:10.1210/endocr/bqab012.

1304   51. J.-H. Cho, K. Zushida, G. P. Shumyatsky, W. A. Carlezon Jr, E. G. Meloni, V. Y. Bolshakov,
1305   Pituitary adenylate cyclase-activating polypeptide induces postsynaptically expressed potentiation in
1306   the intra-amygdala circuit. *J. Neurosci.* **32**, 14165–14177 (2012).

1307   52. E. Y. Lee, L. C. Chan, H. Wang, J. Lieng, M. Hung, Y. Srinivasan, J. Wang, J. A. Waschek, A. L.
1308   Ferguson, K.-F. Lee, N. Y. Yount, M. R. Yeaman, G. C. L. Wong, PACAP is a pathogen-inducible
1309   resident antimicrobial neuropeptide affording rapid and contextual molecular host defense of the
1310   brain. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.1917623117.

1311   53. Y. Kita, H. Nishibe, Y. Wang, T. Hashikawa, S. S. Kikuchi, M. U, A. C. Yoshida, C. Yoshida, T.
1312   Kawase, S. Ishii, H. Skibbe, T. Shimogori, Cellular-resolution gene expression profiling in the
1313   neonatal marmoset brain reveals dynamic species- and region-specific differences. *Proc. Natl. Acad.*
1314   *Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2020125118.

1315   54. T. Shimogori, A. Abe, Y. Go, T. Hashikawa, N. Kishi, S. S. Kikuchi, Y. Kita, K. Niimi, H. Nishibe,
1316   M. Okuno, K. Saga, M. Sakurai, M. Sato, T. Serizawa, S. Suzuki, E. Takahashi, M. Tanaka, S.

1317 Tatsumoto, M. Toki, M. U, Y. Wang, K. J. Windak, H. Yamagishi, K. Yamashita, T. Yoda, A. C.
1318 Yoshida, C. Yoshida, T. Yoshimoto, H. Okano, Digital gene atlas of neonate common marmoset
1319 brain. *Neurosci. Res.* **128**, 1–13 (2018).

1320 55. D. P. Orquera, M. B. Tavella, F. S. J. de Souza, S. Nasif, M. J. Low, M. Rubinstein, The
1321 Homeodomain Transcription Factor NKX2.1 Is Essential for the Early Specification of Melanocortin
1322 Neuron Identity and Activates Pomc Expression in the Developing Hypothalamus. *J. Neurosci.* **39**,
1323 4023–4035 (2019).

1324 56. R. Murcia-Ramón, V. Company, I. Juárez-Leal, A. Andreu-Cervera, F. Almagro-García, S.
1325 Martínez, D. Echevarría, E. Puelles, Neuronal tangential migration from Nkx2.1-positive
1326 hypothalamus. *Brain Struct. Funct.* **225**, 2857–2869 (2020).

1327 57. C. E. Collins, D. C. Airey, N. A. Young, D. B. Leitch, J. H. Kaas, Neuron densities vary across and
1328 within cortical areas in primates. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15927–15932 (2010).

1329 58. B. Schuman, R. P. Machold, Y. Hashikawa, J. Fuzik, G. J. Fishell, B. Rudy, Four Unique
1330 Interneuron Populations Reside in Neocortical Layer 1. *J. Neurosci.* **39**, 125–139 (2019).

1331 59. M. Valero, T. J. Viney, R. Machold, S. Mederos, I. Zutshi, B. Schuman, Y. Senzai, B. Rudy, G.
1332 Buzsáki, Sleep down state-active ID2/Nkx2.1 interneurons in the neocortex. *Nat. Neurosci.* **24**, 401–
1333 411 (2021).

1334 60. B. Zingg, H. Hintiryan, L. Gou, M. Y. Song, M. Bay, M. S. Bienkowski, N. N. Foster, S. Yamashita,
1335 I. Bowman, A. W. Toga, H.-W. Dong, Neural networks of the mouse neocortex. *Cell.* **156**, 1096–
1336 1111 (2014).

1337 61. H. B. M. Uylings, H. J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex? *Behav. Brain Res.*
1338 **146**, 3–17 (2003).

1339 62. R. L. Buckner, D. S. Margulies, Macroscale cortical organization and a default-like apex transmodal
1340 network in the marmoset monkey. *Nat. Commun.* **10**, 1976 (2019).

1341 63. M. K. Lin, Y. S. Takahashi, B.-X. Huo, M. Hanada, J. Nagashima, J. Hata, A. S. Tolpygo, K. Ram,
1342 B. C. Lee, M. I. Miller, M. G. Rosa, E. Sasaki, A. Iriki, H. Okano, P. Mitra, A high-throughput
1343 neurohistological pipeline for brain-wide mesoscale connectivity mapping of the common marmoset.
1344 *Elife*. **8** (2019), doi:10.7554/eLife.40042.

1345 64. P. Majka, S. Bai, S. Bakola, S. Bednarek, J. M. Chan, N. Jermakow, L. Passarelli, D. H. Reser, P.
1346 Theodoni, K. H. Worthy, X.-J. Wang, D. K. Wójcik, P. P. Mitra, M. G. P. Rosa, Open access
1347 resource for cellular-resolution analyses of corticocortical connectivity in the marmoset monkey.
1348 *Nat. Commun.* **11**, 1133 (2020).

1349 65. A. Hladnik, D. Džaja, S. Darmopil, N. Jovanov-Milošević, Z. Petanjek, Spatio-temporal extension in
1350 site of origin for cortical calretinin neurons in primates. *Front. Neuroanat.* **8**, 50 (2014).

1351 66. E. Y. Choi, B. T. T. Yeo, R. L. Buckner, The organization of the human striatum estimated by
1352 intrinsic functional connectivity. *J. Neurophysiol.* **108**, 2242–2263 (2012).

1353 67. K. M. Anderson, F. M. Krienen, E. Y. Choi, J. M. Reinen, B. T. T. Yeo, A. J. Holmes, Gene
1354 expression links functional networks across cortex and striatum. *Nat. Commun.* **9**, 1428 (2018).

1355 68. M. Matamales, J. Götz, J. Bertran-Gonzalez, Quantitative Imaging of Cholinergic Interneurons

1356   Reveals a Distinctive Spatial Organization and a Functional Gradient across the Mouse Striatum.
1357   *PLoS One*. **11**, e0157682 (2016).

1358   69.   E. Fino, M. Vandecasteele, S. Perez, F. Saudou, L. Venance, Region-specific and state-dependent
1359   action of striatal GABAergic interneurons. *Nat. Commun.* **9**, 3339 (2018).

1360   70.   A. B. M. Manchado, C. B. Gonzales, A. Zeisel, H. Munguba, B. Bekkouche, N. G. Skene, P.
1361   Lonnerberg, J. Ryge, K. D. Harris, S. Linnarsson, J. Hjerling-Leffler, Diversity of Interneurons in
1362   the Dorsal Striatum Revealed by Single-Cell RNA Sequencing and PatchSeq. *Cell Rep.* **24**, 2179–
1363   2190.e7 (2018).

1364   71.   T. Zerucha, T. Stühmer, G. Hatch, B. K. Park, Q. Long, G. Yu, A. Gambarotta, J. R. Schultz, J. L.
1365   Rubenstein, M. Ekker, A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of
1366   cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**, 709–
1367   721 (2000).

1368   72.   T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with
1369   Signac. *Nat. Methods*. **18**, 1333–1341 (2021).

1370   73.   N. L. Jorstad, J. H. T. Song, D. Exposito-Alonso, H. Suresh, N. Castro, F. M. Krienen, A. M. Yanny,
1371   J. Close, E. Gelfand, K. J. Travaglini, S. Basu, M. Beaudin, D. Bertagnolli, M. Crow, S.-L. Ding, J.
1372   Eggermont, A. Glandon, J. Goldy, T. Kroes, B. Long, D. McMillen, T. Pham, C. Rimorin, K. Siletti,
1373   S. Somasundaram, M. Tieu, A. Torkelson, K. Ward, G. Feng, W. D. Hopkins, T. Höllt, C. Dirk
1374   Keene, S. Linnarsson, S. A. McCarroll, B. P. Lelieveldt, C. C. Sherwood, K. Smith, C. A. Walsh, A.
1375   Dobin, J. Gillis, E. S. Lein, R. D. Hodge, T. E. Bakken, Comparative transcriptomics reveals human-
1376   specific cortical features. *bioRxiv* (2022), p. 2022.09.19.508480.

1377   74.   D. Velmeshev, L. Schirmer, D. Jung, M. Haeussler, Y. Perez, S. Mayer, A. Bhaduri, N. Goyal, D. H.
1378   Rowitch, A. R. Kriegstein, Single-cell genomics identifies cell type–specific molecular changes in
1379   autism. *Science*. **364**, 685–689 (2019).

1380   75.   W. B. Ruzicka, S. Mohammadi, J. F. Fullard, J. Davila-Velderrain, S. Subburaju, D. R. Tso, M.
1381   Hourihan, S. Jiang, H.-C. Lee, J. Bendl, G. Voloudakis, V. Haroutunian, G. E. Hoffman, P. Roussos,
1382   M. Kellis, PsychENCODE Consortium, Single-cell multi-cohort dissection of the schizophrenia
1383   transcriptome. *bioRxiv* (2022), , doi:10.1101/2022.08.31.22279406.

1384   76.   C. Nagy, M. Maitra, A. Tanti, M. Suderman, J.-F. Théroux, M. A. Davoli, K. Perlman, V. Yerko, Y.
1385   C. Wang, S. J. Tripathy, P. Pavlidis, N. Mechawar, J. Ragoussis, G. Turecki, Single-nucleus
1386   transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte
1387   precursor cells and excitatory neurons. *Nat. Neurosci.* **23**, 771–781 (2020).

1388   77.   A. Sequeira, F. Mamdani, C. Ernst, M. P. Vawter, W. E. Bunney, V. Lebel, S. Rehal, T. Klempan,
1389   A. Gratton, C. Benkelfat, G. A. Rouleau, N. Mechawar, G. Turecki, Global brain gene expression
1390   analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One*.
1391   **4**, e6585 (2009).

1392   78.   D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr.
1393   Opin. Neurobiol.* **56**, 144–152 (2019).

1394   79.   R. N. Delgado, D. E. Allen, M. G. Keefe, W. R. Mancia Leon, R. S. Ziffra, E. E. Crouch, A.
1395   Alvarez-Buylla, T. J. Nowakowski, Individual human cortical progenitors can produce excitatory
1396   and inhibitory neurons. *Nature*. **601**, 397–403 (2022).

80. 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).

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

82. 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).

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

84. 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).

85. 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).

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

87. 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).

88. 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.

89. 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).

90. L. Feng, T. Zhao, J. Kim, neuTube 1.0: A New Design for Efficient Neuron Reconstruction Software

1439   Based on the SWC Format. *eNeuro*. **2** (2015), doi:10.1523/ENEURO.0049-14.2014.

1440 91. R. C. Challis, S. Ravindra Kumar, K. Y. Chan, C. Challis, K. Beadle, M. J. Jang, H. M. Kim, P. S.
1441   Rajendran, J. D. Tompkins, K. Shivkumar, B. E. Deverman, V. Gradinaru, Systemic AAV vectors
1442   for widespread and targeted gene delivery in rodents. *Nat. Protoc.* **14**, 379–414 (2019).

1443 **Acknowledgements:**

1463 **Author Contributions:**
1464 RNA/ATAC Data Generation: FMK, MG, AL
1465 Spatial Data Generation: KML, HZ
1466 Data Analysis: FMK, KML, HZ, RCHdR, MES, MG, AL, KXL, VFB-G, TWS, SAM
1467 Data Interpretation: FMK, KML, HZ, RCHdR, MES, MG, KXL, VFB-G, SB, EB, SAM, GF
1468 Tissue Samples/Tissue Processing: FMK, KML, HZ, MG, AL, QZ, GC, SB
1469 AAV design, generation & experiments: MW, CC, QZ, JS, SJV, JD, KC, BED
1470 Morphology Data generation/analysis: KML, HZ, VFB-G, TWS, AM, EB
1471 Software/Data management: FMK, RCHdR, MG, AW, JN, SK
1472 Writing: FMK, KML, HZ, MES, VFB-G, SAM, GF
1473

1474 **Competing interests:** Authors declare that they have no competing interests.

1475

1476 **Data and materials availability:**
1477 Raw sequence data were produced as part of the BRAIN Initiative Cell Census Network are available for
1478 download from the Neuroscience Multi-omics Archive
1479 (https://assets.nemoarchive.org/dat-1je0mn3) and the Brain Cell Data Center (https://biccn.org/data).
1480 Morphological reconstructions and single molecule FISH of interneuron types are available for download
1481 through the Brain Image Library (https://submit.brainimagelibrary.org/search?grant_number=1-U01-

1482   MH114819-01). The AAV-BI103 rep-cap plasmid will be made available through Addgene upon

1483   publication of the characterization of this capsid and through direct requests to Ben Deverman.
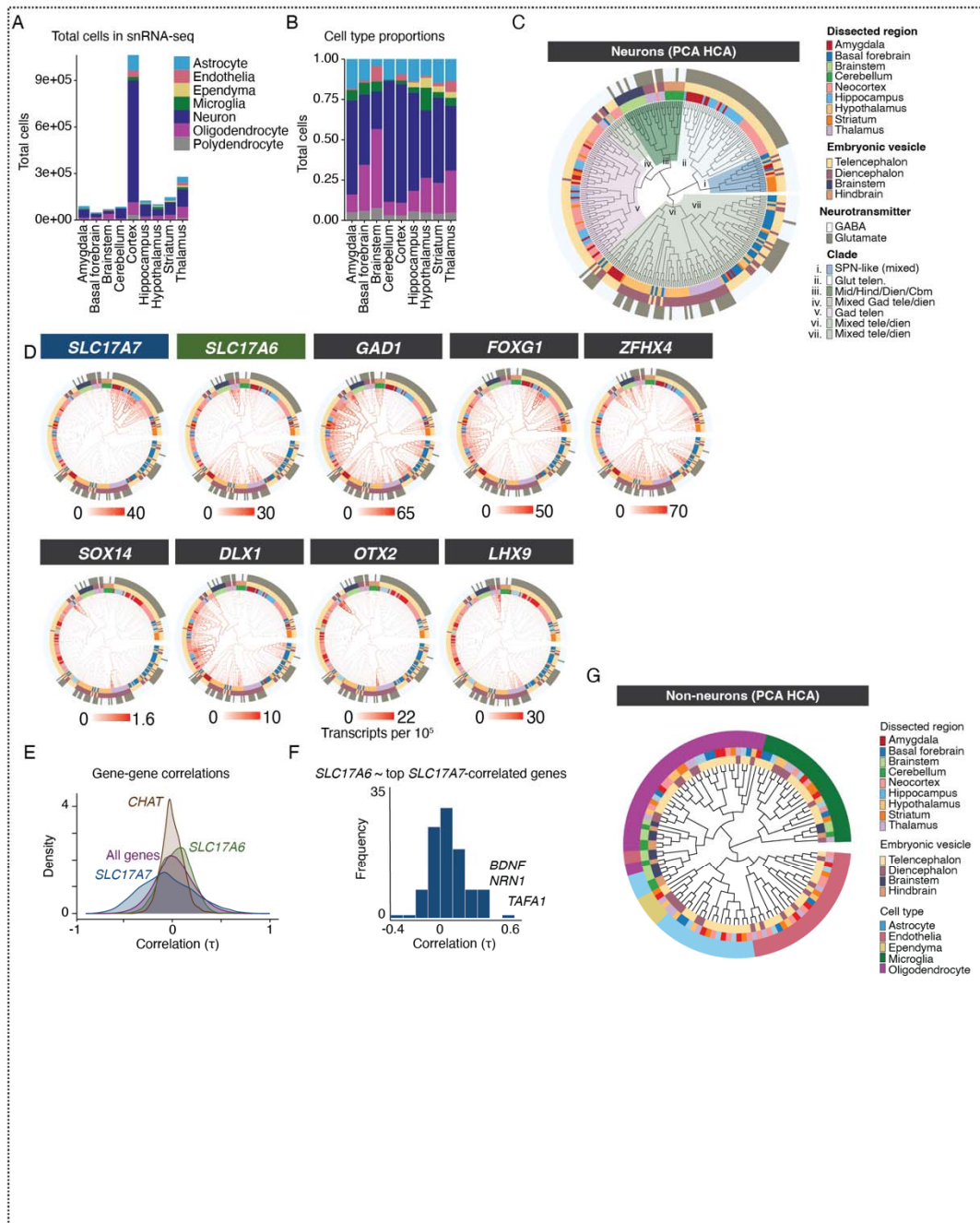

1484   **Supplementary Materials:**

1485   Figs. S1-S11

1486   Table S1-S7

1487   Data S1-S2

1488   Legends:

FIGURE 1

Figure 1. Single nucleus RNA sequencing of marmoset brain. (A) Number of nuclei per brain structure. Colors indicate cell classes. **(B)** Proportions of cell classes across brain structures. **(C)** Neurons in each dissected region are clustered separately, then pseudo-bulked "metacells" of all 288 clusters are arranged in the dendrogram using hierarchical clustering of top 100 PCA scores of expressed genes. Outer rings colored by dissected subregion (1st ring), cephalic vescicle (2nd ring) and major neurotransmitter usage (3rd ring). Seven major "clades" are colored and labeled according to cell type and regional composition of clade. **(D)** Expression of markers for glutamatergic neurons (*SLC17A7*, *SLC17A6*), GABAergic neurons (*GAD1*), and transcription factors (*FOXG1*, *ZFHX4*, *SOX14*, *DLX1*, *OTX2*, *LHX9*) are plotted as heatmaps on dendrogram shown in (C). These transcription factors are largely restricted to specific clades or are associated with particular cephalic vesicles. **(E)** Gene-gene correlation (Spearman tau) distributions across all neuron clusters in (C) for *SLC17A7*, *SLC17A6*, and *CHAT* (marker for cholinergic neurons) as well as background (all genes). The distribution of pairwise correlations to *CHAT* had a lower 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. **(F)** Distribution of cross-cell-type correlation to *SLC17A6* of genes most correlated with *SLC17A7* (top 116 genes with Spearman tau > 0.5 to *SLC17A7*). **(G)** Same hierarchical clustering procedure as (C) but for non-neuronal cell types. Outer ring colors indicate major non-neuronal cell class. Inner ring colors indicate region dissection and vesicle; colors as in (C).
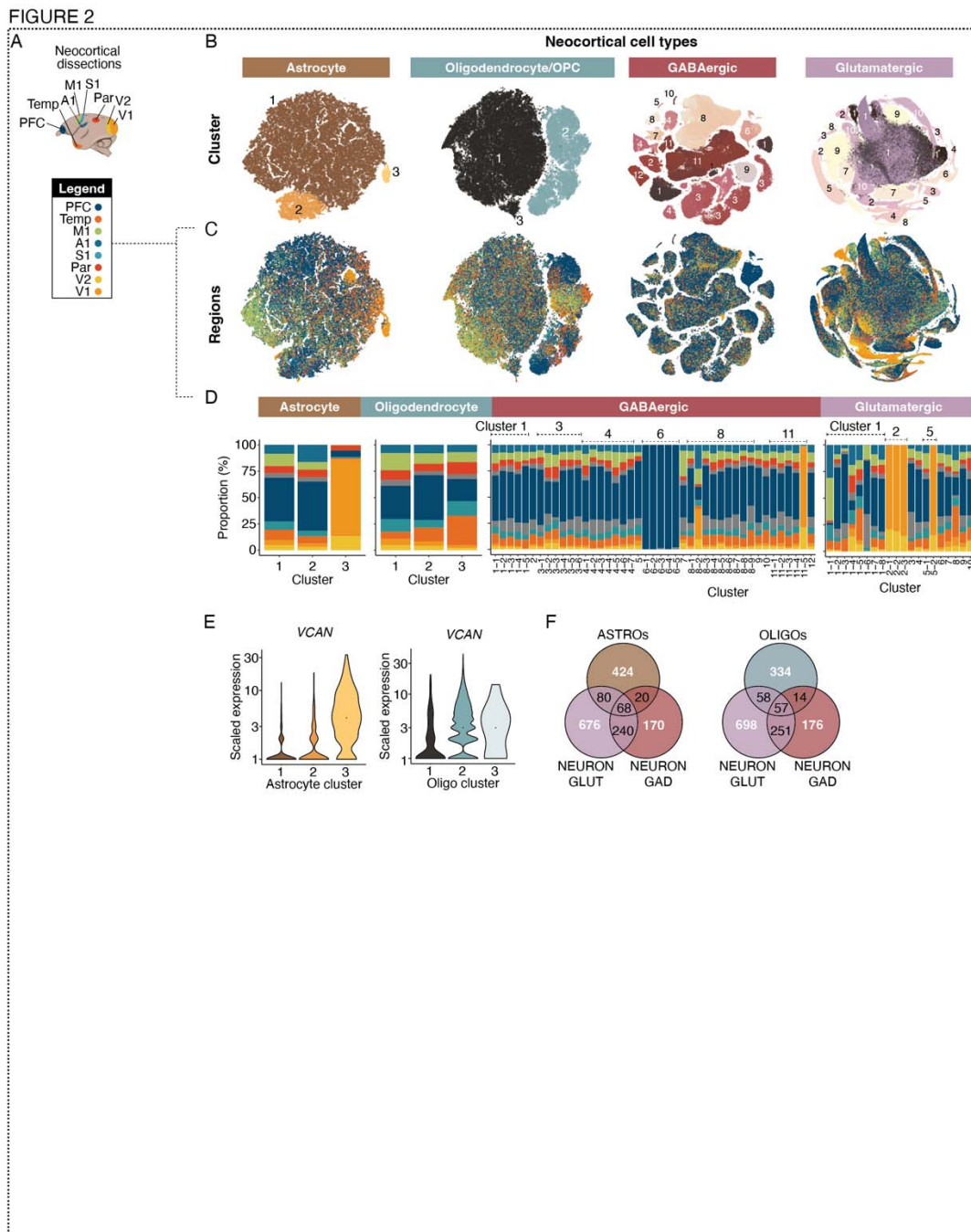
1509
Figure 2. Regional variation in neocortical cell types and expression patterns. (A) Cortical regions
sampled. **(B)** t-SNE embeddings of neocortical macroglia (astrocytes, oligodendrocyte lineage types) and
neurons (GABAergic interneurons, glutamatergic neurons). Colors represent clusters (numbered). **(C)**

1513   Same as (B) but cells colored by neocortical dissection. **(D)** Regional proportions of each cluster, colors

1514   same as (A). **(E)** *VCAN* expression across astrocyte clusters and oligodendrocyte lineage clusters. Colors

1515   as in (B). **(F)** Venn diagrams showing overlap of neocortical regionally differentially expressed genes

1516   (rDEGs) across GABAergic neurons, glutamatergic neurons, astrocytes and oligodendrocyte lineage cells.

1517   rDEGs are defined as > 3 fold expression difference in homologous cell types across pairs of cortical

1518   regions.

FIGURE 3
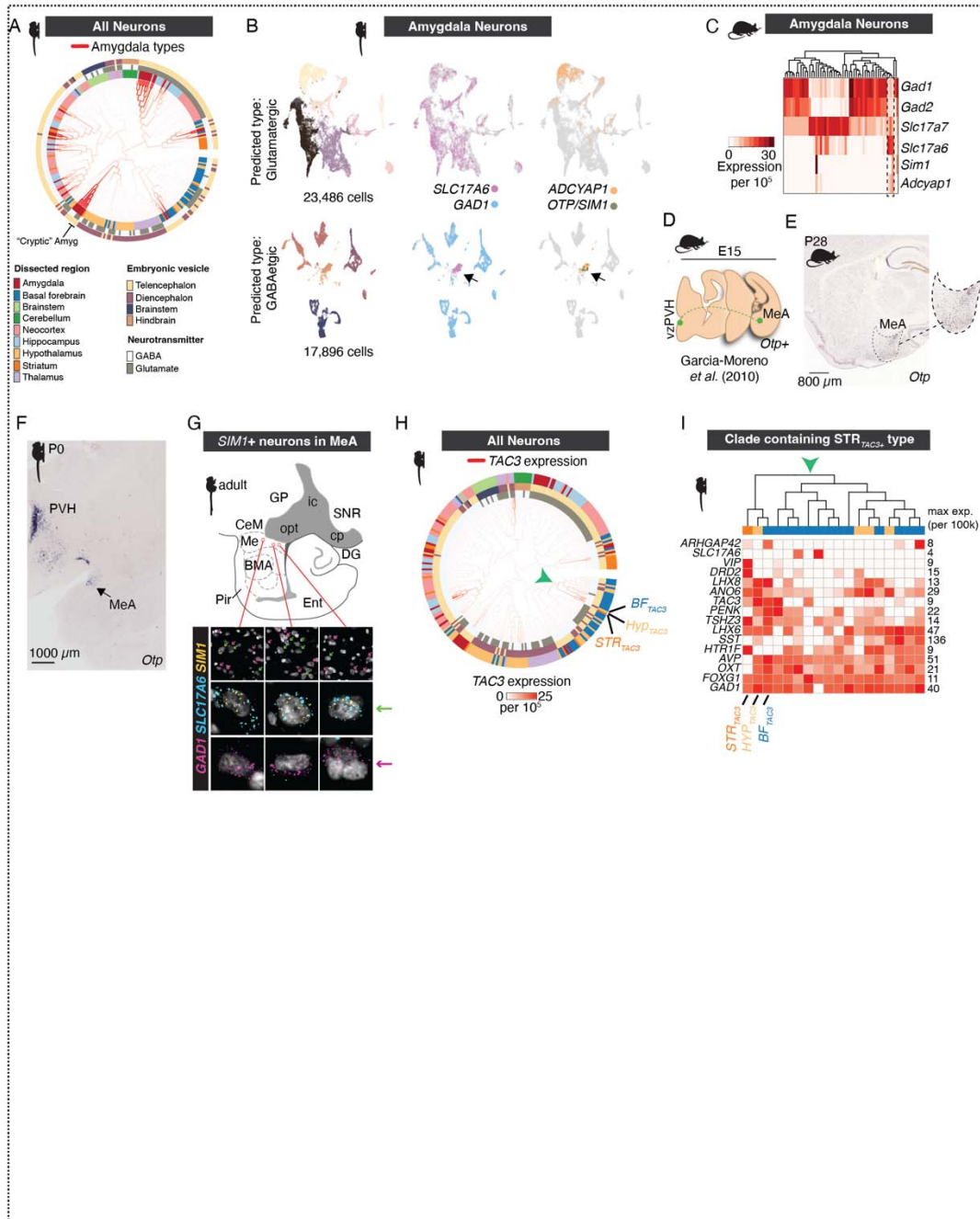
1519
Figure 3. Examples suggesting cross-cephalic-boundary cell type migration. (A) Locations of amygdala clusters in the dendrogram from **Fig. 1C**. **(B)** Clustering of marmoset amygdala cells (n =

1522   44,165 neurons) predicted to be glutamatergic (top row) or GABAergic (bottom row) based on a linear

1523   classifier (scPred) trained on supervised cell type models. t-SNEs for each class are colored by cluster

1524   (first column) or genes (*SLC17A6*, *GAD1, ADCYAP1, OTP, SIM1*). Arrowhead indicates cluster of

1525   neurons that are classified as GABAergic yet do not express *GAD1* and do express *SLC17A6*. **(C)**

1526   snRNA-seq from mouse amygdala neurons. Heatmap shows normalized, scaled expression of

1527   neurotransmitters in neurons from snRNA-seq of adult mouse amygdala (n = 25,930 nuclei, 3 replicates

1528   pooled). Dendrogram shows hierarchical clustering of neuron types. Dotted outline shows presence of 3

1529   *Slc17a6+* subtypes that preferentially cluster with GABAergic (*GAD1+/GAD2+*) subtypes and that

1530   express *Sim1* and/or *Adcyap1*. **(D)** Cartoon of embryonic migration of *Otp+* cells that migrate from

1531   proliferative zones around the 3rd ventricle to periventricular hypothalamus (vzPZH) and medial

1532   amygdala (MeA), following data in Garcia-Moreno et al. (*45*). **(E)** ISH for *Otp* in sagittal section of P28

1533   mouse brain. Dotted outline indicates borders of medial amygdala (MeA). **(F)** Marmoset P0 coronal

1534   section showing ISH staining for *OTP*. Data from (https://gene-atlas.brainminds.jp/; (Shimogori et al.

1535   2018; Kita et al. 2021)). **(G)** Cartoon of marmoset amygdala the sagittal plane. FISH staining for *GAD1*

1536   (magenta), *SLC17A6* (cyan), and *SIM1* (yellow) in the medial amygdala (Me). Magenta arrows highlight

1537   *GAD1* expressing nuclei. Green arrows highlight *SLC17A6* and *SIM1* dual expressing nuclei. **(H)**

1538   Marmoset neuronal dendrogram shown in **Fig. 1C** with clades colored by *TAC3* expression.

1539   Unexpectedly, a novel primate-specific striatal *TAC3* population (*3*) clusters with hypothalamic and basal

1540   forebrain populations, and not with other telencephalic GABAergic interneurons. Red arrowhead

1541   indicates clade containing three *TAC3+* populations found in striatum, basal forebrain, and hypothalamus.

1542   **(I)** Heatmap of genes expressed in the clade indicated by red arrowhead in (H). Expression values of each

1543   gene is normalized to its max within the clade. The three *TAC3+* populations are labeled and each show

1544   distinct expression differences.

FIGURE 4
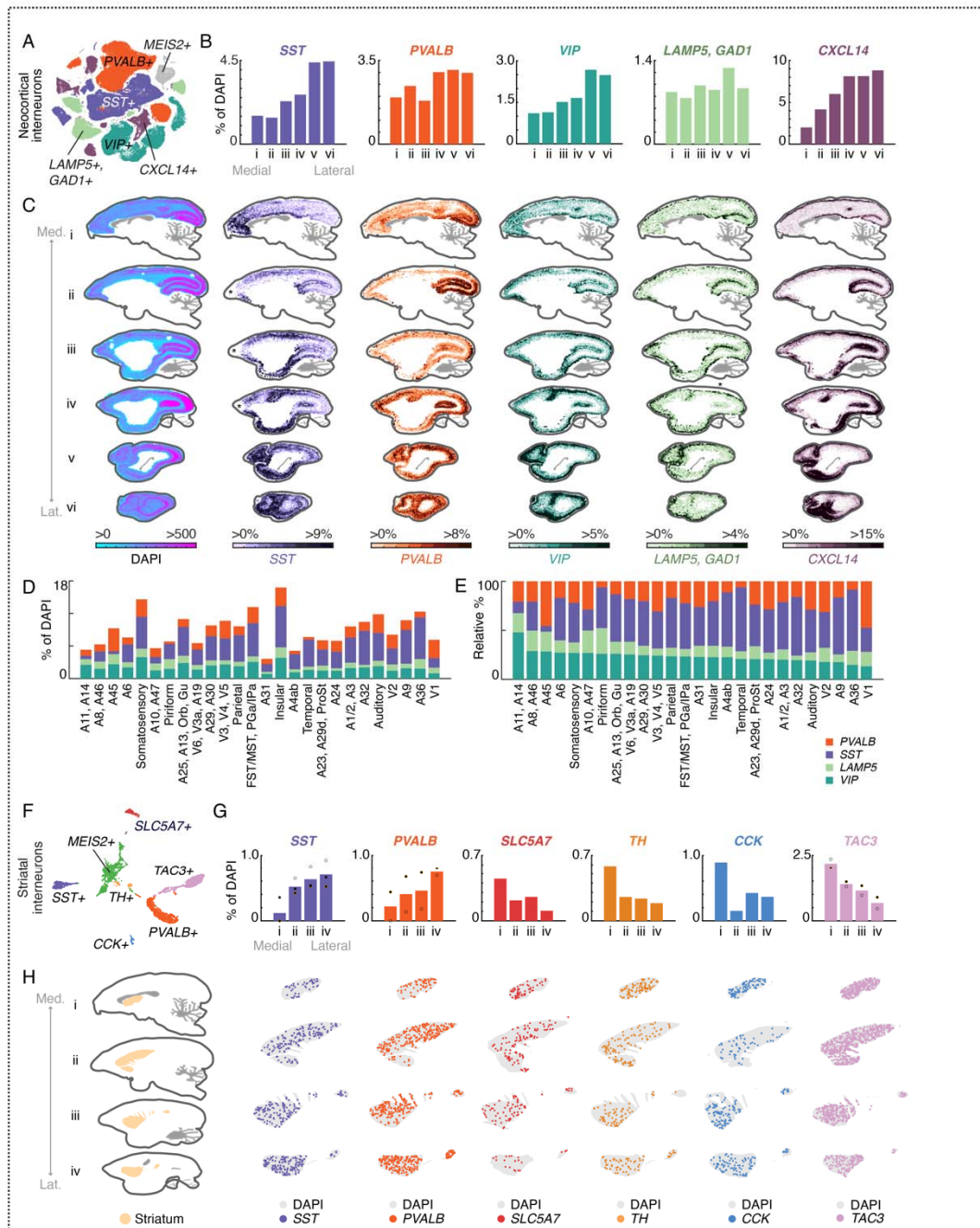
1545
Figure 4. Cell-type-specific distributions of interneurons in primate telencephalon using quantitative smFISH. (A) t-SNE of GABAergic neocortical interneurons as in **Fig. 2B**, colored by

1548   subclass marker (*PVALB, SST, VIP, LAMP5, CXCL14*). Gray points are the *MEIS2+* population that is

1549   restricted to orbiomedial prefrontal cortex and was not spatially profiled. **(B)** Medial-lateral proportions of

1550   each major class as percentage of all cells (DAPI+). Barplots quantify positive cells as proportion of all

1551   (DAPI+) cells from medial to lateral sections of smFISH performed on 6 thin (16 µm) sagittal sections of

1552   marmoset neocortex, each section being 1600 µm apart and covering 9600 µm of neocortex. Colors as in

1553   (A). **(C)** smFISH for neocortical interneuron subclass markers showing locations of cells positive for each

1554   marker across 6 sagittal sections of the marmoset neocortex. First column shows density of all DAPI+

1555   nuclei per unit area (approximately 387 µm per bin) profiled from one series. Heatmap scale in

1556   subsequent columns show percentage of marker-positive cells relative to DAPI+ cells. Average

1557   proportions across section shown in (B). Med. = medial, Lat. = Lateral. Asterisks in section ii and iii of

1558   the SST series and iv of the LAMP5 series denote regions where tissue or staining artifacts caused loss of

1559   signal. These can be seen in the DAPI-only series of these experiments, which show lower overall cell

1560   counts at these locations (**Fig S7**). **(D)** Quantitation of interneuron proportions by cortical area parcellated

1561   according to **Fig. S8B-C**. **(E)** Relative percentages of interneuron proportions by cortical area shown in

1562   Fig. S8B-C, sorted by max relative proportion of *VIP+* interneurons. **(F)** t-SNE of striatal, cholinergic

1563   neurons (*SLC5A7*) and GABAergic interneurons (*SST*, *PVALB*, *TH*, *CCK*, *TAC3*). Green points

1564   correspond to the *MEIS2+* population and were not spatially profiled. **(G)** Medial-lateral gradients of

1565   striatal cell type proportions across 4 sagittal sections. Dots = individual replicates. Colors as in (F). **(H)**

1566   smFISH for striatal cholinergic and interneuron subclass markers showing locations of cells positive for

1567   each subclass marker across 4 (16 µm) sagittal sections of the marmoset striatum, each 1600 µm apart

1568   covering 6400 µm in total. First column shows the anatomical context of the four sagittal striatal planes.

1569   Average proportions across each section are shown in (G).
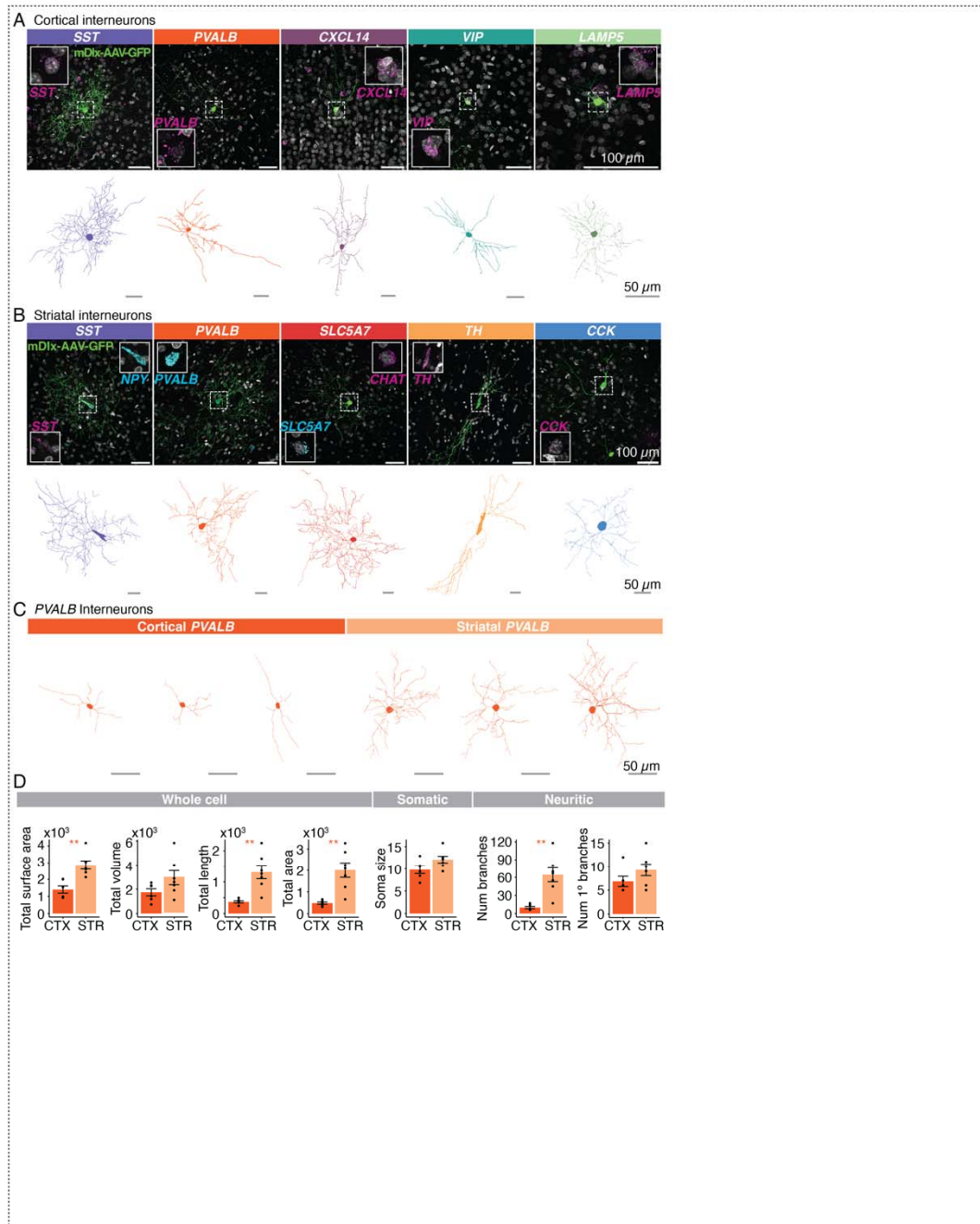
1570
1571 **Figure 5. Morphological characteristics of marmoset cortical and striatal interneurons. (A-B)** *Top*
1572 *rows*, Examples of AAV9-hDlx5/6-GFP-fGFP labeled neocortical **(A)** and striatal **(B)** interneurons. Insets

1573   show magnified cell nucleus of GFP+ cell along with smFISH staining for interneuron type marker to

1574   confirm molecular identity of labeled cell. Scale Bar = 100 µm. *Bottom rows,* Reconstructed skeletonized

1575   morphology (Imaris) of GFP+ cells depicted in *top rows*. Scale Bar = 50 µm. **(C)** Representative

1576   reconstructed neocortical and striatal *PVALB+* interneurons from one marmoset (Cj 17-154; **Table S1,**

1577   **S6**). Scale Bar = 50 µm. **(D)** Quantification of morphological characteristics of *PVALB+* cells from Cj 17-

1578   154. Means ± SEM. *$*P < 0.05$ and $**P < 0.01$. See the Methods section for detailed statistical

1579   information.

FIGURE 6

1580
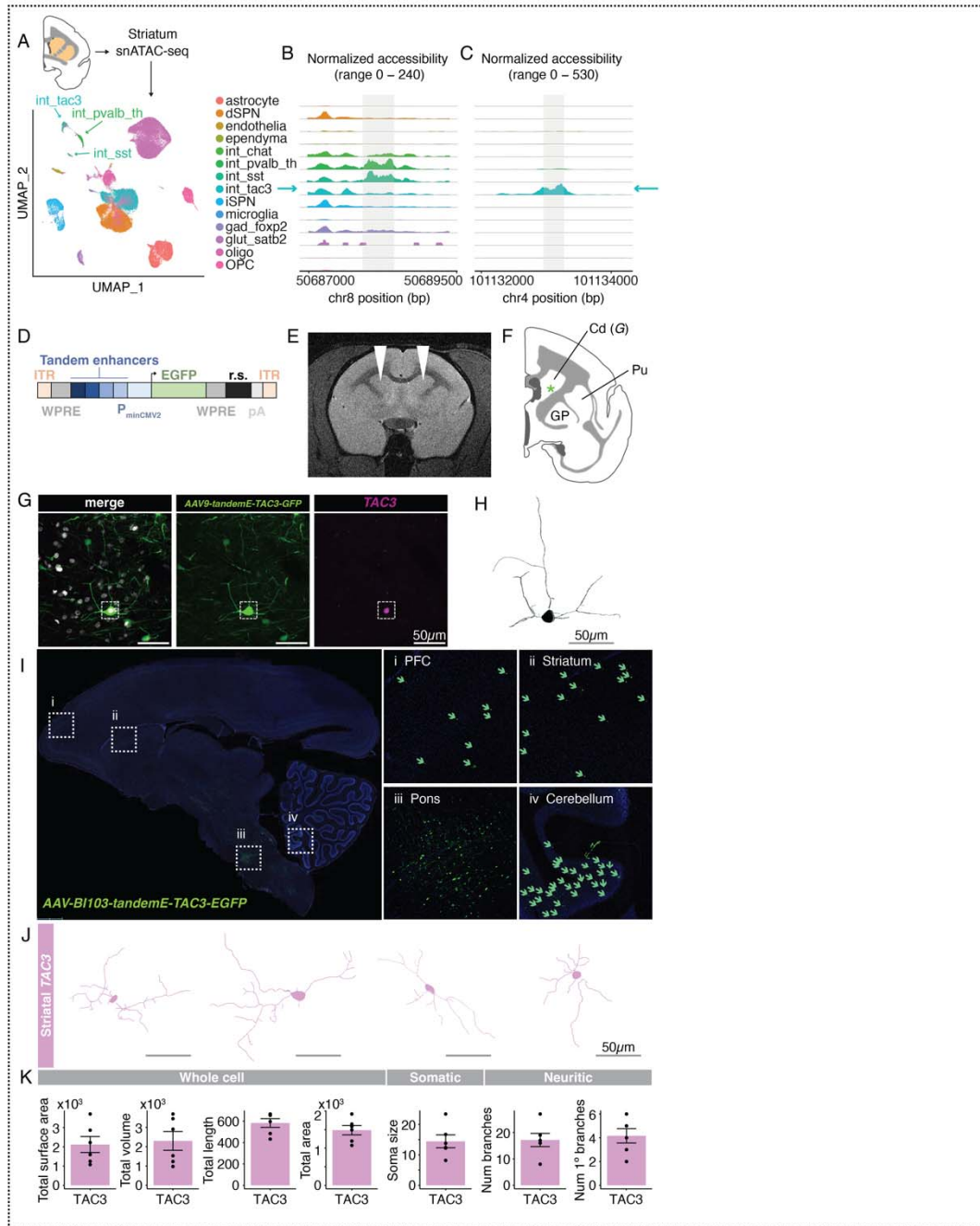
1581 **Figure 6. Development of a novel enhancer-AAV for *TAC3*+ primate-specific striatal interneurons.**
1582 **(A)** snATAC-seq (69,808 nuclei) from fresh marmoset striatum (1 male, Cj 18-153). UMAP shows major
1583 clusters with labels transferred from striatal snRNA-seq data. **(B)** Chromatin accessibility at the locus
1584 corresponding to the mDlx sequence in marmoset shows read pileups in *PVALB*, *SST*, and *CHAT*+
1585 neurons, but not in the *TAC3*+ type. **(C)** Chromatin accessibility of a candidate selective enhancer for the
1586 *TAC3*+ type. **(D)** Sequence construct design for *TAC3* interneuron specific AAV. Four *TAC3* interneuron-
1587 specific regulatory elements (example in (C)) are packaged in tandem upstream of a minimal promoter
1588 driving EGFP expression. ITR = inverted terminal repeats. WPRE = Woodchuck Hepatitis Virus
1589 Posttranscriptional Regulatory Element. PminCMV2 = minimal CMV2 promoter. EGFP = enhanced
1590 green fluorescent protein. R.S. = 300 bp random sequence. pA = polyadenylation sequence. **(E)** MRI
1591 showing injection location (white arrowheads) of the AAV9-tandemE-TAC3-EGFP virus into bilateral
1592 dorsal striatum (caudate) in one animal (Cj 19-207, **Table S1**). **(F)** Cartoon showing location of cell
1593 shown in (G). **(G)** Main: EGFP antibody-amplified confocal image of a labeled cell (position shown in
1594 (F)). Insets: smFISH for *TAC3* showing colocalization. Scale bar = 50 µm. **(H)** Morphological
1595 reconstruction (Imaris) of cell shown in (G). Scale bar = 50 µm. **(I)** Whole sagittal section (20x image) of
1596 adult marmoset showing cells transduced by the AAV-BI103-tandemE-TAC3-EGFP virus in marmoset
1597 Cj 20-214. GFP+/*TAC3*+ cells were detected sparsely in neocortex as well as striatum, cerebellum,
1598 substantia nigra, superior colliculus, and brainstem. (*i*) Prefrontal cortex, (*ii*) Striatum, (*iii*) Pons, (*iv*)
1599 Cerebellum. Green arrows highlight cells in areas with sparse expression. **(J)** Four reconstructions
1600 (Imaris) of striatal GFP+/*TAC3*+ cells from similar section shown in (I). **(K)** Quantification of
1601 morphological parameters of reconstructed cells from striatal GFP+/*TAC3*+ cells from the local injection
1602 in (E) as well as the systemic injection in (I) (**Table S6**).
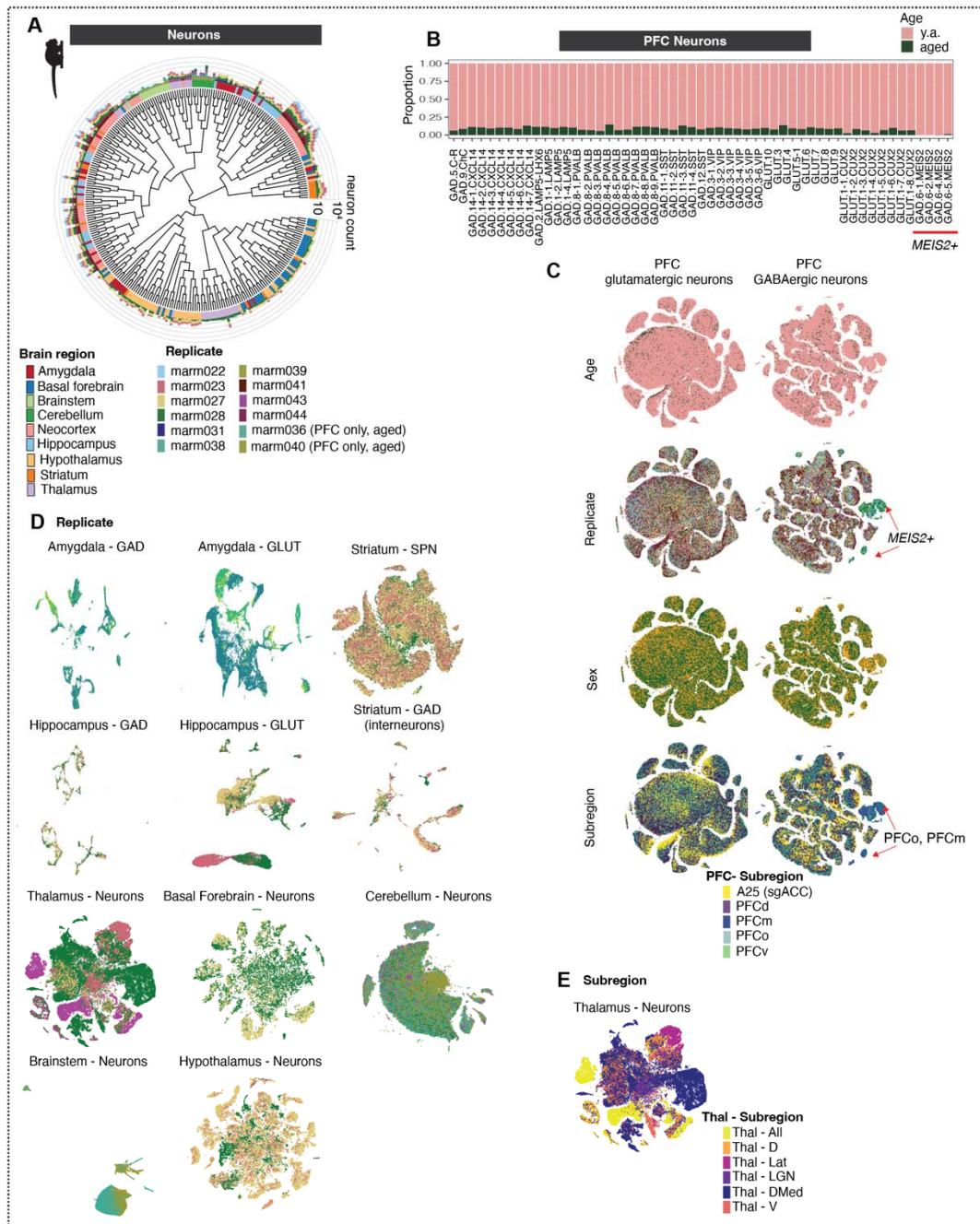
Figure S1. Neuron counts by donor across brain regions. (A) Neuronal dendrogram as in **Fig. 1C**, with outer barplots depicting number of nuclei per cell type and replicate. Ring colors are brain regions,

1606   colors in barplots correspond to replicates. **(B)** Proportional per-cluster representation of PFC neurons

1607   between young adult donors and aged (n=2) donors. While our snRNA-seq collection focused on post-

1608   sexual maturity young adults, we acquired an additional dataset of PFC sampled from 2 aged animals (1

1609   M, 11y0m; 1F, 14y4m, 37,260 cells total; **Table S1**). Individual replicates contributed similar proportions

1610   of neurons to each prefrontal neuron subtype, and clusters generally had proportional representation

1611   across young adults and aged animals, as well as across males and females (**Fig. S1B-C**), suggesting that

1612   these variables do not dramatically impact neuronal ensembles and identities in prefrontal cortex. **(C)** t-

1613   SNE embeddings of PFC neurons (*top row*, GABAergic; *bottom row*, glutamatergic) with colors

1614   representing different metadata: age (young vs aged), replicate, PFC subregion. There was notable

1615   enrichment of *MEIS2+* GABAergic neurons in medial prefrontal and orbital prefrontal dissections **(Fig.**

1616   **S1C)**. Based on their gene expression profiles, these cells likely correspond to the recently described

1617   population of LGE-derived *MEIS2+* neurons that populate the olfactory bulb in mice, and which are

1618   instead directed to medial prefrontal cortex in macaques and humans (*33*). **(D)** t-SNEs of neurons in each

1619   brain structure, with cells colored by replicate (colors as in (A)). Telencephalic neurons are plotted

1620   separately by class: GABAergic and glutamatergic classes (neocortex, hippocampus, amygdala), or

1621   GABAergic interneurons and spiny projection neurons (striatum). Compared with neocortex, greater

1622   cross-donor variability was observed in some subcortical structures such as hypothalamus and thalamus,

1623   though this was likely driven more by dissection variability than by donor variability, as the donor-

1624   specific clusters tended to be from subregions that were only sampled in one individual (see E). **(E)** t-

1625   SNEs of thalamic neurons with cells colored by thalamic subdivision.

1626
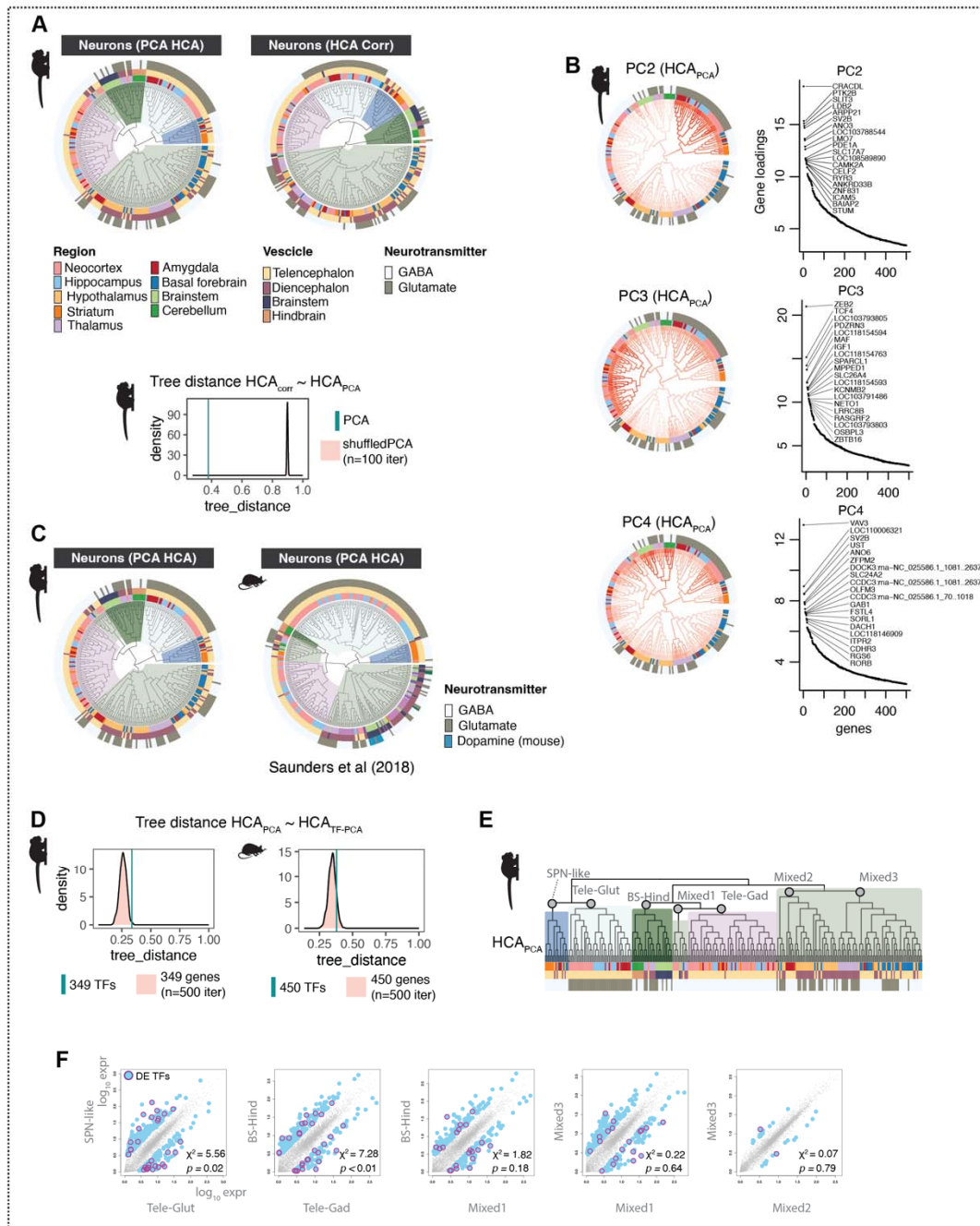
1627

Figure S2. Conservation of neuronal hierarchy across species and clustering methods. (A)
Comparison of marmoset and mouse neurons (mouse atlas data from (*6*)) using three different distance

1630   calculations for hierarchical clustering: HCA_{corr} distance = gene expression correlations of top 5904 genes
1631   in marmoset or 3528 in mouse (mouse genes include all genes with 1:1 orthologs to the 5904 marmoset
1632   genes and that are expressed in at least 10 transcripts per 100,000 in at least one mouse neuron type);
1633 HCA_{PCA} distance = top 100 PCA scores across the same genes; HCA_{TF-PCA} distance = top 100 PCA scores
1634   using expressed transcription factors only (marmoset = 349 TFs, mouse = 450 TFs). **(B)** Distance between
1635   hierarchical clustering (HC) dendrogram trees computed using different methods. Cyan line = tree
1636   distance (R package TreeDist) between hierarchical clustering using distance = HCA_{corr} and using
1637   distance = HCA_{PCA}. Pink distribution is tree distance scores between HCA_{corr} and shuffled PCA scores (n
1638   = 100 shuffling iterations). Lower values of tree_distance (x-axis) mean higher agreement between
1639   dendrogram tree structures. **(C)** PCA loadings and top genes for PC2-PC4. PC scores are plotted on the
1640 HCA_{PCA} dendrogram. Ranked gene loading plots show top 20 genes per PC. **(D)** Tree distances computed
1641   as in (B) between HCA_{PCA} and HCA_{TF-PCA}. The tree distances between these two trees is low, but not
1642   different from distributions of random, same-sized sets of genes. **(E)** Marmoset dendrogram in **Fig. 1C**
1643   (HCA_{PCA}) indicating major clades compared in (F). **(F)** Ancestral reconstruction (AR; R package
1644   phytools) of gene expression profiles of major clades of marmoset neuron types from dendrogram in (E).
1645   Maximum likelihood estimates of gene expression (fastAnc) were computed for 7 major internal nodes
1646   (gray circles) of the HCA_{PCA} dendrogram. Scatterplots show pairwise comparisons between AR of
1647   internal nodes of major clades. Blue dots = genes with >3 foldchange difference between the two ARs.
1648   Magenta circles = differentially expressed transcription factors (DE-TFs). Chi-square and p-values
1649   describe whether TFs are significantly differentially enriched between the AR pairs.
1650

Supp S3
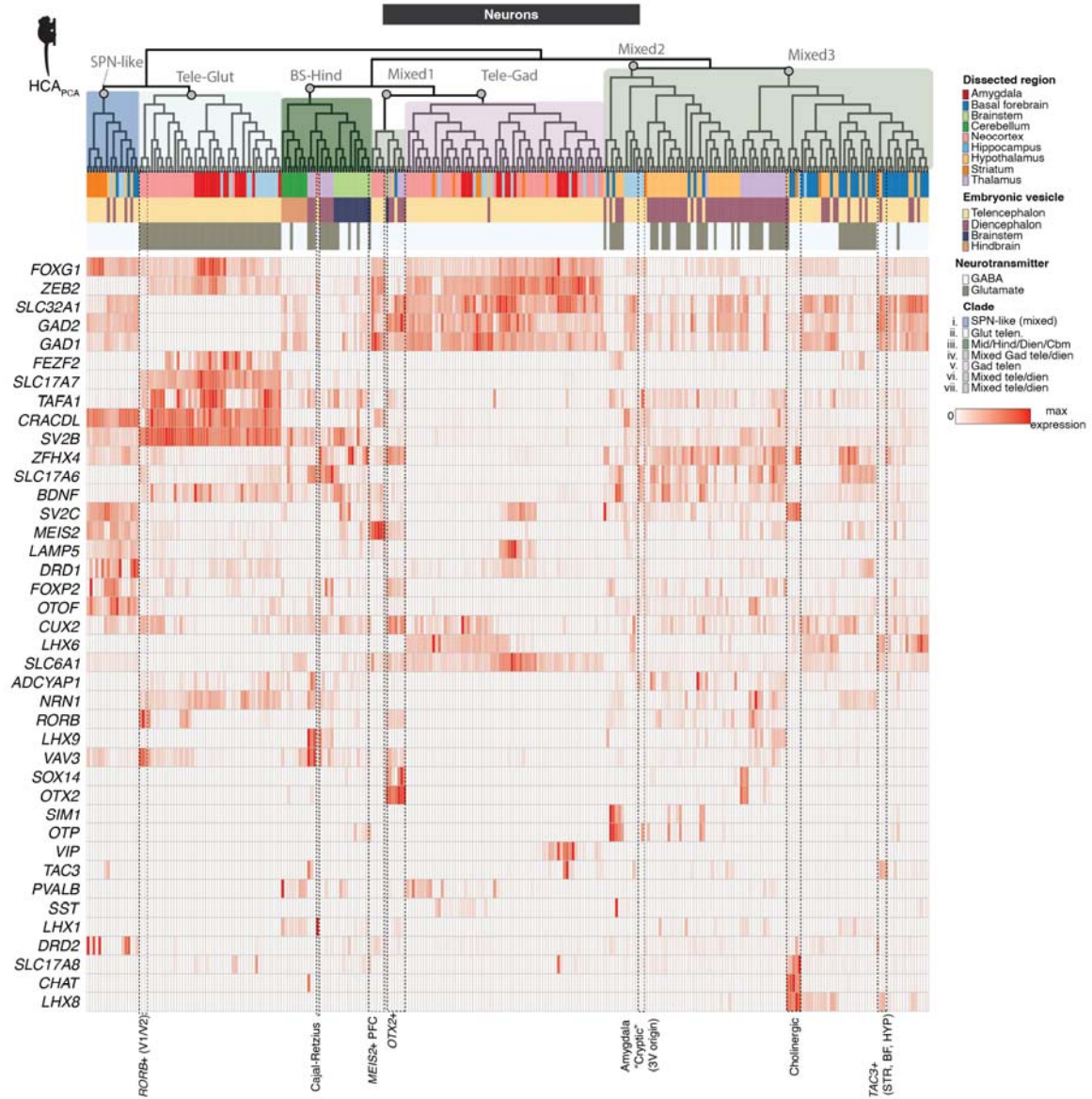
1651
Figure S3. Gene expression across neural populations. Expression of broad class marker genes and other genes of interest across all neurons sampled by snRNA-seq. Heatmap colors are scaled to max

1654   normalized expression for each row (gene). Dendrogram ordering and metadata colors as in **Fig. 1C**. Cell

1655   types discussed in the main text are labeled at bottom.

1656

Supp S4 related to Fig 2
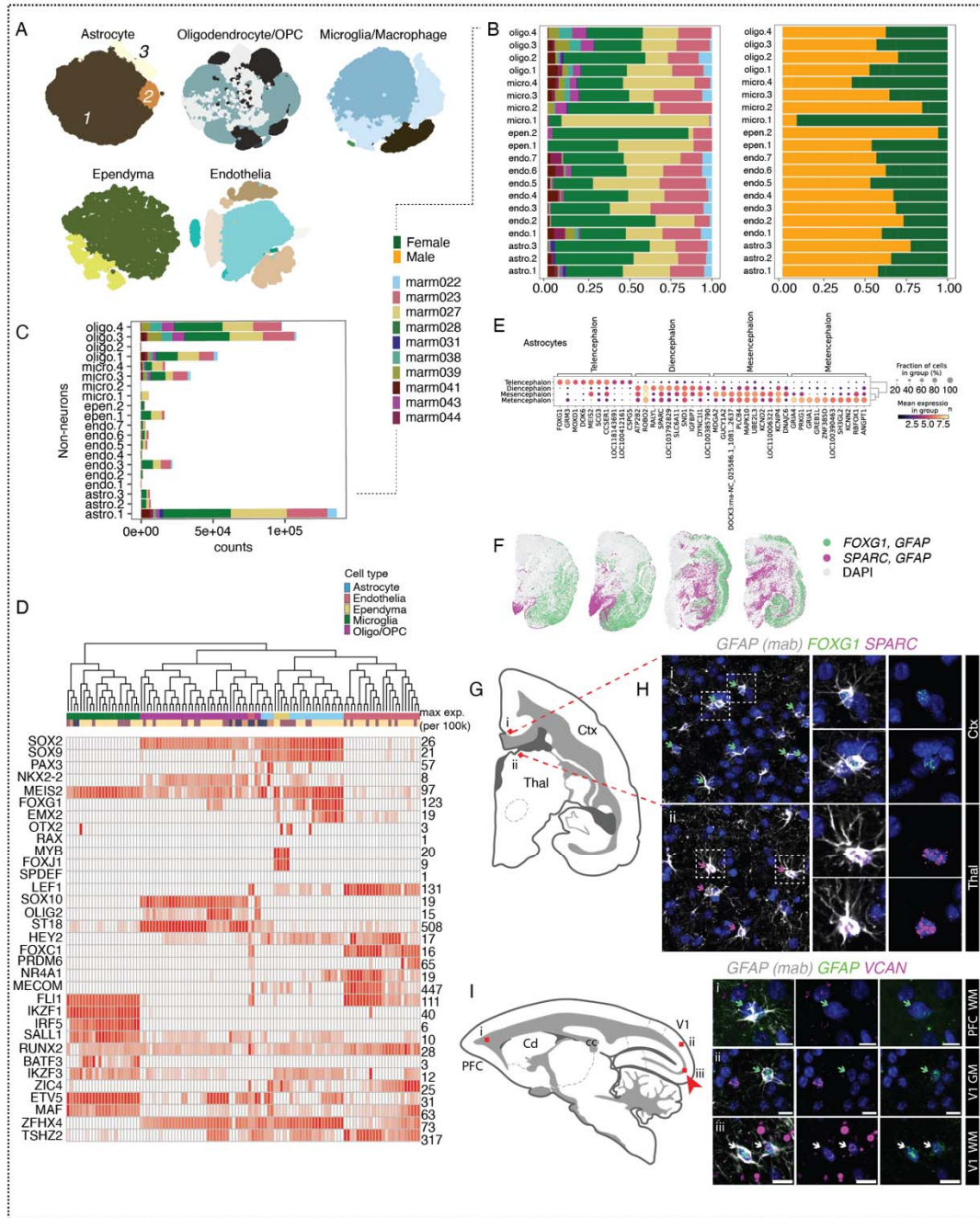
1657

Figure S4. Glia diversity across regions. (A) t-SNE embeddings of major non-neuronal types colored
by cluster. **(B)** Barplots of glial proportions colored by donor and by sex. **(C)** Non-neuronal nuclei counts.

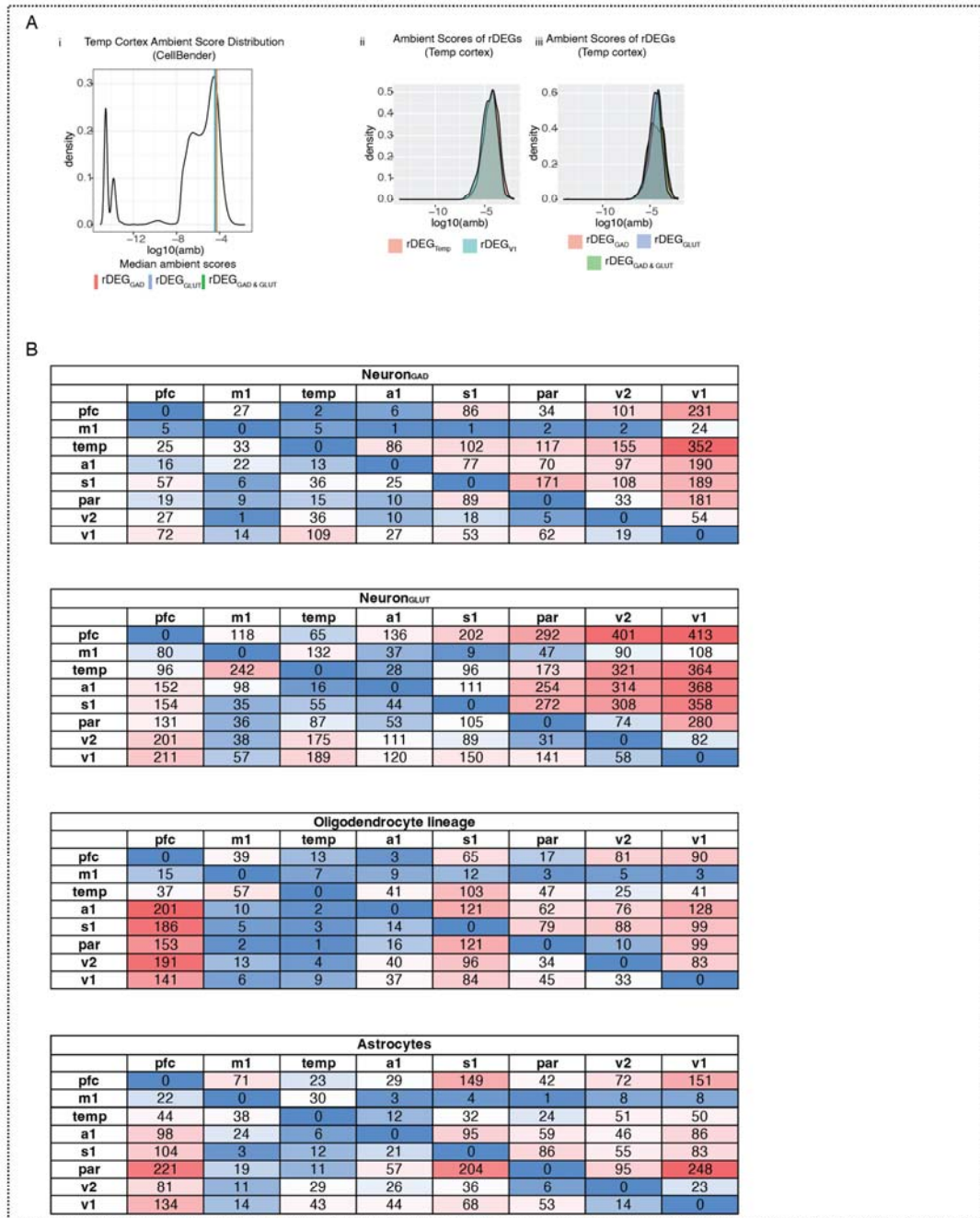1660 Colors indicate donor, same as (B). **(D)** Expression of marker genes in non-neurons. Genes as in (*8*).

1661 Heatmap colors are scaled to max normalized expression for each row (gene). Dendrogram and metadata

1662 colors as in **Fig. 1G**. **(E)** Differentially expressed genes in astrocytes across cephalic compartments. **(F)**

1663 Tissue validation for astrocyte differentially expressed genes (*FOXG1, SPARC*) in coronal sections of

1664 marmoset brain. Green dots indicate locations of cells that stain positive for *GFAP* (IHC, mAb **Table S4**)

1665 and *FOXG1* (smFISH). Magenta dots indicate cell positions for *GFAP* (IHC) and *SPARC* (smFISH). **(G)**

1666 Cartoon of coronal section imaged; Red boxes (*i-ii*) correspond to tissue validation in (H). **(H)** (*Left)*

1667 Fields of view from neocortex and thalamus stained for *GFAP* antibody (gray), *FOXG1* (green), and

1668 *SPARC* (magenta). Green arrows highlight *GFAP* cells colocalized with *FOXG1*, magenta arrows

1669 highlight *GFAP* cells colocalized with *SPARC*. (*Right*) Magnified examples of double positive cells in

1670 neocortex and thalamus. Ctx = cortex, Thal = thalamus. **(I)** (*Left*) Cartoon of sagittal section imaged; red

1671 boxes (*i-iii*) correspond to (*Right*) tissue validation of increased abundance of *VCAN*+ astrocytes in adult

1672 marmoset V1-adjacent white matter (*iii*) compared with PFC-adjacent white matter (*i*) and V1 gray

1673 matter (*ii*). GFAP antibody (gray) combined with smFISH probes against *VCAN* (magenta) and *GFAP*

1674 (green). Green arrows correspond to *GFAP*+ (antibody), *GFAP*+ (smFISH) cells. White arrows

1675 correspond to *GFAP*+ (antibody), *GFAP*+ (smFISH), and *VCAN*+ cells. V1 = visual cortex V1, PFC =

1676 prefrontal cortex, GM = gray matter, WM = white matter. Red arrow highlights locale of *VCAN*+ *GFAP*+

1677 images. Scale bar = 10 µm.

Supp S5 related to Fig 2

Neuron_GAD

	pfc	m1	temp	a1	s1	par	v2	v1
pfc	0	27	2	6	86	34	101	231
m1	5	0	5	1	1	2	2	24
temp	25	33	0	86	102	117	155	352
a1	16	22	13	0	77	70	97	190
s1	57	6	36	25	0	171	108	189
par	19	9	15	10	89	0	33	181
v2	27	1	36	10	18	5	0	54
v1	72	14	109	27	53	62	19	0

Neuron_GLUT

	pfc	m1	temp	a1	s1	par	v2	v1
pfc	0	118	65	136	202	292	401	413
m1	80	0	132	37	9	47	90	108
temp	96	242	0	28	96	173	321	364
a1	152	98	16	0	111	254	314	368
s1	154	35	55	44	0	272	308	358
par	131	36	87	53	105	0	74	280
v2	201	38	175	111	89	31	0	82
v1	211	57	189	120	150	141	58	0

Oligodendrocyte lineage

	pfc	m1	temp	a1	s1	par	v2	v1
pfc	0	39	13	3	65	17	81	90
m1	15	0	7	9	12	3	5	3
temp	37	57	0	41	103	47	25	41
a1	201	10	2	0	121	62	76	128
s1	186	5	3	14	0	79	88	99
par	153	2	1	16	121	0	10	99
v2	191	13	4	40	96	34	0	83
v1	141	6	9	37	84	45	33	0

Astrocytes

	pfc	m1	temp	a1	s1	par	v2	v1
pfc	0	71	23	29	149	42	72	151
m1	22	0	30	3	4	1	8	8
temp	44	38	0	12	32	24	51	50
a1	98	24	6	0	95	59	46	86
s1	104	3	12	21	0	86	55	83
par	221	19	11	57	204	0	95	248
v2	81	11	29	26	36	6	0	23
v1	134	14	43	44	68	53	14	0

1678

1679 **Figure S5. Cortical rDEGs are defined across cell types and do not reflect ambient RNA**

1680 **contamination. (A)** *(i)* rDEG ambient score distributions (from CellBender) in temporal cortex samples,

1681   which had amongst the highest numbers of rDEGs compared to other neocortical regions. Despite

1682   glutamatergic neurons being more numerous and having more expressed genes/transcripts per cell, the

1683   median ambient contamination scores for glutamatergic rDEGs were not higher than median

1684   contamination scores for GABAergic rDEGs. rDEGs shared between glutamatergic and GABAergic

1685   neurons had indistinguishable scores compared with rDEGs private to one neuronal class. *(ii)* Ambient

1686   scores in temporal cortex of temporal cortex rDEGs are indistinguishable from V1 rDEGs. *(iii)*

1687   Distributions of temporal cortex rDEG ambient scores by neuron class, again showing no difference

1688   between rDEGs that are shared or private to a neuronal class. **(B)** Numbers of regionally differentially

1689   expressed genes (rDEGs) between pairs of cortical regions for neurons, astrocytes, and oligodendrocyte

1690   lineage types.

FIGURE S6
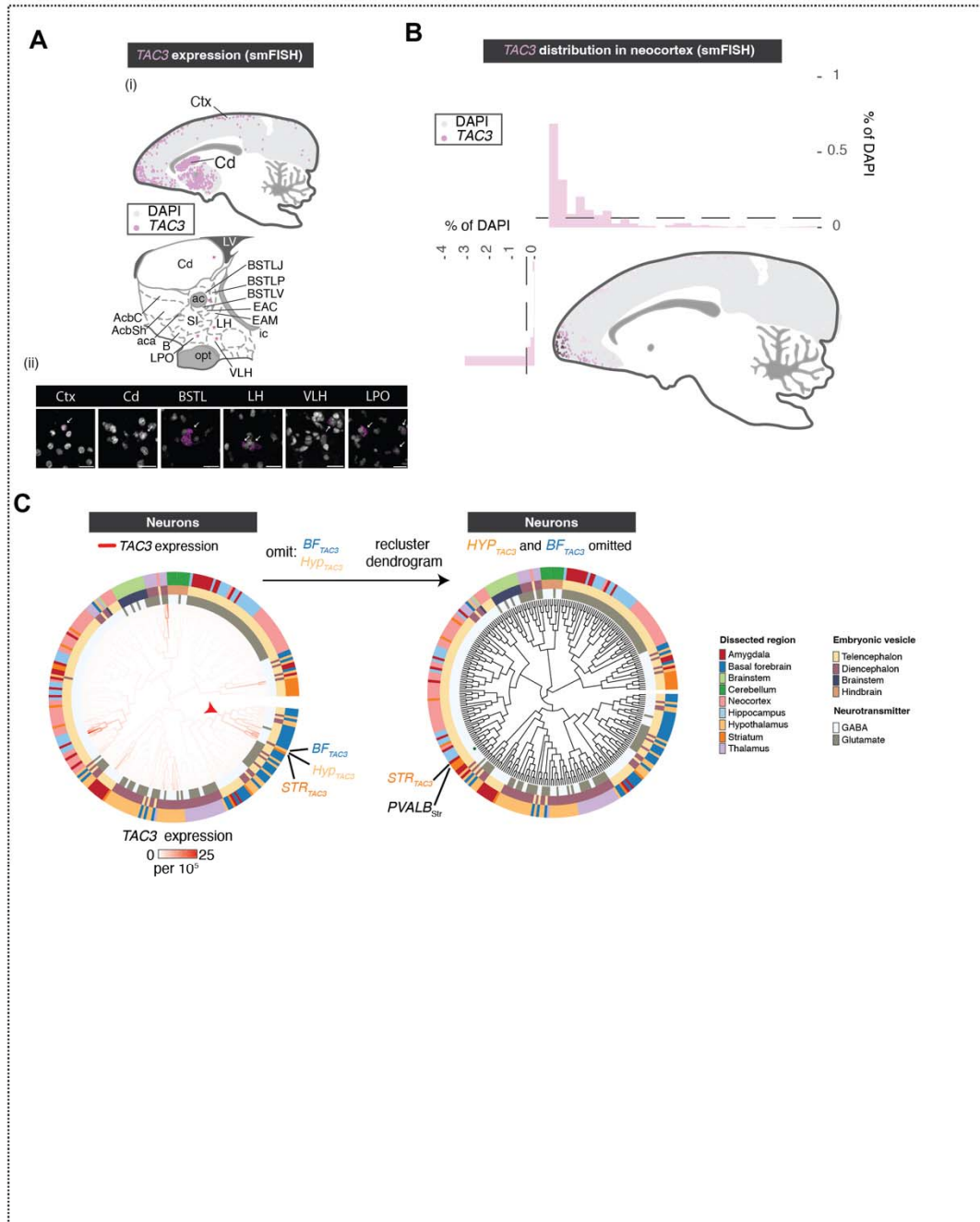
1691
Figure S6. Locations of *TAC3+* cells in marmoset forebrain. (A) smFISH reveals anatomical locations
and expression levels *TAC3+* types in different brain regions. *(i)* Schematic of *TAC3+* cells imaged across

1694   cortex, dorsal striatum. Ctx = neocortex, Cd = Caudate. *(ii)* Cartoon close-up of nuclei in striatum, basal
1695   forebrain and hypothalamus. Magenta stars = locations of *TAC3+* cells in lower image panel. Cd =
1696   Caudate, AcbC = nucleus accumbens core, AcbSh = nucleus accumbens shell, SI = Substantia
1697   innominata, B = basal nucleus of Meynert, EAM = extended amygdala, medial, EAC = extended
1698   amygdala, central, ac = anterior commissure, BSTLP = bed nuc st, lateral posterior, BSTLJ = bed nuc st,
1699   juxtacap, BSTLV = bed nuc st, lateral ventral, LH = lateral hypothalamus, VLH = ventrolateral
1700   hypothalamus, LPO = lateral preoptic area. **(B)** Density and location of *TAC3+* cells as proportion of all
1701   DAPI+ cells. Barplots show percentages in bins (approximately 1,290 µm per bin) taken across the
1702   anterior-posterior (top) and dorsal-ventral (left side) axes. **(C)** Effect on placement of the *TAC3+* striatal
1703   type on the neuronal dendrogram when omitting the two *TAC3+* types in hypothalamus and basal
1704   forebrain. When these types are omitted and hierarchical clustering is repeated (using HCA$_{PCA}$), the
1705   *TAC3+* striatal type is most similar to *PVALB+* striatal interneurons, consistent with previous reports that
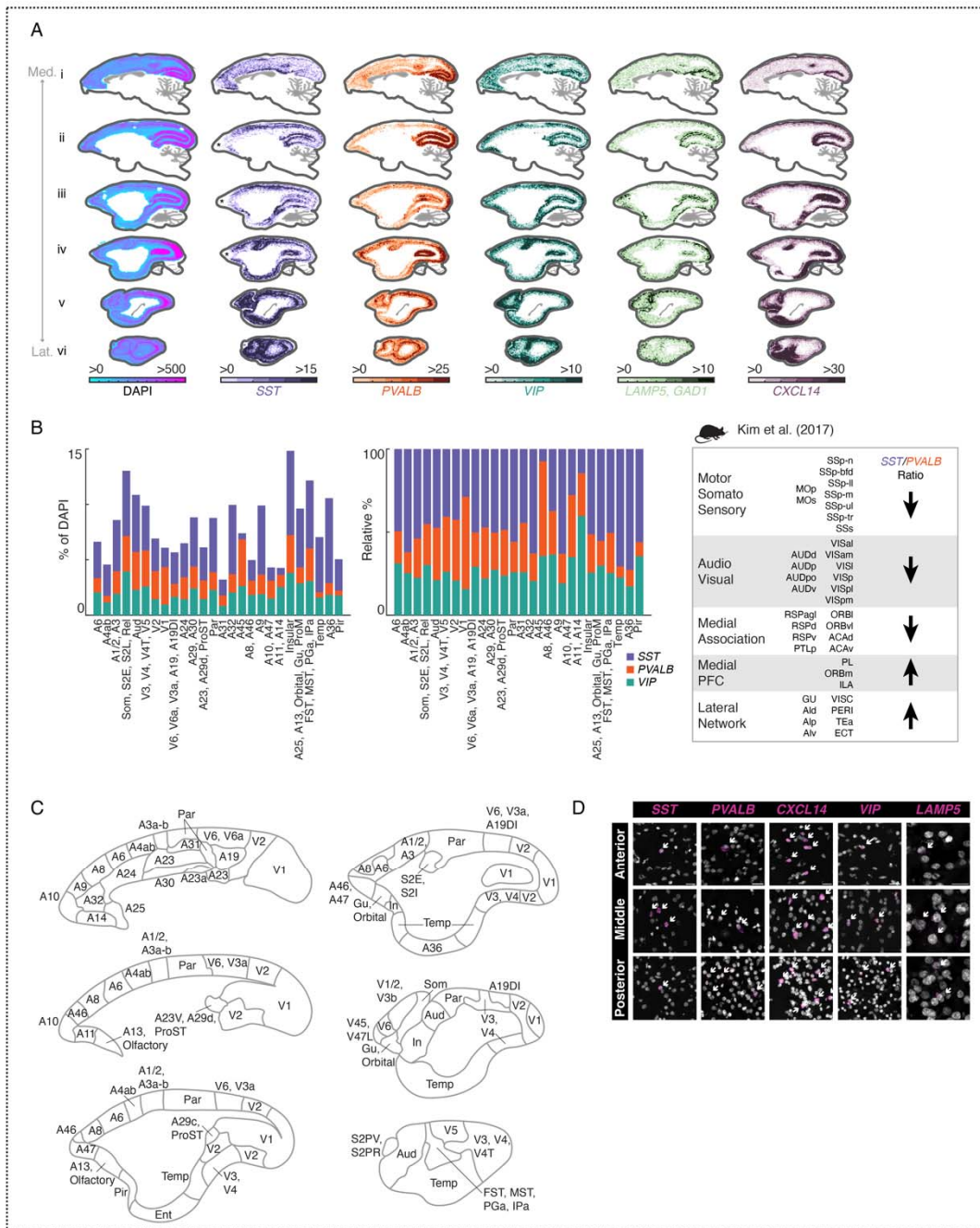1706   only compared telencephalic interneurons (*3*).
1707

FIGURE S7
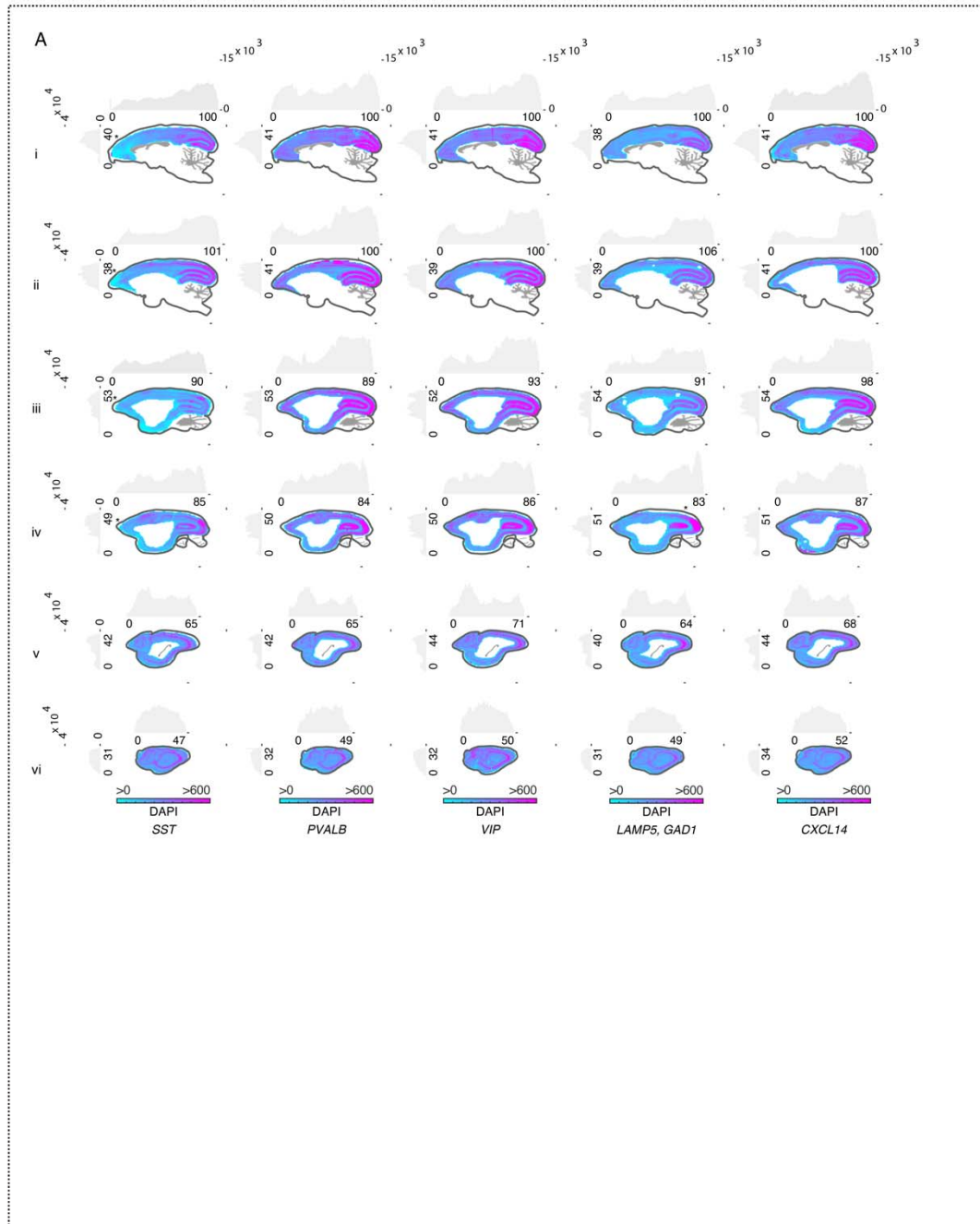
1708
Figure S7. Total cell numbers across marmoset neocortex. (A) Total numbers of DAPI+ cells per unit
1710 area (approximately 387 µm per bin) for each of the sections shown in **Fig. 4C**.

FIGURE S8

1713   **Figure S8. Interneuron numbers across marmoset neocortex. (A)** smFISH for neocortical interneuron
1714   subclass markers showing locations of cells positive for each marker across 6 sagittal sections of the
1715   marmoset neocortex. Heatmap scale shows absolute density per unit area (approximately 387 µm per
1716   bin). First column shows DAPI and area profiled. **(B)** (*Left, middle)* Quantitation of interneuron
1717   proportions by cortical area in marmoset parcellated according to Fig. S8C. *(Right)* Quantitation of
1718   interneuron proportions by cortical area in mouse reproduced from Kim et al. (2017). (*Left*) Absolute
1719   percentages of *SST, PVALB,* and *VIP* populations. (*Middle*) Same as (*left),* but scaled as proportions to
1720   100%. (*Right*) Schematic describing ratios of major interneuron types (*Sst+, Pvalb+*) in mouse from Kim
1721   et al. (2017). **(C)** Cartoons of cortical areas and areal groupings used to bin smFISH neocortical
1722   interneuron proportions in (B) and **Fig. 4D-E**. Neocortical parcellation from
1723   https://doi.org/10.24475/bma.4520. **(D)** Examples of smFISH images quantitated in **Fig. 4B-E**. Panels for
1724   each marker show example positive cells in anterior, middle, and posterior locations across neocortex.
1725   White arrows indicate positive cells. Scale bar = 20 µm.
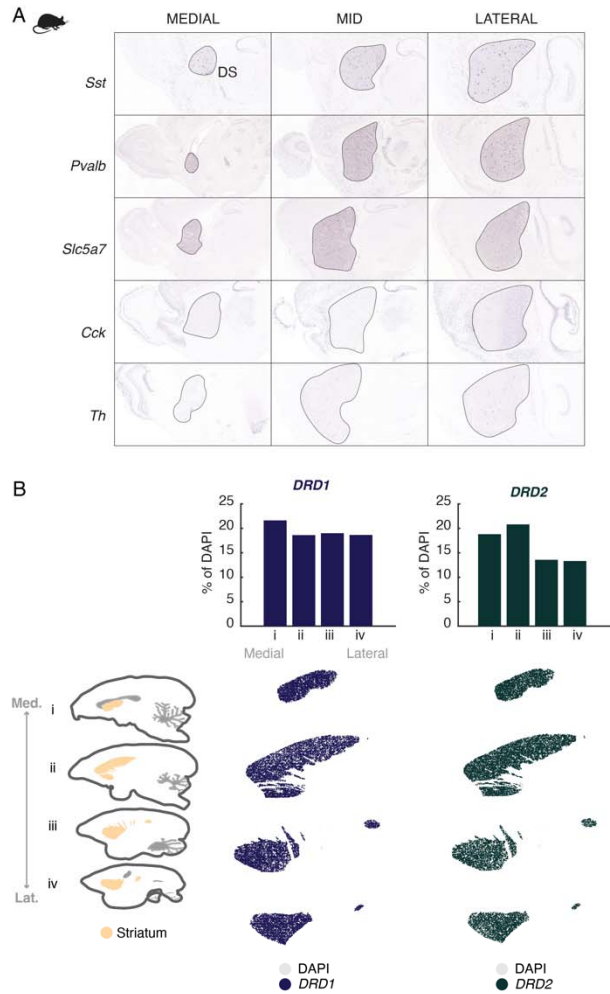
FIGURE S9

1726
Figure S9. Interneuron numbers across marmoset striatum. (A) Mouse striatal *in situ* from Allen
Brain Atlas for *Sst, Pvalb, Slc5a7, Cck, Th*. **(B)** Cartoon of marmoset striatum illustrates area profiled,

1729   medial to lateral. Proportions of *DRD1+* (dark blue) and *DRD2+* (dark green) cells across primate
1730   striatum calculated as in **Fig. 4H**.

FIGURE S10

1731
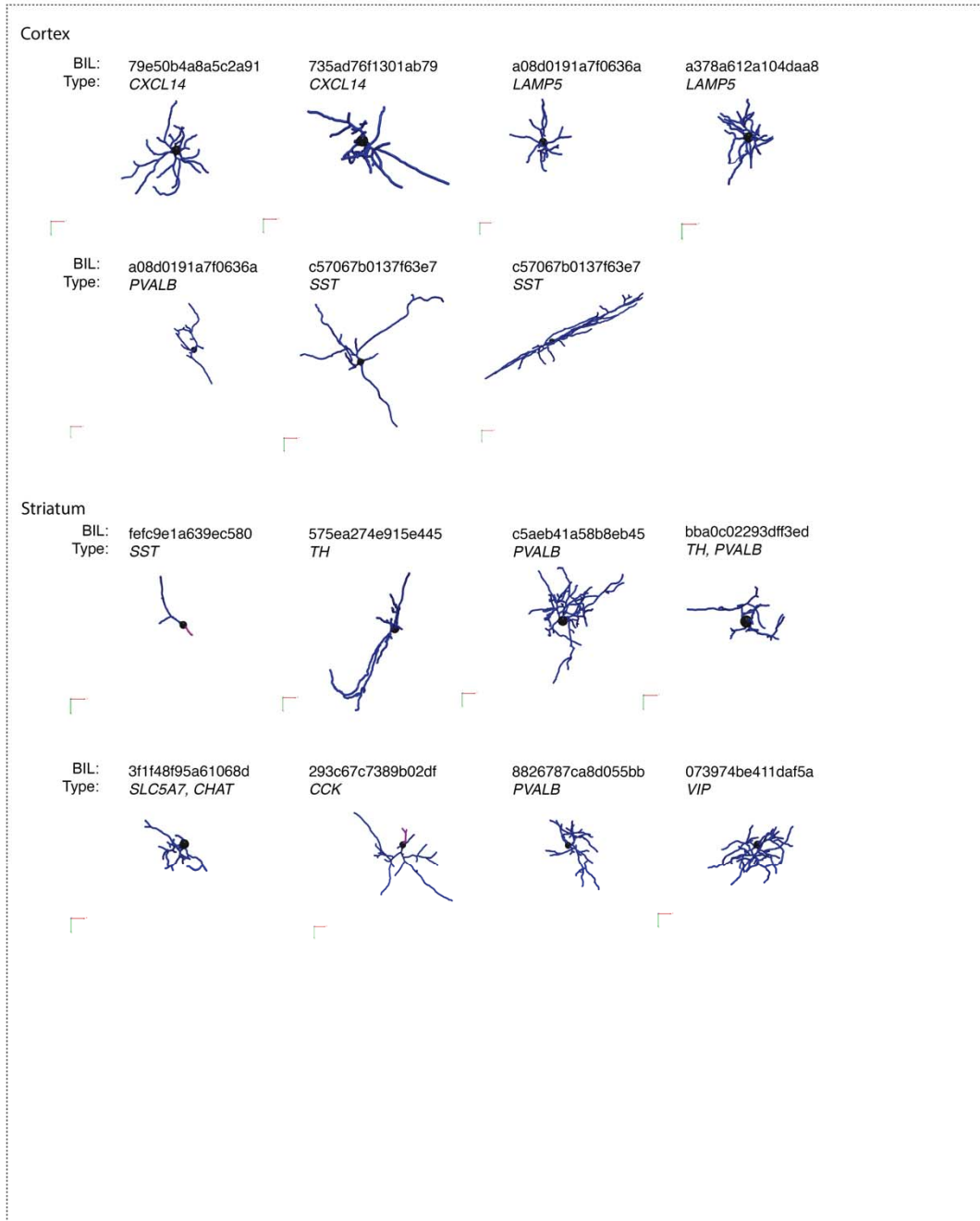
1732 **Figure S10. Morphology examples using NeuTube reconstructions. (A)** Example morphological
1733 reconstructions of striatal and neocortical interneurons using the NeuTube pipeline. Each cell, along with
1734 associated smFISH staining, is available for download at https://doi.org/10.35077/g.609.
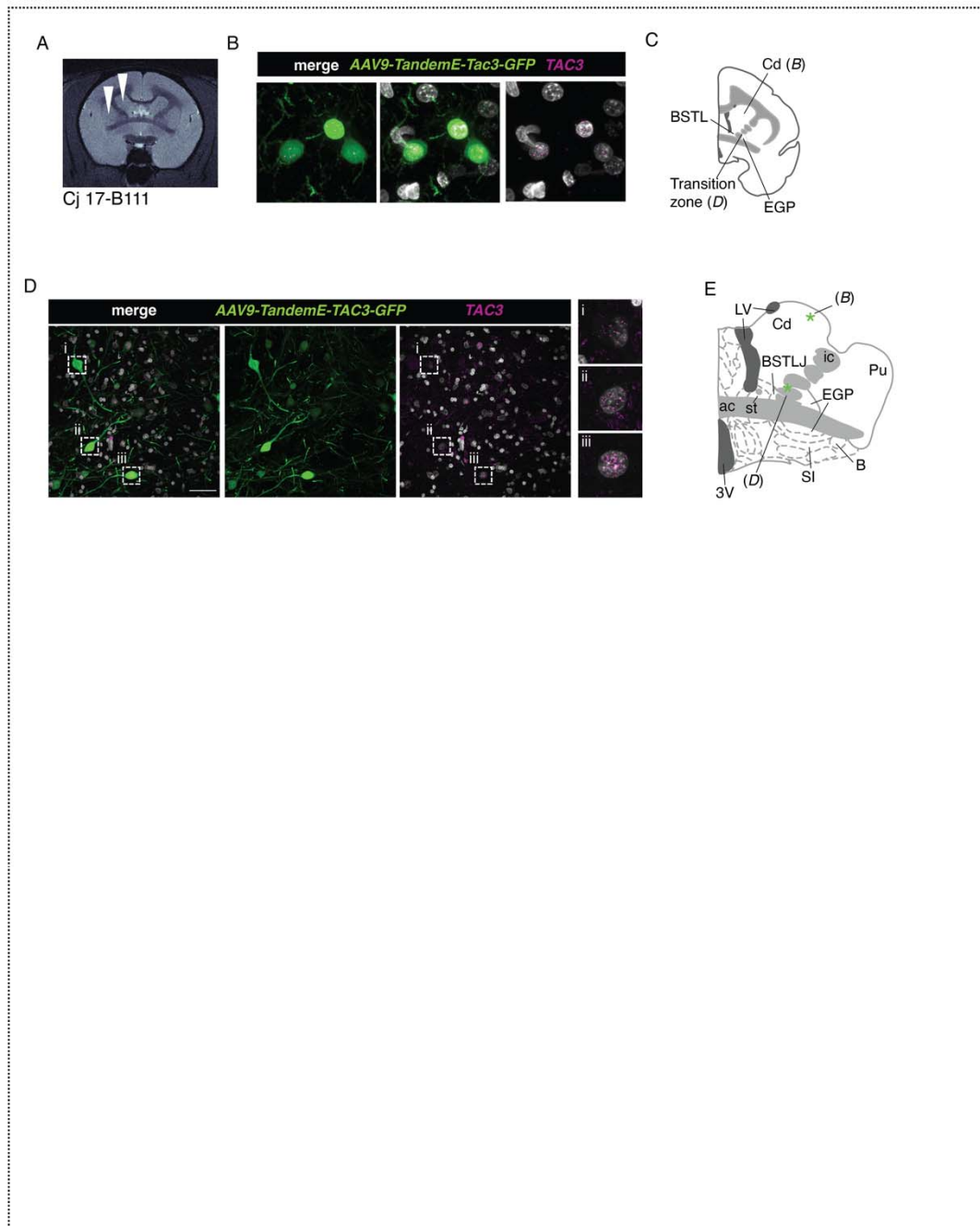
Figure S11

1735
Figure S11. Examples of striatal and peri-striatal *TAC3*+ neurons labeled by AAV-tandemE-TAC3-EGFP. (A) MRI showing injection location (white arrowheads) of virus into bilateral dorsal striatum

1738 (caudate) in one animal (Cj 17-B111). **(B)** EGFP antibody-amplified confocal image of a labeled cell
1739 (position shown in (C) with smFISH for *TAC3* showing colocalization). **(C)** Cartoon showing location of
1740 positive cells in (B) as well as labeled cells in transition zone (D). **(D)** Examples of extra-striatal labeled
1741 cells from injections in (A). **(E)** Position of cells in (D). Cd = Caudate, SI = Substantia innominata, B =
1742 basal nucleus of Meynert, ac = anterior commissure, BSTLP = bed nuc st, lateral posterior, BSTLJ = bed
1743 nuc st, juxtacap, 3V = third ventricle, LV = lateral ventricle.
1744
1745 **Table S1. Marmoset sample information.** Table of animals, metadata, and experimental information.
1746
1747 **Table S2. snRNA-seq dataset by donor and brain area.** Tables of the number of cells per brain
1748 structure and samples per donor.
1749
1750 **Table S3. Neocortical rDEGs across three donors.** Neocortical rDEGs across three donors with
1751 pairwise comparisons between neocortical locations for major clusters of cortical excitatory neurons,
1752 inhibitory neurons, astrocytes and oligodendrocyte lineage types.
1753
1754 **Table S4. Fluorescent *in situ* hybridization probes and antibodies used.** A list of all FISH probes
1755 (RNA-Scope and Molecular Instruments) and antibodies used for validation of gene and protein
1756 expression *in situ*.
1757
1758 **Table S5. Morphological reconstructions performed with Neutube.** A list of all cells reconstructed
1759 with Neutube with their corresponding morphological measurements.
1760
1761 **Table S6. Morphological reconstructions performed with Imaris.** A list of all cells reconstructed with
1762 Imaris with their corresponding morphological measurements.
1763
1764 **Table S7. Links and DOIs.** A list of all links and DOIs referenced.
1765
1766 **Data S1. Striatal and cortical medio-lateral quantification.** Tabular data for striatal and cortical
1767 medio-lateral histograms/bar plots.
1768
1769 **Data S2. *TAC3* medio-lateral quantification.** Tabular data for *TAC3* medio-lateral histograms and bar
1770 plots.