1    **Innovations in Primate Interneuron Repertoire**

2

3    Fenna M. Krienen[1,2], Melissa Goldman[1,2], Qiangge Zhang[2,3], Ricardo del Rosario[2], Marta Florio[1,2],
4    Robert Machold[4], Arpiar Saunders[1,2], Kirsten Levandowski[2,3], Heather Zaniewski[2,3], Benjamin
5    Schuman[4], Carolyn Wu[3], Alyssa Lutservitz[1,2], Christopher D. Mullally[1,2], Nora Reed[1,2], Elizabeth
6    Bien[1,2], Laura Bortolin[1,2], Marian Fernandez-Otero[2,5], Jessica Lin[2,5], Alec Wysoker[2], James
7    Nemesh[2], David Kulp[2], Monika Burns[3], Victor Tkachev[6], Richard Smith[6,7], Christopher A.
8    Walsh[6,7], Jordane Dimidschstein[2], Bernardo Rudy[4,8], Leslie Kean[6], Sabina Berretta[5,9,10], Gord
9    Fishell[2,5], Guoping Feng[2,3], Steven A. McCarroll[1,2]

10

11    [1] Department of Genetics, Harvard Medical School, Boston, MA 02115

12    [2] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA

13    02142, USA

14    [3] McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences,

15    Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

16    [4] NYU Neuroscience Institute, Langone Medical Center, New York, New York 10016, USA.

17    [5] Program in Neuroscience, Harvard Medical School, Boston, MA 02215, USA

18    [6] Department of Pediatrics, Boston Children's Hospital, Boston, MA 02115

19    [7] Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA 02115

20    [8] Department of Anesthesiology, Perioperative Care and Pain Medicine, New York University
21    School of Medicine, New York, NY 10016
22
23    [9] McLean Hospital, Belmont, MA 02478, USA

24    [10] Department of Psychiatry, Harvard Medical School, Boston, MA 02215, USA

25

26  **ABSTRACT**

27  Primates and rodents, which descended from a common ancestor more than 90 million years

28  ago, exhibit profound differences in behavior and cognitive capacity. Modifications,

29  specializations, and innovations to brain cell types may have occurred along each lineage. We

30  used Drop-seq to profile RNA expression in more than 184,000 individual telencephalic

31  interneurons from humans, macaques, marmosets, and mice. Conserved interneuron types

32  varied significantly in abundance and RNA expression between mice and primates, but varied

33  much more modestly among primates. In adult primates, the expression patterns of dozens of

34  genes exhibited spatial expression gradients among neocortical interneurons, suggesting that

35  adult neocortical interneurons are imprinted by their local cortical context. In addition, we found

36  that an interneuron type previously associated with the mouse hippocampus—the "ivy cell", which

37  has neurogliaform characteristics—has become abundant across the neocortex of humans,

38  macaques, and marmosets. The most striking innovation was subcortical: we identified an

39  abundant striatal interneuron type in primates that had no molecularly homologous cell population

40  in mouse striatum, cortex, thalamus, or hippocampus. These interneurons, which expressed a

41  unique combination of transcription factors, receptors, and neuropeptides, including the

42  neuropeptide *TAC3*, constituted almost 30% of striatal interneurons in marmosets and humans.

43  Understanding how gene and cell-type attributes changed or persisted over the evolutionary

44  divergence of primates and rodents will guide the choice of models for human brain disorders and

45  mutations and help to identify the cellular substrates of expanded cognition in humans and other

46  primates.

47

48  **INTRODUCTION**

49

50  Vertebrate brains contain many specialized brain structures, each with its own evolutionary

51  history. For example, the six-layer neocortex arose in mammals around 200 million years ago[1],

52  whereas distinct basal ganglia nuclei were already present in the last common ancestor of

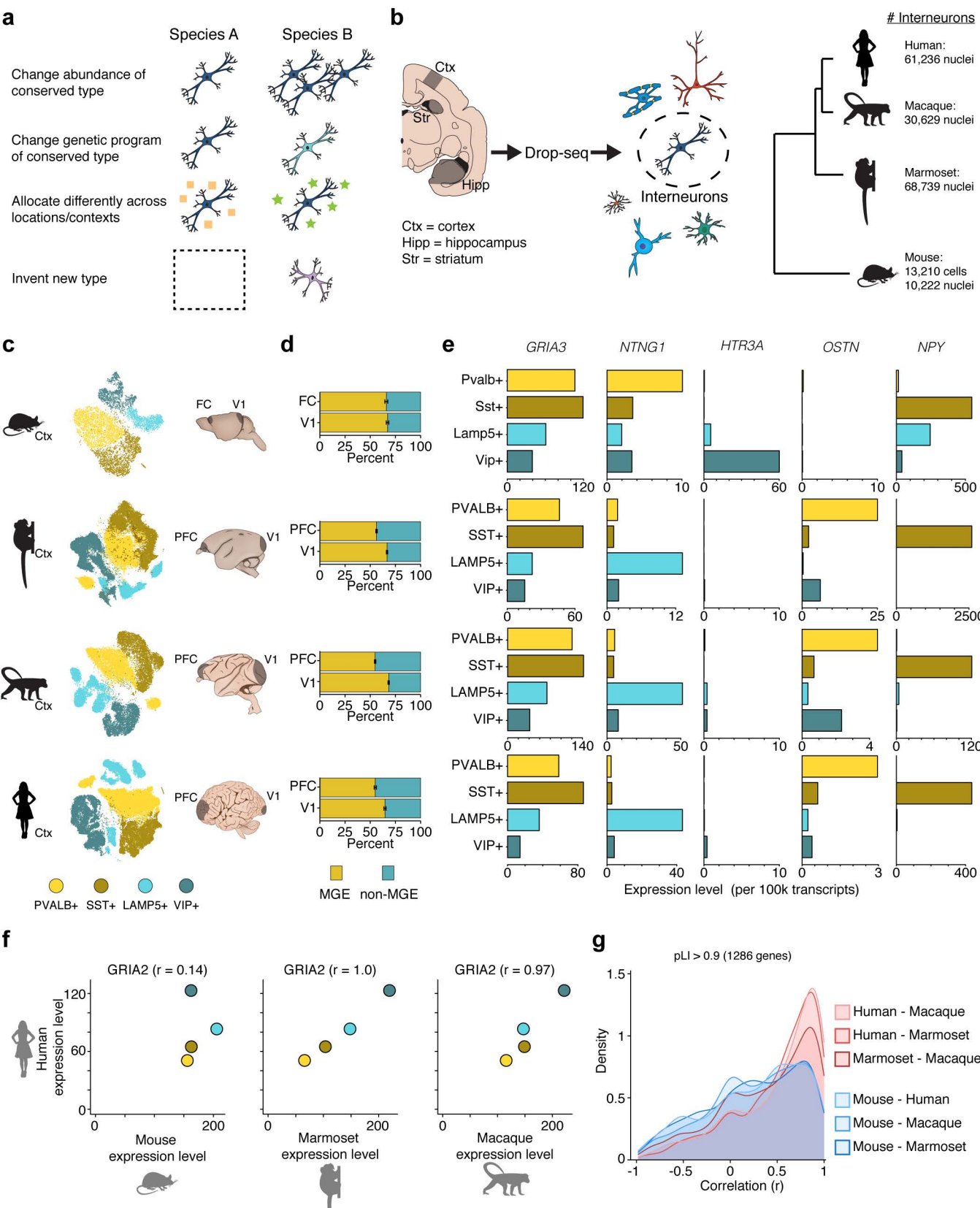53  vertebrates more than 500 million years ago[2].

54

55  Brain structures, circuits, and cell types have acquired adaptations and new functions along

56  specific evolutionary lineages. Numerous examples of modifications to specific cell types within

57  larger conserved brain systems have been discovered, including hindbrain circuits that control

58  species-specific courtship calls in frogs[3], the evolution of trichromatic vision in primates[4], and

59  neurons that have converted from motor to sensory processing to produce a novel swimming

60  behavior in sand crabs[5]. Evolution can modify brain structures through a wide range of

61  mechanisms, including increasing or reducing production of cells of a given type, altering the

62  molecular and cellular properties of shared cell types, reallocating or redeploying cell types to

63  new locations in the brain, or inventing entirely new cell types (Fig. 1a).

64

65  Single-cell RNA sequencing, which systematically measures gene expression in thousands of

66  individual cells, has recently enabled detailed comparisons of cell types and expression patterns

67  between homologous brain structures separated by millions of years of evolution[4,6,7] (non-single

68  cell approaches have also yielded important insights in this domain, e.g. [8]). For example, one

69  recent study compared neocortical cells between humans and mice, identifying conserved and

70  diverged features of many cell types[7]. It is not yet known which of the many differences between

71  mouse and human brains are specific to humans, and which are shared among primates.

72

# Figure 1



**a**

Change abundance of conserved type

Change genetic program of conserved type

Allocate differently across locations/contexts

Invent new type

Species A    Species B

**b**

Ctx
Str
Hipp

→ Drop-seq →

Interneurons

Ctx = cortex
Hipp = hippocampus
Str = striatum

# Interneurons

Human: 61,236 nuclei
Macaque: 30,629 nuclei
Marmoset: 68,739 nuclei
Mouse: 13,210 cells 10,222 nuclei

**c**

Ctx    FC V1
Ctx    PFC V1
Ctx    PFC V1
Ctx    PFC V1

● PVALB+  ● SST+  ● LAMP5+  ● VIP+

**d**

FC
V1
Percent
0 25 50 75 100

PFC
V1
Percent
0 25 50 75 100

PFC
V1
Percent
0 25 50 75 100

PFC
V1
Percent
0 25 50 75 100

■ MGE  ■ non-MGE

**e**

GRIA3    NTNG1    HTR3A    OSTN    NPY

Pvalb+
Sst+
Lamp5+
Vip+
0 ... 120    0 ... 10    0 ... 60    0 ... 10    0 ... 500

PVALB+
SST+
LAMP5+
VIP+
0 ... 60    0 ... 12    0 ... 10    0 ... 25    0 ... 2500

PVALB+
SST+
LAMP5+
VIP+
0 ... 140    0 ... 50    0 ... 10    0 ... 4    0 ... 120

PVALB+
SST+
LAMP5+
VIP+
0 ... 80    0 ... 40    0 ... 10    0 ... 3    0 ... 400

Expression level  (per 100k transcripts)

**f**

GRIA2 (r = 0.14)
Human expression level
120
60
0
0    200
Mouse expression level

GRIA2 (r = 1.0)
0    200
Marmoset expression level

GRIA2 (r = 0.97)
0    200
Macaque expression level

**g**

pLI > 0.9 (1286 genes)

Density
1.5
1.0
0.5
0

Correlation (r)
-1    -0.5    0    0.5    1

Human - Macaque
Human - Marmoset
Marmoset - Macaque

Mouse - Human
Mouse - Macaque
Mouse - Marmoset

73

**Figure 1. Analysis of cortical interneurons in mouse, marmoset, macaque, and human. a,** Schematic showing possible modes of change in cellular assemblies across species. **b,** Schematic of experimental workflow and numbers of interneurons sampled in each species. Region abbreviations: Ctx = neocortex, Hipp = hippocampus, Str = striatum. **c,** t-distributed stochastic neighbor embedding (t-SNE) of cortical interneurons in each species. Cells are colored by membership in one of the four major neocortical classes: *SST+, PVALB+, VIP+,* or *LAMP5+* (dark brown or dark green cells represent the minority of cells that co-expressed *SST* and *PVALB* or *LAMP5* and *VIP*, respectively)*.* **d,** The proportion of medial ganglionic eminence (MGE)-derived (*SST+* or *PVALB+*) and non-MGE–derived (*VIP+* or *LAMP5+*) types across two cortical regions (frontal/prefrontal cortex, PFC, and visual cortex, V1) for each species. Error bars represent 95% binomial confidence intervals. **e,** Examples of markers with different enrichment patterns across species (see also Extended Data Fig. 1b). Values are scaled expression levels (number of transcripts per 100k) for each of the four main cortical interneuron classes. **f,** Scaled expression levels (number of transcripts per 100k) for *GRIA2*, a gene encoding an AMPA receptor subunit, in human vs. mouse, marmoset, and macaque for the four major interneuron classes. Dots are colored as in (**e**). **g,** Density histograms showing correlation distribution of expressed genes between pairs of species (red = primate–primate pairs, blue = primate–mouse pairs). pLI = probability of loss of function intolerance.

92

In this study, we compared interneurons, a major class of neurons present in all vertebrates, in mice and three primates: marmoset, macaque, and human, which span ~90 million years of evolutionary divergence (Fig. 1b). Interneurons contribute to local circuit assemblies and provide the main source of inhibition in neuronal circuits by releasing the inhibitory neurotransmitter GABA. Interneurons are born subcortically from progenitors that reside in transient proliferative zones called the ganglionic eminences, including the medial and caudal ganglionic eminences (MGE and CGE), and migrate to the neocortex and to subcortical structures during development[9]. Interneurons are particularly interesting for comparative analysis because they are morphologically and physiologically diverse within any one species, but major types are shared across the amniotes[10]. In mice, the same interneuron types recur across functionally distinct neocortical regions[11,12]. An understanding of interneurons' evolution in primates and rodents could guide the choice of models for studying how microcircuits and excitatory/inhibitory balance are affected in human brain disorders. Moreover, although the main developmental origins for

4

106   interneurons appear conserved, we still do not know how interneurons are qualitatively and

107   quantitatively allotted to their destinations, nor the extent to which local cues shape interneuron

108   gene expression in different species.

109

110   **Identifying interneurons in mice and primates**

111   We used Drop-seq[13], a single-cell RNA sequencing technology, to measure RNA expression in

112   nuclei of telencephalic brain cells (i.e., cells from brain regions including neocortex, hippocampus,

113   and striatum) from adult animals of four species: mouse, common marmoset, rhesus macaque,

114   and human. By applying unsupervised methods to the data from each individual species and from

115   combinations of species, we classified transcriptionally distinct and similar groups of cells. We

116   identified interneurons using canonical, conserved markers (e.g., *GAD1* and *GAD2,* which encode

117   the glutamate decarboxylase required for synthesis of GABA) as well as class-specific molecular

118   markers. In total, we sampled 68,739 telencephalic interneurons from marmoset, 61,236 from

119   human, 30,629 from macaque, and 23,432 from mouse.

120

121   It was not previously known whether the broad classes of interneurons identified in mouse[11,12,14]

122   also exist in all three primates, and if so, whether they could be delineated with the same set of

123   markers. Across all four species, the same four genes (*SST, PVALB, VIP,* and *LAMP5*) exhibited

124   mutually exclusive expression while together accounting for almost 100% of neocortical

125   interneurons, suggesting that these markers stably delineate a core repertoire of interneuron

126   types among rodents and primates (Fig. 1c; see also[7]).

127

128   **Interneuron abundances and local specialization in neocortex**

129   Within conserved brain structures, evolutionary changes in cell numbers or proportions can have

130   profound functional consequences; such alterations appear to be major effectors of brain

131   evolution[5,15,16]. Mammalian neocortex is patterned into functionally specialized fields, called

132    areas, that differ in cytoarchitecture, cell number, and connectivity. There is a fundamental

133    distinction between primary sensory areas of the neocortex, which process visual, auditory and

134    tactile information as part of well-defined hierarchies, and association areas such as prefrontal

135    cortex, which perform higher-order functions. Primates, and particularly humans, have

136    neocortices that are disproportionately enlarged relative to those of other mammals[17]. The

137    accompanying changes in cellular composition are not well understood, although recent

138    quantitative stereological methods have begun to relate cortical specialization to cell-type

139    composition across the neocortex[16].

140

141    In primates, but not mice, frontal association areas (FC/PFC) differed from primary visual cortex

142    (V1) in the extent to which interneurons were derived from MGE as opposed to other eminences

143    (Fig. 1d). In all four species, V1 contained similar proportions of MGE (marked by *SST* or *PVALB*)

144    and non-MGE interneurons (marked by *VIP* or *LAMP5)*: ~66% MGE and ~34% non-MGE.  In

145    primates, however, PFC harbored a significantly higher proportion of non-MGE interneurons

146    (~55% MGE, ~45% non-MGE; Fig. 1c). Like PFC, association areas in temporal and parietal

147    cortex contained proportionally more non–MGE-derived interneurons than V1 (Extended Data

148    Fig. 1a). In primates, the upper neocortical layers have enlarged, particularly in association

149    cortex[18]. Because MGE-derived interneurons preferentially populate deep layers[19], the

150    proportional increase in non–MGE-derived interneurons in primates is consistent with

151    enlargement of upper-layer neocortical compartments and suggests greater recruitment of

152    interneurons from the CGE to association cortex in primates.

153

154    **Genetic programs within conserved interneuron types**

155    Homologous cell types can acquire species-specific functions through changes in gene

156    expression[9,16,17].  To evaluate the extent to which gene-expression specializations distinguishing

157    interneuron types are shared across species, we compared the expression level of each gene

158    across the four principal interneuron classes (*PVALB+*, *SST+*, *LAMP5+*, *VIP+*) within each

159    species, and then compared these class-specific relative expression profiles across the species

160    surveyed. (Comparing within and then across species corrects for species-specific [e.g.

161    sequence-related] influences on mRNA sampling, as well as for latent technical variables that

162    might distinguish brains from different species.) This analysis enabled identification of conserved

163    gene expression patterns. For example, *GRIA3*, which encodes an AMPA receptor subunit, was

164    expressed in similar patterns across the four classes (*SST+ > PVALB+ > ID2+ > VIP+*) in all four

165    species (Fig. 1e).

166

167    We first focused on genes that were selectively expressed at least one interneuron type (relative

168    to the others) in at least one species (examples in Fig. 1e and Extended Data Fig. 1b). A clear

169    pattern emerged: the great majority of human–mouse gene expression differences were shared

170    among all three primates.  For example, the neuropeptide Y (*NPY*) gene, a commonly used

171    marker for specific interneurons, was expressed in both *SST+* and *LAMP5+* interneurons in

172    mouse but was selectively expressed in *SST+* interneurons in marmoset, macaque, and humans.

173    The netrin G1 (*NTNG1*) gene was, in mouse neocortex, selectively expressed in *PVALB+*

174    interneurons; in primate neocortex, *NTNG1* was instead expressed by *LAMP5+* interneurons. In

175    rare cases, a gene was enriched as a specific marker in one cell class in primates (e.g. *OSTN* in

176    *PVALB+* interneurons) but not detected at all in mouse interneurons[20], or vice versa (e.g., *HTR3A*,

177    which encodes serotonin receptor 3a, see also[7]). Other examples included synuclein gamma

178    (*Sncg*), the short transient receptor potential channel 3 (*Trpc3*), and the IQ motif containing

179    GTPase activating protein 2 (*Iqgap2*), which were expressed specifically in certain classes of

180    interneurons in mice, but were either widely expressed (*SNCG*) or enriched in a different

181    population (*TRPC3, IQGAP2*) among neocortical interneurons from primates (Extended Data Fig.

182    1b). Such cross-species expression variation has implications for choosing selective markers to

183    define or characterize conserved cell types.

184

185    Far more genes were expressed in many or all neuronal types at quantitatively distinct levels. We

186    know little about the extent to which the precise expression levels of such genes contribute to the

187    specialized functions of neuronal types. However, understanding such relationships will be

188    important for interpreting the significance of noncoding genetic variation in humans, as well as

189    for selecting appropriate models for heterozygous mutations ascertained in human patients. A

190    comparative lens can reveal the extent to which evolution has maintained a gene's quantitative

191    expression level in different cell types relative to one another. To evaluate conservation at this

192    level, we identified 4051 expressed genes that exhibited at least 1.5-fold expression variation

193    among the four main interneuron classes, and calculated the cross-species correlation of each

194    gene's expression measurements across those classes. Illustrating one of the main patterns

195    revealed by this analysis, expression levels of *GRIA2*, a member of an AMPA receptor subunit

196    family, exhibited relatively little variation (± 25%) among the four main interneuron classes in mice,

197    but varied by 2–3-fold across the homologous interneuron types in primates, in a pattern (*VIP+ >*

198    *ID2+* and *SST+ > PVALB+)* that was highly correlated across the three primates (r = 0.95–1.0;

199    Fig. 1f).

200

201    Genes that are dosage-sensitive in humans might have particularly strong evolutionary constraint

202    on their expression levels. To evaluate this, we further focused on 1,286 genes that exhibit

203    evidence of haploinsufficiency in humans as determined using pLI, a metric based on sequence

204    variation across 60,706 human genomes that describes the probability that a given gene is

205    intolerant of loss of function in human populations[21]. Relative expression of these human-

206    haploinsufficient genes (pLI > 0.9) in interneuron types appeared to be much more constrained

207    among primates than in the primate–mouse comparisons (Fig. 1g). (A control analysis revealed

208    that the subset of genes with low pLI scores (pLI < 0.1) had correlation values around 0 for all

209    species pairs, implying that genes that are tolerant of loss of function in humans are less

210  constrained in their expression levels.)  This relationship suggests that even dosage-sensitive

211  genes have undergone substantial evolutionary change in their quantitative expression levels,

212  and that these levels are more similar among primates than between primates and mice.

213

214  We also compared pairs of cell types in each pair of species, evaluating the extent to which

215  differential-expression relationships were conserved between species across genes that are

216  meaningfully expressed (> 10 transcripts per 100,000) in interneurons (Extended Data Fig. 2).

217  Such comparisons offered abundant evidence that the relative expression levels (between

218  interneuron subtypes) of a vast number of mutually expressed genes have been conserved. The

219  overall correlation of relative expression levels between cell-type pairs was stronger for

220  comparisons between primates than for comparisons between mice and any of the primates

221  (Extended Data Fig. 2c).

222

223  **Regional specialization of expression patterns within the neocortex**

224  To resolve types of interneurons at a finer scale, and to compare these types across species and

225  brain regions, we used a computational approach, LIGER[22], that aligns expression patterns

226  across experiments and species (Extended Data Fig. 2), enabling us to compare gene expression

227  programs within interneuron types across cortical regions (Fig. 2). In mouse, the expression

228  programs of interneurons, in contrast to those of excitatory neurons, exhibit few differences across

229  cortical locations[11,12,23]. In marmosets, each of 17 readily resolvable interneuron types was

230  present in all seven cortical regions surveyed, confirming that, as in mouse, different cortical

231  regions contain the same basic interneuron types (Fig. 2a). However, gene expression patterns

232  for these conserved types differed across cortical regions. Across types, the median number of

233  regionally differentially expressed genes (rDEGs, >3-fold difference) between PFC and V1 was

234  55 (Fig. 2b), exceeding the number observed when comparing mouse frontal and posterior

235  cortical areas using the same criteria (median = 12.5, see also[11,12]).
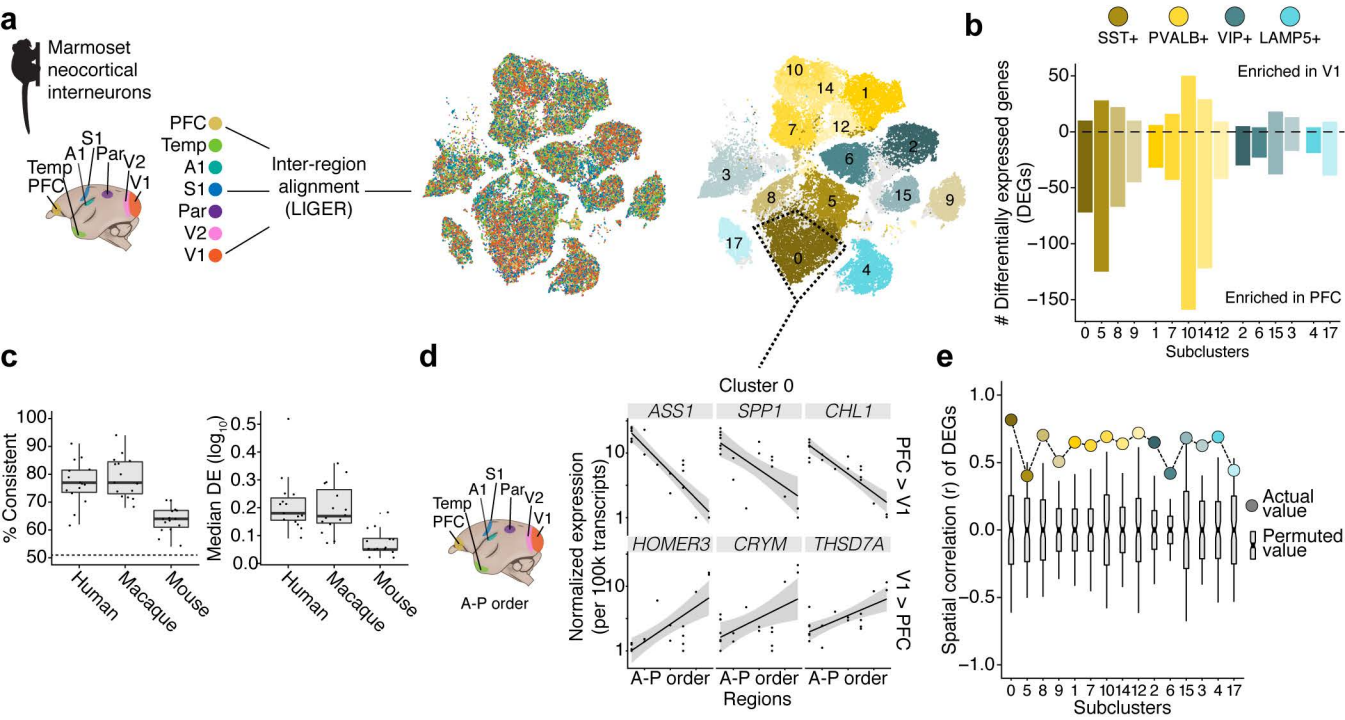
236

237    Although the number of rDEGs varied across clusters (Fig. 2b and Extended Data Fig. 3b), rDEGs

238    identified for any one cluster or type tended to exhibit the same regional bias in the other clusters

239    as well (Extended Data Fig. 3c,d). This suggests that most such differences reflect a common

240    regional signature that is shared by diverse interneurons, rather than being specific to particular

241    interneuron types. This regional bias did not extend to astrocytes (Extended Data Fig. 3e). The

242    specific rDEGs defined in marmosets exhibited shared patterns of regional bias in interneurons

243    in the other species: genes that were enriched in PFC vs V1 in marmoset were more likely to be

244    more highly expressed in PFC than in V1 in the other three species, with greater probability and

245    magnitude of difference in humans and macaques than in mice (Fig. 2c). These results suggest

246    that interneurons acquire region-specific components of their molecular identities (see also[24,25]),

247    and that these details are shared across species and most strongly among close relatives.

248

249

# Figure 2



**a** Marmoset neocortical interneurons

Inter-region alignment (LIGER)

PFC, Temp, A1, S1, Par, V2, V1

**b**
SST+ PVALB+ VIP+ LAMP5+

Enriched in V1 / Enriched in PFC

# Differentially expressed genes (DEGs)

Subclusters: 0 5 8 9 1 7 10 14 12 2 6 15 3 4 17

**c** % Consistent — Human, Macaque, Mouse

Median DE (log₁₀) — Human, Macaque, Mouse

**d** A-P order

Cluster 0

ASS1, SPP1, CHL1 — PFC > V1

HOMER3, CRYM, THSD7A — V1 > PFC

Normalized expression (per 100k transcripts)

A-P order / Regions

**e** Spatial correlation (r) of DEGs

Actual value / Permuted value

Subclusters: 0 5 8 9 1 7 10 14 12 2 6 15 3 4 17

**Figure 2. Comparing cortical interneurons within and across species. a,** Integrative analysis of seven neocortical regions in marmoset using LIGER. t-SNE plots show resultant clusters with cells colored by cortical region of origin (left) and by cluster assignment (right). **b,** Histogram of the number of regionally differentially expressed genes (rDEGs) (>3-fold expression difference) between prefrontal (PFC) and primary visual cortex (V1) in each cell type (cluster) for which there were at least 50 cells per region. **c,** The rDEGs in marmoset (PFC, V1) tended to have the same regional differences in the other three species (left, percent of genes consistent with marmoset pattern; dashed line represents chance), with greater differential expression in humans and macaques than in mice (right). Dots represent rDEGs from each marmoset cluster; values were calculated from the cluster in the other species that had the most DE genes in common with that marmoset cluster. **d,** Normalized expression of rDEGs (identified between PFC and V1) across all seven neocortical regions in marmosets. X-axis arranged by anterior–posterior position of neocortical region. Plot shows the top three differentially expressed genes for the cluster outlined in **a** for each contrast (PFC>V1, V1>PFC). Dots are individual replicates within each region. **e,** Colored dots show averaged spatial correlations across rDEGs identified in each cluster when regions (n = 5, excluding PFC and V1) are arranged in anterior–posterior order. Gray boxplots show averaged correlations of the same rDEGs in each cluster when computed using permuted region orderings (n=120 possible orderings).

Spatial patterns of gene expression, including macro-scale gradients and the distinction between primary and higher-order neocortical areas, configure the layout of neocortical areas during development[26,27] and persist into adulthood[28]. Gradients may contribute to excitatory neuron diversity[29]. Unlike excitatory neurons, which are born just below the neocortical areas in which they ultimately reside, neocortical interneurons are born subcortically and migrate into the neocortex post-mitotically; little is known about whether individual interneurons acquire information or specializations reflecting their ultimate areal locations within the neocortex. To explore this question, we investigated whether the expression of rDEGs identified in comparisons of PFC and V1 also varied across other neocortical regions. This analysis revealed a spatial logic: rDEG expression correlated strongly with anterior–posterior location (Fig. 2d); a control analysis in which region order was permuted yielded correlations distributed around zero (Fig. 2e). This anterior–posterior gradient is also correlated with, and might well reflect, the more complex patterns associated with the distinction between primary and higher-order neocortical areas[30,31]; for example, a number of genes had expression levels in parietal association cortex that were

11

283    more similar to those in temporal and prefrontal cortex than those in the sensory areas more

284    proximal to it. (Definitively parsing the overlapping effects of anterior-posterior, sensory-

285    association, and other topographies would require a comprehensive sampling of neocortical

286    areas.) Together, these results suggest that neocortical interneurons intrinsically detect and

287    encode some aspect of their ultimate spatial position.

288

289    **Re-allocation of a shared interneuron type across brain structures**

290    The main classes (*PVALB, SST, VIP, LAMP5*) of neocortical interneurons each contain many

291    types[32], and our analysis of the mouse and marmoset cortical interneurons (using LIGER)

292    identified at least 15 transcriptionally distinct types (Extended Data Fig. 2b) with clear cross-

293    species homologies in their global patterns of gene expression (Extended Data Fig. 4).  These

294    analyses affirm findings that homologous, molecularly-defined interneuron types can be identified

295    across species spanning vast evolutionary distances, including reptiles, mice, and humans[7,33].

296    We have not attempted a definitive taxonomic classification here because we anticipate that

297    improvements in single-cell technology and deeper ascertainment of neurons will further refine

298    these categories.

299

300    The broad sharing of interneuron types, though, included notable differences. Notably, marmoset

301    neocortex contained a substantial population of *LAMP5+* cells that co-expressed *LHX6* (Fig. 3a)*.*

302    The existence of these interneurons in the neocortex raised intriguing questions because *LHX6*,

303    a transcription factor, participates in cell fate determination of MGE types, whereas *LAMP5+*

304    neocortical interneurons come from the CGE[34]. In mouse, neocortical *Lamp5+* interneurons

305    consist of neurogliaform and single-bouquet types, which are the most numerous type of Layer 1

306    neuron and have distinct morphological, neurochemical, and connectivity properties[9]. Analysis by

307    smFISH revealed that the spatial distribution of primate *LAMP5+/LHX6+* neurons was distinct

308    from that of *LAMP5+/LHX6-* neurons, with the former tending to reside within the deep cortical

12

309    layers (Fig. 3b). The proportion of interneurons that were *LAMP5+/LHX6+* was 10-fold higher in

310    marmoset, macaque, and human cortex than in mouse cortex, in all cortical regions analyzed

311    (Fig. 3c). Thus, this cell type, recently reported to be much more abundant in human temporal

312    lobe than in mouse primary visual cortex[7], appears to have expanded throughout the neocortex

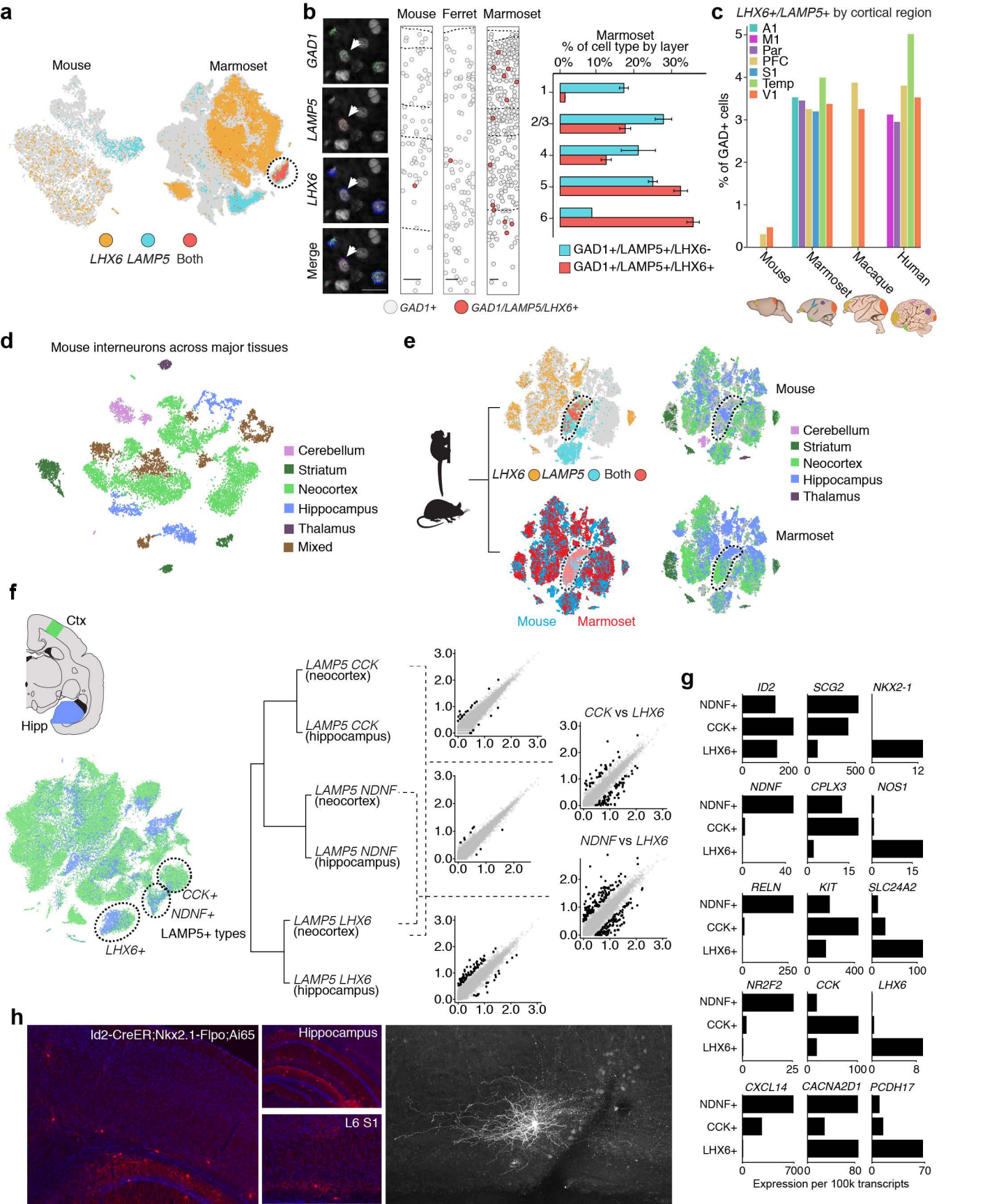313    in an ancestor of diverse primates.

314

315    The *LAMP5/LHX6+* neurons could represent an innovation of primates or an ancestral condition

316    lost by laboratory mice. The ferret, as a carnivore, serves as an outgroup relative to mice and

317    primates.  Analysis by smFISH for *LAMP5*, *LHX6,* and *GAD1* in ferret neocortex showed that, like

318    mice, ferrets lacked a large deep layer *LAMP5+/LHX6+* interneuron population, suggesting that

319    the expansion of *LAMP5+/LHX6+* interneurons is a primate innovation (Fig. 3b).

320

321    To better appreciate the developmental and evolutionary origins of *LAMP5+/LHX6+* interneurons,

322    we sought clues from other brain areas.  Progenitors in the ganglionic eminences give rise to

323    interneurons that migrate to the neocortex, striatum, hippocampus, and other subcortical

324    structures. Comparing the expression profile of the primate cortical *LAMP5+/LHX6+* population

325    to expression profiles of 17,952 interneurons sampled from eight major structures of the mouse

326    brain[12] revealed that primate cortical *LAMP5+/LHX6+* cells most closely resembled

327    *Lamp5+/Lhx6+* interneurons in the mouse hippocampus (Fig. 3d,e). Although neurogliaform

328

13

**a**

Mouse    Marmoset

*LHX6*  *LAMP5*  Both

**b**

Mouse  Ferret  Marmoset

*GAD1*

*LAMP5*

*LHX6*

Merge

*GAD1+*   *GAD1/LAMP5/LHX6+*

Marmoset
% of cell type by layer

0%  10%  20%  30%

1
2/3
4
5
6

GAD1+/LAMP5+/LHX6-
GAD1+/LAMP5+/LHX6+

**c**

*LHX6/LAMP5+* by cortical region

A1
M1
Par
PFC
S1
Temp
V1

% of GAD+ cells

Mouse  Marmoset  Macaque  Human

**d**

Mouse interneurons across major tissues

Cerebellum
Striatum
Neocortex
Hippocampus
Thalamus
Mixed

**e**

*LHX6*  *LAMP5*  Both

Mouse

Cerebellum
Striatum
Neocortex
Hippocampus
Thalamus

Mouse  Marmoset

Marmoset

**f**

Ctx
Hipp

*CCK+*
*NDNF+*
LAMP5+ types
*LHX6+*

LAMP5 CCK (neocortex)
LAMP5 CCK (hippocampus)
LAMP5 NDNF (neocortex)
LAMP5 NDNF (hippocampus)
LAMP5 LHX6 (neocortex)
LAMP5 LHX6 (hippocampus)

*CCK* vs *LHX6*

*NDNF* vs *LHX6*

**g**

*ID2*        *SCG2*        *NKX2-1*
NDNF+
CCK+
LHX6+
0    200    0    500    0    12

*NDNF*      *CPLX3*      *NOS1*
NDNF+
CCK+
LHX6+
0    40    0    15    0    15

*RELN*      *KIT*        *SLC24A2*
NDNF+
CCK+
LHX6+
0    250    0    400    0    100

*NR2F2*     *CCK*        *LHX6*
NDNF+
CCK+
LHX6+
0    25    0    1000    0    8

*CXCL14*    *CACNA2D1*   *PCDH17*
NDNF+
CCK+
LHX6+
0    7000    0    80    0    70

Expression per 100k transcripts

**h**

Id2-CreER;Nkx2.1-Flpo;Ai65    Hippocampus

L6 S1

329 **Figure 3. Cortical *LHX6+/LAMP5+* interneurons are more numerous in primates and are**
330 **molecularly similar to conserved hippocampal interneurons.**
331 **a,** Cells expressing *LHX6, LAMP5,* or both are plotted on t-SNEs from mouse and marmoset
332 neocortical data. **b,** (left) Single molecule fluorescence *in situ* hybridization (smFISH) in marmoset
333 neocortex showing an example of an *LHX6+/LAMP5+/GAD1+* cell. (right) Quantification by layer
334 of *LAMP5+/LHX6-/GAD1+* cells (blue) and *LAMP5+/LHX6+/GAD1+* cells (red) in marmoset
335 neocortex. **c,** Abundances of *LHX6+/LAMP5+* cells, expressed as proportions of *GAD1+*
336 interneurons sampled by Drop-seq, in marmoset, macaque, and human. Each sampled
337 neocortical region is plotted separately. **d,** Mouse interneurons across major brain structures
338 from [12]. Clusters are colored based on their dominant region of origin. **e,** (left, top) Integrative
339 cross-species analysis (using LIGER) of marmoset neocortical, hippocampal, and striatal
340 interneurons and mouse interneurons from (**d**). (left, bottom) The same t-SNE, colored to show
341 cells expressing *LHX6*, *LAMP5,* or both genes. Cluster outlined in t-SNE contains cells that
342 express both *LAMP5* and *LHX6*. (right) The same t-SNE, colored for each species separately by
343 cell region of origin. **f,** (left) Clustering of marmoset neocortex and hippocampal interneurons.
344 (right) Hierarchical clustering of *LAMP5+* subtypes, separated by region (neocortex and
345 hippocampus). Scatter plots of relative gene expression in pairs of subtypes ($log_{10}$). **g,** Scaled,
346 normalized expression of select gene markers that distinguish the three main *LAMP5+* types. **h,**
347 To identify this population in mouse, and to determine whether the same Nkx2.1 lineage gives
348 rise to such cells in hippocampus and neocortex, *Id2-CreER; Nkx2.1-Flpo; Ai65* animals were
349 examined. (*left*) Overview including neocortex and hippocampus. (*middle, top*) Hippocampus was
350 abundantly labeled. (*middle, bottom*) In neocortex, labeling was extremely sparse and mostly
351 restricted to Layer 6. Labeled cells could be found rarely in L2/3, but not at all in L1. (*right*) A
352 biocytin-filled mouse *Id2;Nkx2.1* interneuron in neocortical layer 6.
353

354 interneurons in the neocortex are thought to derive solely from the CGE, mouse hippocampal

355 *Lamp5+/Lhx6+* interneurons arise from the MGE; such neurons comprise the closely related ivy

356 and neurogliaform subtypes, the most numerous hippocampal *Nos1+* interneurons[35].

357

358 To directly evaluate the similarity of the *LAMP5+/LHX6+* populations in hippocampus and

359 neocortex, we analyzed marmoset neocortical and hippocampal interneurons together (Fig. 3f).

360 The neocortical and hippocampal *LAMP5+/LHX6+* populations formed a common cluster,

361 indicating that they were more similar to each other than to the other two *LAMP5+* neocortical

362 subtypes, which did not express *LHX6* (Fig. 3f,g). Notably marmoset hippocampal and neocortical

363 *LAMP5+/LHX6+* populations expressed *NKX2-1* (Fig. 3g), which in mouse is obligately

364 downregulated in MGE-derived interneurons destined for the neocortex[36] but persists in some

14

365    human cortical interneurons[37]. Fate mapping of interneurons identified by *Id2/Nkx2-1* in mice

366    confirmed that the cortical and hippocampal populations arose from a common (MGE) origin (Fig.

367    3h).

368

369    Primates might have evolved customized allocation for these *LAMP5+/LHX6+* cells from the MGE

370    to neocortex specifically or could have simply expanded the generation of these cells for all brain

371    structures, for example by expanding their progenitor pool. To evaluate these possibilities, we

372    asked whether the *LAMP5+/LHX6+* population has also increased in primate hippocampus.

373    smFISH analyses in in marmoset and mouse indicated that *LAMP5+/LHX6+* neurons populate

374    the same hippocampal layers in the CA1/CA2 region (Extended Data Fig. 5, see also[38]). Thus,

375    the ten-fold expansion of *LAMP5+/LHX6+* neurons in primates appears selective to the neocortex

376    and is likely to represent differentially customized allocation (potentially via different rates or cues

377    for migration, or via different rates of cell death) between neocortex and hippocampus relative to

378    mice. Intriguingly, the *LAMP5+/LHX6+* cells are distinct from, but molecularly most closely related

379    to, recently described LAMP5+ cells that have acquired a distinct "rosehip" morphology and

380    distinct physiological properties in human neocortex relative to their molecularly homologous cell

381    population in mouse cortex[7,39].

382

383    **A novel molecular interneuron type in primate striatum**

384    Although the neocortex has greatly expanded and specialized in the primate lineage[18,40,41], the

385    basal ganglia are deeply conserved collections of subcortical nuclei – so much so that the

386    lamprey, which shared a last common ancestor with mammals more than 500 million years ago,

387    retains nuclei, circuitry, and basic cell types homologous to those observed in mice[2]. Therefore,

388    we expected that interneuron types in the striatum (the largest part of the basal ganglia) would be

389    highly conserved between primates and mice. To our surprise, marmoset striatum revealed, in

390    addition to all the major classes of striatal interneurons found in mice[12,42], a transcriptionally

15

391   distinct type that expressed *VIP* and *TAC3.* This subtype did not appear to have a molecularly

392   homologous population among mouse striatal interneurons (Fig. 4a-b). Because *VIP* is also

393   sparsely expressed in other marmoset striatal interneuron types [Fig. 4b], we hereafter refer to

394   the population as *TAC3+*. The *TAC3+* interneuron population was present in male and female

395   marmosets and in multiple striatal nuclei, including the caudate nucleus, putamen, and nucleus

396   accumbens. Surprisingly, it constituted ~30% of all interneurons in the striatum.

397

398   *TAC3+* interneurons expressed unique combinations of transcription factors, neuropeptides,

399   transporters, and receptors that were not observed in other interneuron subtypes in marmosets

400   (or in any subtype in mice) (Fig. 4c–f). These included genes encoding the androgen receptor

401   (*AR*), serotonin receptor 1A (*HTR1A*), and sugar transporter *SLC35D3*.

402

403

16

**a** Striatal interneurons — Striatal interneurons

CHAT+
CCK/VIP+
SST/NPY/NOS1+
TH+
TAC3+
PVALB+

Mouse
Marmoset

**b** PVALB, CHAT, TH, VIP, CCK, SST, CNR1, TAC3

Mouse
Marmoset

**c** Transcription factors

MEF2C, MAF, SOX1, SATB1, BCL11A, ZNF462, PBX1, ZEB1, TRPS1, DLX1, MAFB, NFIX, AR, ZNF385B, TSHZ3, LHX8, GBX2

CHAT+, CCK+, SST+, TAC3+, TH+, PVALB+

**d** MGE transcription factors — CGE markers

NKX2-1, LHX6, ADARB2, CXCL14
TAC3, VIP, SORCS1, HTR1A
PTPRZ1, SLC35D3, PARM1, STXBP6

CHAT+, CCK+, SST+, TAC3+, PVALB+, TH+

**e** NKX2-1, merge, VIP

**f**

PVALB+ (expression per 100k transcripts)
Neuropeptides
Transcription factors
SCG2, NXPH1, MEF2C, MAF, ZNF462, SOX1, NFIB, VIP, NXPH2, PNOC, NFIX, CBLN4, ZNF385B, TSHZ3, TAC3
TAC3+ (expression per 100k transcripts)

SST+ (expression per 100k transcripts)
NPY, SST, SCG2, BCL11A, PNOC, FOS, SOX2, PBX, DACH1, B11B, SATB1, CBLN4, VGF, NFIB, NR3C1, NFIA, ZNF385B, NKX2-1, CXXC5, TAC3, ETV1, PRDM, DLX2, ST18, VIP
TAC3+ (expression per 100k transcripts)

**g** Marmoset types

CHAT+, CCK+, SST+, TAC3+, PVALB+, TH+

**h**

Mouse striatum
Marmoset striatum
All mouse Vip+ or Tac2+

**i** Striatal interneurons — Striatal interneurons

Human
Marmoset

404  **Figure 4. A primate striatal interneuron type not observed in mouse. a,** Integrative cross-
405  species alignment (using LIGER) of marmoset and mouse striatal interneurons. **b,** Cells
406  expressing markers for each interneuron cluster plotted for marmoset (red) and mouse (blue). **c,**
407  Heat map of expressed transcription factors in marmoset for each striatal subtype. Each gene is
408  scaled to its maximum value across types (black = max). **d,** Beeswarm plots showing additional
409  markers that distinguish *TAC3+* interneurons from other interneuron types in marmoset, including
410  MGE transcription factors (shaded yellow) and CGE markers (shaded blue). Dots are individual
411  cells; bars indicate median expression. **e,** smFISH for *VIP* and *NKX2-1* in marmoset striatum.
412  Cells that co-express both genes identified by arrows. **f,** Scatterplots showing gene expression
413  ($log_{10}$) between *TAC3+* and *PVALB+* or *TAC3+* and *SST+* populations. Differentially expressed
414  (> 3-fold difference) neuropeptides and transcription factors are labeled. **g,** Hierarchical clustering
415  of all expressed genes within marmoset striatal subtypes. **h,** The analysis in (**a**) repeated, but
416  additionally including all mouse extra-striatal interneurons from [12]. For display, the t-SNE shows
417  marmoset striatal interneurons (red), mouse striatal interneurons (blue), and any extra-striatal
418  mouse interneuron that expressed *Vip* or *Tac2* in the Saunders et al. [12] dataset (gray). Circled
419  cells indicate marmoset *TAC3+* population. **i,** LIGER analysis pooling marmoset striatal
420  interneurons with caudate interneurons isolated from human postmortem donors (n=2). Circled
421  cluster indicates aligned marmoset and human *TAC3+* populations.
422

423  The genes expressed by this novel population of striatal interneurons could provide hints about

424  their developmental origins. Although *Vip* and *Tac2* (the mouse homologue of *TAC3*) are

425  neuropeptide genes associated with CGE origin in mouse[9], the marmoset *TAC3+* population

426  expressed the MGE-associated transcription factors *LHX6* and *NKX2-1* (Fig. 4d). Because

427  transcription factor and neuropeptide expression offered divergent clues, we used the entire

428  genome-wide expression pattern to identify the interneurons most similar to the *TAC3+*

429  interneurons. Hierarchical clustering situated the *TAC3+* population between the *SST+* and the

430  *TH+* and *PVALB+* populations, all of which are MGE-derived (Fig. 4g). This suggests that despite

431  expressing some CGE-associated neuropeptides, the *TAC3+* cells are more similar to striatal

432  MGE-derived types than CGE-derived types.

433

434  To evaluate the possibility that a population of cells homologous to the *TAC3+* marmoset striatal

435  interneurons might reside elsewhere in the mouse brain, as in the case of *LAMP5+/LHX6+*

436  neocortical interneurons, we jointly analyzed the RNA expression profiles of marmoset striatal

17

437   interneurons and all interneurons from eight regions of the mouse brain (from [12]). The marmoset

438   *TAC3*+ population still formed its own cluster, suggesting that no homologous cell population

439   existed in any of the mouse brain regions sampled (Fig. 4h). We also confirmed, by single-nucleus

440   RNA sequencing, that no such population is present in ferret striatum, consistent with the

441   interpretation that this cell population was introduced in the lineage leading to primates rather

442   than being lost in mice.

443

444   The *TAC3*+ interneuron population appeared to be shared between marmosets and humans:

445   comparing marmoset striatal interneurons to interneurons obtained from the caudate nucleus from

446   adult human postmortem donors identified a human striatal interneuron population with a

447   homologous pattern of gene expression (Fig. 4i and Extended Data Fig. 6). The genes that were

448   differentially expressed in marmoset between the *TAC3*+ population and other striatal interneuron

449   types also tended to be differentially expressed in the corresponding comparisons in human

450   (Extended Data Fig. 6c).  The *TAC3*+ population constituted 38% of the interneurons sampled in

451   human striatum.

452

453   The abundance of the novel, TAC3+ interneuron population raised the question of whether it had

454   replaced, or added to, conserved interneuron populations. The primates exhibited expanded

455   representation of interneurons in the striatum overall: while interneurons were 4.1% of all striatal

456   neurons in mice, they were 13.1% of all striatal neurons in marmosets and 10.8% in humans,

457   consistent with stereological estimates of higher interneuron proportions in primate striatum[43].

458   Thus, compared to mice, primate striatum has more than doubled the proportion of interneurons,

459   including an interneuron type with no clear homolog in mice.

460

461

462   **DISCUSSION**

463

464    "Cell types" have been defined as collections of cells that change together over the course

465    evolution[44,45]. In this study, we found that although most of the major molecularly defined types of

466    cortical interneurons are conserved across mice, humans, marmosets, and macaques, these

467    interneurons nonetheless have undergone surprising levels of evolutionary change in the genes

468    they express and the relative levels at which they express pan-neuronal genes.  The significance

469    of these changes for the detailed physiological and connectivity properties of interneurons will be

470    important to understand[46].

471

472    We found that the primate striatum contains an abundant interneuron type that has no

473    homologous cell population in mice. These TAC3+ interneurons constituted 30% of interneurons

474    in human and marmoset striatum, and expressed a suite of transcription factors and

475    neuropeptides that distinguished them from other striatal interneurons.  This innovation in primate

476    striatum was accompanied by a broader expansion in the numbers of interneurons that doubled

477    their representation as a fraction of all striatal neurons.  These observations raise questions about

478    how expanded numbers and types of interneurons have changed the functions of striatal

479    microcircuits and their roles within the larger cortico-striatal circuits that contribute to primate

480    cognition and potentially to neuropsychiatric disorders[47].

481

482    We found that an interneuron type that is abundant in the mouse hippocampus—the ivy cell, which

483    has properties similar to neurogliaform cells and is defined by co-expression of *Lamp5* and

484    *Lhx6*[35]—has expanded throughout the neocortex in primates. Primates have retained ivy cells in

485    the hippocampus but appear to have also greatly upregulated the production of these cells,

486    increased their recruitment to the neocortex, and expanded their distribution throughout

487    neocortical areas and layers. Other neurogliaform cell types in mouse neocortex are preferentially

19

488    found in upper layers and signal by volume transmission, the diffuse release of the inhibitory

489    neurotransmitter GABA in the absence of conventional synapses[48].  Given that these cells were

490    proportionally most numerous in the deep layers of the primate neocortex, one possibility is that

491    in primates, these cells now contribute diffuse inhibitory signaling in new neocortical contexts.

492

493    The qualitative and quantitative deployment of gene expression across conserved interneuron

494    types indicated that, even for genes that are pan-neuronally expressed, evolution has strongly

495    constrained quantitative gene expression levels on evolutionary time scales, although substantial

496    differences in gene expression and cellular proportions still clearly distinguish primates from mice.

497

498    Efforts to model the effects of specific genes and mutations on human brain function and illness

499    could be facilitated by systematic data sets that reveal the extent to which each gene's cell-type–

500    specific pattern of expression is shared by humans with each other species (e.g., [8]).  We hope

501    that the data from the experiments reported here will inform the design and interpretation of such

502    studies. Accordingly, we have developed a simple, web-based data resource to enable cross-

503    species comparisons of interneurons (http://interneuron.mccarrolllab.org/).

504

505    Our results reveal the ways in which the cellular and molecular repertoires of mouse and primate

506    neurons have adapted over time. These evolutionary paths likely involved diverse developmental

507    mechanisms, including alteration of neurogenesis rates, the creation of novel migratory pathways,

508    and changes in gene regulation.  Their effects on circuitry, cytoarchitecture, and physiology will

509    be important and interesting to understand.

510

511    The innovations among interneurons are notable because the single-cell expression studies of

512    tetrapod species performed to date – in lizards, turtles, mice, and primates – had suggested that

513    the known interneuron types are conserved across a broad taxonomic range. In this study,

514    however, we identified surprising variation in interneurons within and across species, which was

515    discordant with expectations in key ways. For example, the primate CGE might be expected to

516    harbor evolutionary novelties in its interneuron repertoire because it generates a larger proportion

517    of interneurons in primates than in rodents[49], because CGE interneurons are born later than MGE

518    interneurons, and because CGE interneurons preferentially occupy the expanded upper

519    neocortical layers[50]. However, the most striking cellular aspects of rodent–primate divergence in

520    the cortex, hippocampus, and striatum involved interneurons whose RNA expression patterns

521    indicated that they originated in the MGE. Similarly, although the neocortex has attracted intense

522    interest because it is a highly evolved and specialized structure that is thought to underlie

523    expansions in primate cognitive capability, it was in the striatum that we identified a primate

524    interneuron type with no mouse homolog. The systematic analysis of many more species and cell

525    types may reveal more such examples of evolutionary flexibility and innovation.

526
527    **REFERENCES**

528    1.        Northcutt, R. G. & Kaas, J. H. The emergence and evolution of mammalian
529           neocortex. *Trends in Neurosciences* **18,** 373–379 (1995).
530    2.        Laclef, C. & Métin, C. Conserved rules in embryonic development of cortical
531           interneurons. *Seminars in Cell and Developmental Biology* **76,** 86–100 (2018).
532    3.        Barkan, C. L., Kelley, D. B. & Zornik, E. Premotor neuron divergence reflects vocal
533           evolution. *J. Neurosci.* **38,** 5325–5337 (2018).
534    4.        Peng, Y.-R. *et al.* Molecular Classification and Comparative Taxonomics of Foveal
535           and Peripheral Cells in Primate Retina. *Cell* **176,** 1222–1237.e22 (2019).
536    5.        Katz, P. S. & Harris-Warrick, R. M. The evolution of neuronal circuits underlying
537           species-specific behavior. *Current Opinion in Neurobiology* **9,** 628–633 (1999).
538    6.        Tosches, M. A. *et al.* Evolution of pallium, hippocampus, and cortical cell types
539           revealed by single-cell transcriptomics in reptiles. *Science* **360,** 881–888 (2018).
540    7.        Hodge, R. D. *et al.* Conserved cell types with divergent features between human
541           and mouse cortex. *biorxiv.org*
542    (2018). doi:10.1101/384826
543    8.        Xu, X. *et al.* Species and cell-type properties of classically defined human and
544           rodent neurons and glia. *eLIFE* **7,** 1–47 (2018).
545    9.        Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505,** 318–
546           326 (2014).
547    10.      Tosches, M. A. & Laurent, G. Evolution of neuronal identity in the cerebral cortex.
548           *Current Opinion in Neurobiology* **56,** 199–208 (2019).

549  11.  Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical
550       areas. 1–37 (2017). doi:10.1101/229542
551  12.  Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the
552       Adult Mouse Brain. *Cell* **174,** 1015–1030 (2018).
553  13.  Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual
554       Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
555  14.  Paul, A. *et al.* Transcriptional Architecture of Synaptic Communication Delineates
556       GABAergic Neuron Identity. *Cell* **171,** 522–525.e20 (2017).
557  15.  Herculano-Houzel, S. The remarkable, yet not extraordinary, human brain as a
558       scaled-up primate brain and its associated cost. *Proc Natl Acad Sci USA* **109,**
559       10661–10668 (2012).
560  16.  Kim, Y. *et al.* Brain-wide Maps Reveal Stereotyped Cell-Type- Based Cortical
561       Architecture and Subcortical Sexual Dimorphism. *Cell* **171,** 456–460.e22 (2017).
562  17.  Florio, M., Borrell, V. & Huttner, W. B. Human-specific genomic signatures of
563       neocortical expansion. *Current Opinion in Neurobiology* **42,** 33–44 (2017).
564  18.  Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and Evolution of the
565       Human Neocortex. *Cell* **146,** 18–36 (2011).
566  19.  Brito, R. B. & (null), G. F. The developmental integration of cortical interneurons into
567       a functional network. *Current Topics in Developmental Biology* **87,** 81–118 (2009).
568  20.  Ataman, B. *et al.* Evolution of Osteocrin as an activity- regulated factor in the
569       primate brain. *Nature* **539,** 242–247 (2016).
570  21.  Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature*
571       **536,** 285–291 (2016).
572  22.  Welch, J., Kozareva, V., Ferreira, A., bioRxiv, C. V.2018. Integrative inference of
573       brain cell similarities and differences from single-cell genomics. *biorxiv.org*
574   doi:10.1101/459891
575  23.  Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical
576       areas. *Nature* **563,** 1–41 (2018).
577  24.  Wester, J. C. *et al.* Neocortical Projection Neurons Instruct Inhibitory Interneuron
578       Circuit Development in a Lineage- Dependent Manner. *Neuron* **102,** 960–975.e6
579       (2019).
580  25.  Lodato, S. *et al.* Excitatory Projection Neuron Subtypes Control the Distribution of
581       Local Inhibitory Interneurons in the Cerebral Cortex. *Neuron* **69,** 763–779 (2011).
582  26.  Sur, M. & Rubenstein, J. L. R. Patterning and plasticity of the cerebral cortex.
583       *Science* **310,** 805–810 (2005).
584  27.  Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature*
585       **508,** 199–206 (2014).
586  28.  Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human
587       brain transcriptome. *Nature* **489,** 391–399 (2012).
588  29.  Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal
589       developmental hierarchies of the human cortex. *Science* **358,** 1318–1323 (2017).
590  30.  Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient
591       of macroscale cortical organization. *Proc Natl Acad Sci USA* **113,** 12574–12579
592       (2016).
593  31.  Krienen, F. M., Yeo, B. T. T., Ge, T., Buckner, R. L. & Sherwood, C. C.
594       Transcriptional profiles of supragranular-enriched genes associate with
595       corticocortical network architecture in the human brain. *Proc Natl Acad Sci USA*
596       **113,** E469–E478 (2016).
597  32.  Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell
598       transcriptomics. *Nat. Neurosci.* **19,** 335–346 (2016).

599  33.  Tosches, M. A. Developmental and genetic mechanisms of neural circuit evolution.
600       *Developmental Biology* **431,** 16–25 (2017).
601  34.  Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons.
602       *Nature* **555,** 457–462 (2018).
603  35.  Tricoire, L. *et al.* Common Origins of Hippocampal Ivy and Nitric Oxide Synthase
604       Expressing Neurogliaform Cells. *Journal of Neuroscience* **30,** 2165–2176 (2010).
605  36.  Nóbrega-Pereira, S. *et al.* Postmitotic Nkx2-1 Controls the Migration of
606       Telencephalic Interneurons by Direct Repression of Guidance Receptors. *Neuron*
607       **59,** 733–745 (2008).
608  37.  Jakovcevski, I., Mayer, N. & Zecevic, N. Multiple Origins of Human Neocortical
609       Interneurons Are Supported by Distinct Expression of Transcription Factors.
610       *Cerebral Cortex* **21,** 1771–1782 (2010).
611  38.  Tricoire, L. *et al.* A Blueprint for the Spatiotemporal Origins of Mouse Hippocampal
612       Interneuron Diversity. *Journal of Neuroscience* **31,** 10948–10970 (2011).
613  39.  Boldog, E. *et al.* Transcriptomic and morphophysiological evidence for a specialized
614       human cortical GABAergic cell type. *Nat. Neurosci.* **21,** 1185–1195 (2018).
615  40.  Buckner, R. L. & Krienen, F. M. The evolution of distributed association networks in
616       the human brain. *Trends Cogn. Sci.* **17,** 648–665 (2013).
617  41.  Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor
618       amplification and neocortex expansion. *Science* **347,** 1465–1470 (2015).
619  42.  Muñoz Manchado, A. B. *et al.* Diversity of Interneurons in the Dorsal Striatum
620       Revealed by Single-Cell RNA Sequencing and PatchSeq. *Cell Reports* **24,** 2179–
621       2190.e7 (2018).
622  43.  Graveland, G. A. & Difiglia, M. The frequency and distribution of medium-sized
623       neurons with indented nuclei in the primate and rodent neostriatum. *Brain Res.* **327,**
624       307–311 (1985).
625  44.  Arendt, D. *et al.* The origin and evolution of cell types. *Nat Rev Genet* **17,** 744–757
626       (2016).
627  45.  Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities
628       and the path forward. *Nat Rev Neurosci* **18,** 530–546 (2017).
629  46.  Huang, Z. J. & Paul, A. The diversity of GABAergic neurons and neural
630       communication elements. *Nat Rev Neurosci* 1–10 (2019). doi:10.1038/s41583-019-
631       0195-4
632  47.  Middleton, F. A. & Strick, P. L. Basal Ganglia Output and Cognition: Evidence from
633       Anatomical, Behavioral, and Clinical Studies. *Brain and Cognition* **42,** 183–200
634       (2000).
635  48.  Oláh, S. *et al.* Regulation of cortical microcircuits by unitary GABA-mediated
636       volume transmission. *Nature* **461,** 1278–1281 (2009).
637  49.  Hladnik, A., Džaja, D., Darmopil, S., Jovanov-Milošević, N. & Petanjek, Z. Spatio-
638       temporal extension in site of origin for cortical calretinin neurons in primates. *Front*
639       *Neuroanat* **8,** 50 (2014).
640  50.  Miyoshi, G. *et al.* Genetic Fate Mapping Reveals That the Caudal Ganglionic
641       Eminence Produces a Large and Diverse Population of Superficial Cortical
642       Interneurons. *Journal of Neuroscience* **30,** 1582–1594 (2010).
643

23

644

655

656    **Author Contributions**

657    F.M.K, S.A.M., G. Feng, and G. Fishell designed the study. F.M.K. prepared and dissected tissue;

658    L.B. and M.G. developed the nuclei Drop-seq protocol. M.G., A.L., C.D.M., N.R., E.B., and L.B.

659    performed Drop-seq and prepared sequencing libraries. M.G. performed sequencing, alignment,

660    and QC analysis. F.M.K., A.S., J.N., A.W., D.K. R.d.R., and S.A.M. developed analysis pipelines.

661    F.M.K. analyzed the data with input from S.A.M, G. Fishell, M.F., A.L. and A.S. D.K. developed the

662    web resource. Q.Z., C.W., M.B., V.T., R.S., C.A.W., L.K., S.B., and G. Feng provided tissue for Drop-

663    seq and smFISH experiments. K.L., H.Z., C.D.K., N.R., E.B., M.F-O., J.L., F.M.K. and J.D. performed

664    and analyzed smFISH experiments. R.M., B.S., and B.R. contributed fate-mapping experiments.

665    F.M.K and S.A.M wrote the paper with input from coauthors.

666

667    **Extended Data Table 1.** Specimen information table.

668      **Extended Data Table 2.** Reagent and resource table.
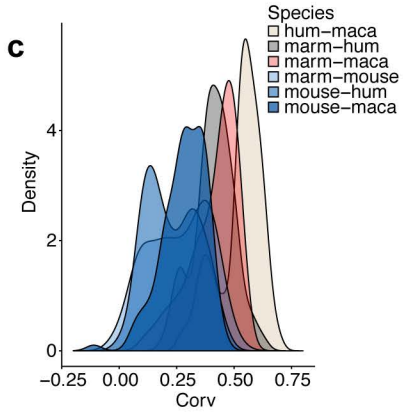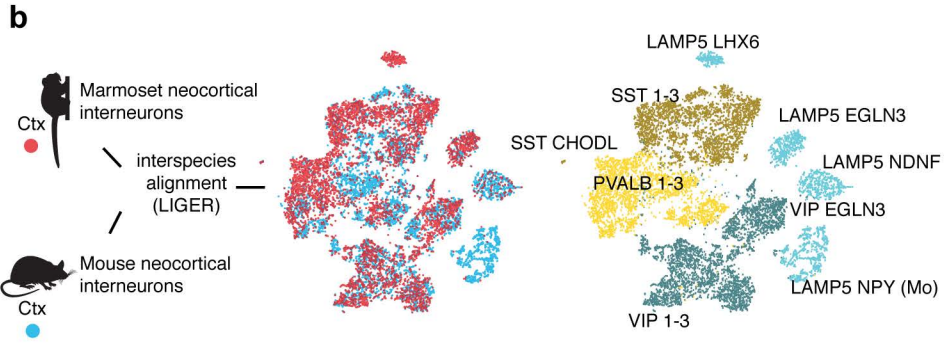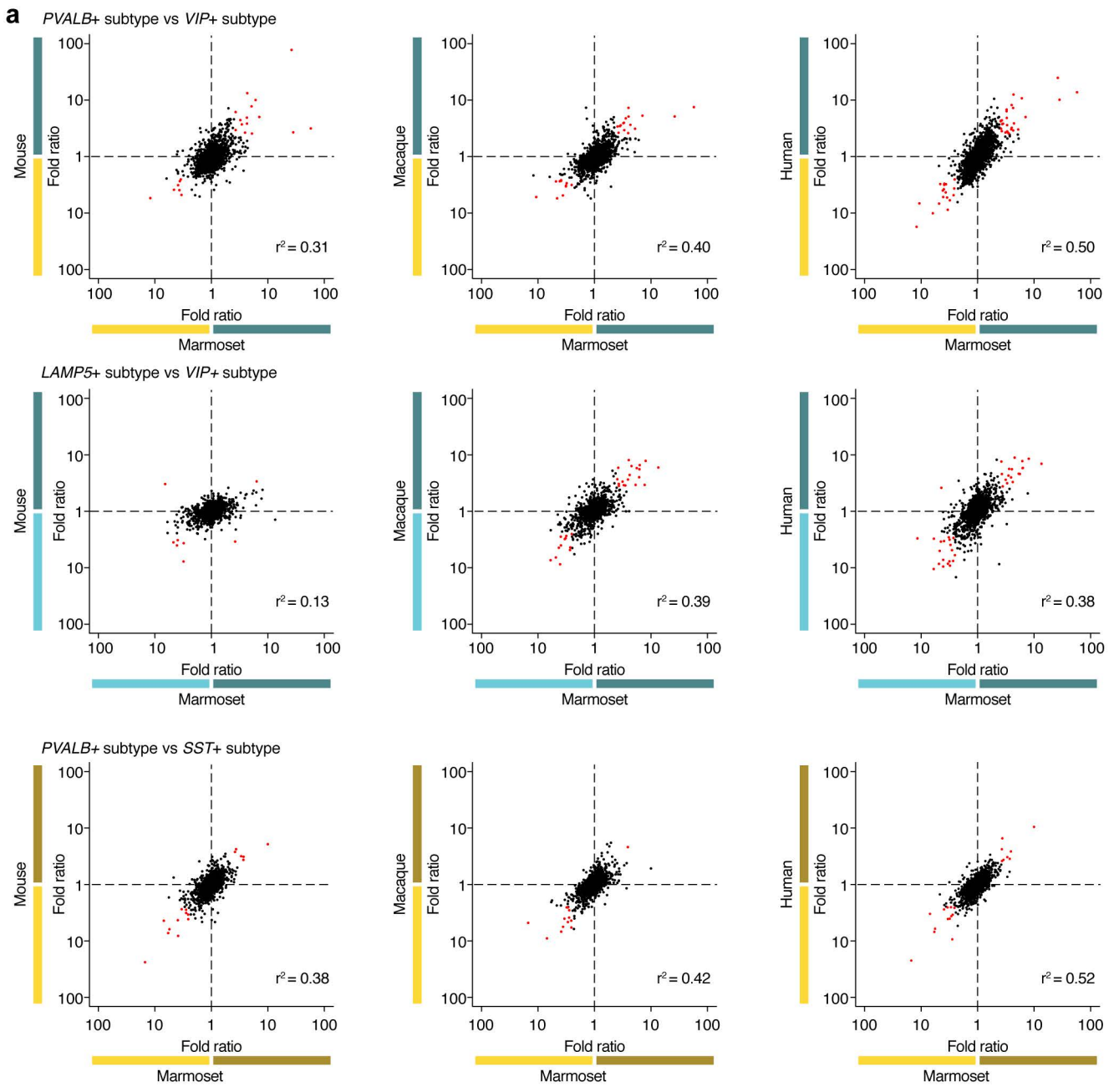
669
670

# Extended Data Figure 1
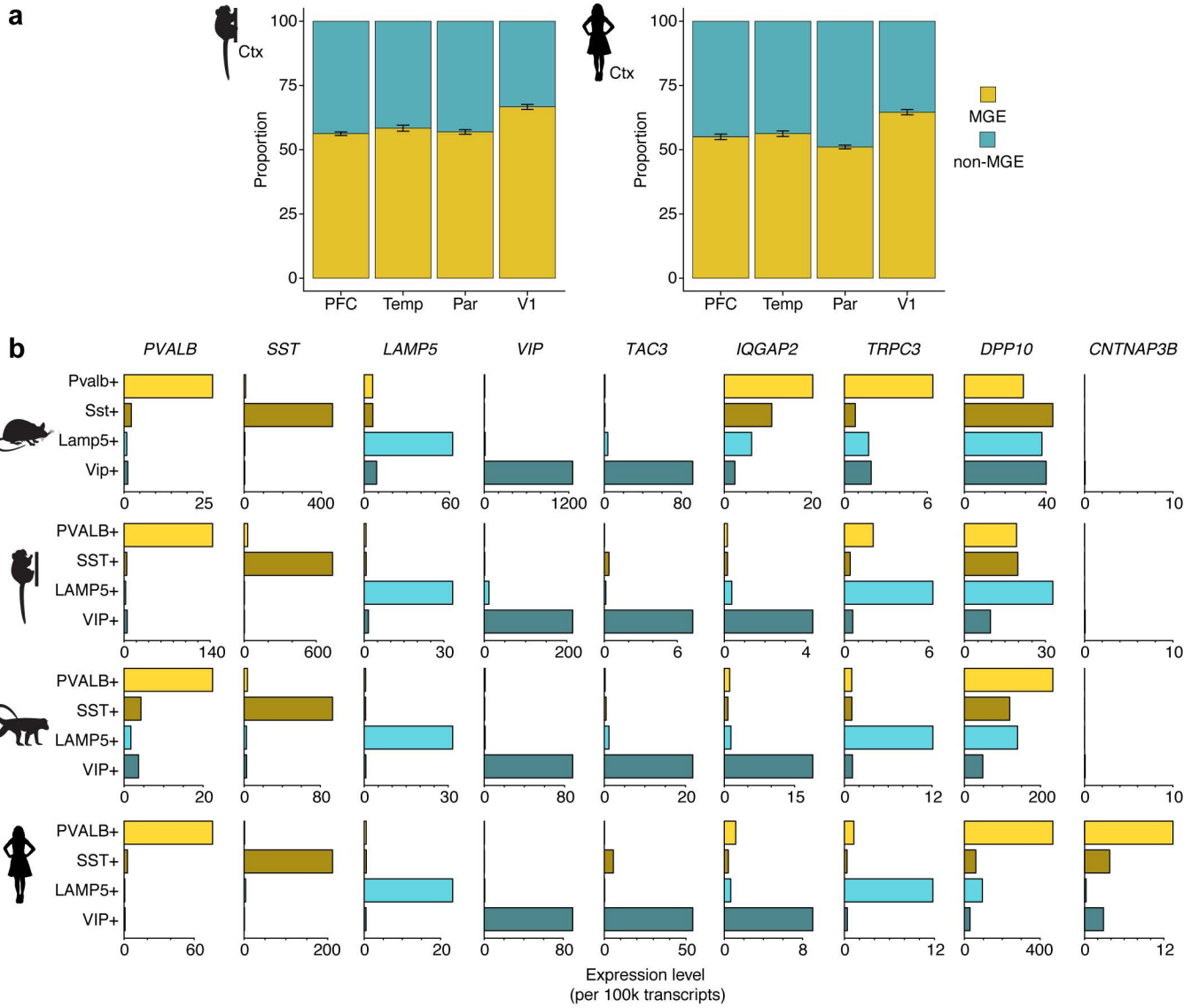
671     **Extended Data Figure 1. Interneuron abundances and gene expression in neocortex.**

672     **a,** Proportion of MGE and non-MGE interneurons in cortical association areas (prefrontal cortex,

673     temporal pole, and lateral parietal association cortex) and in primary visual area (V1) in marmoset

674     and human. Error bars represent binomial confidence intervals. **b,** Examples of markers that are

675     consistent, or that vary across species, within the four primary interneuron classes. Values are

676     scaled counts per 100k transcripts.

677

678

679     **Extended Data Figure 2. Pairwise comparisons of neocortical interneuron types across**

680     **species. a,** Fold difference of each expressed gene between types of MGE-derived interneurons

681     and types of CGE-derived interneuron across pairs of species. Genes in red have >3-fold

682     expression difference in either cell type in each species pair.    **b,** (left) LIGER integration of

683     marmoset (red dots) and mouse (teal dots) neocortical interneurons. (right) Same t-SNE with

684     clusters colored by interneuron class (colors as in Fig. 1). **c,** Density histogram of correlation (*r*)

685     values of fold differences computed for each possible cluster pair from LIGER species-integrated

686     analyses. Each density trace corresponds to a species pair; blue traces indicate primate–mouse

687     comparisons, and other colors indicate primate–primate comparisons.
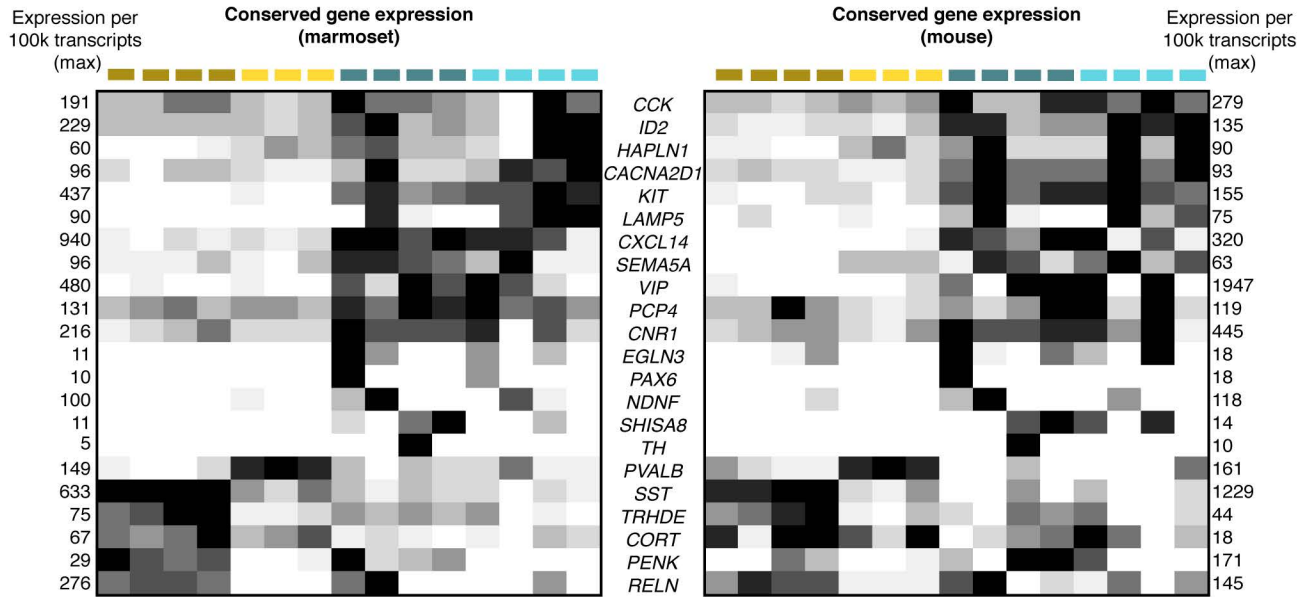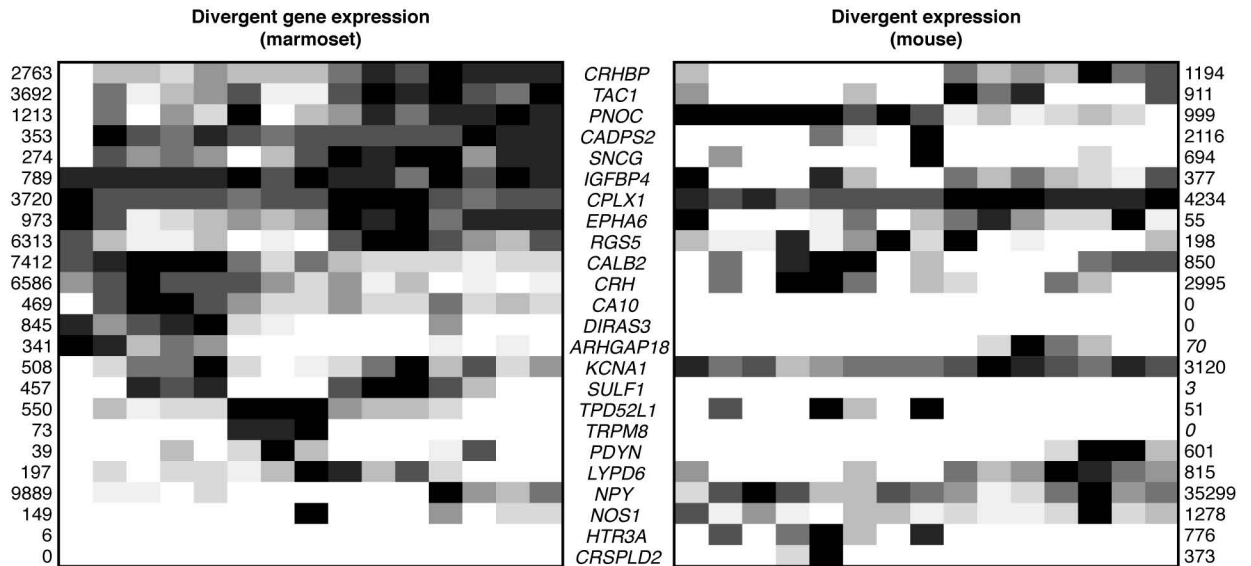
688

689

# Extended Data Figure 3
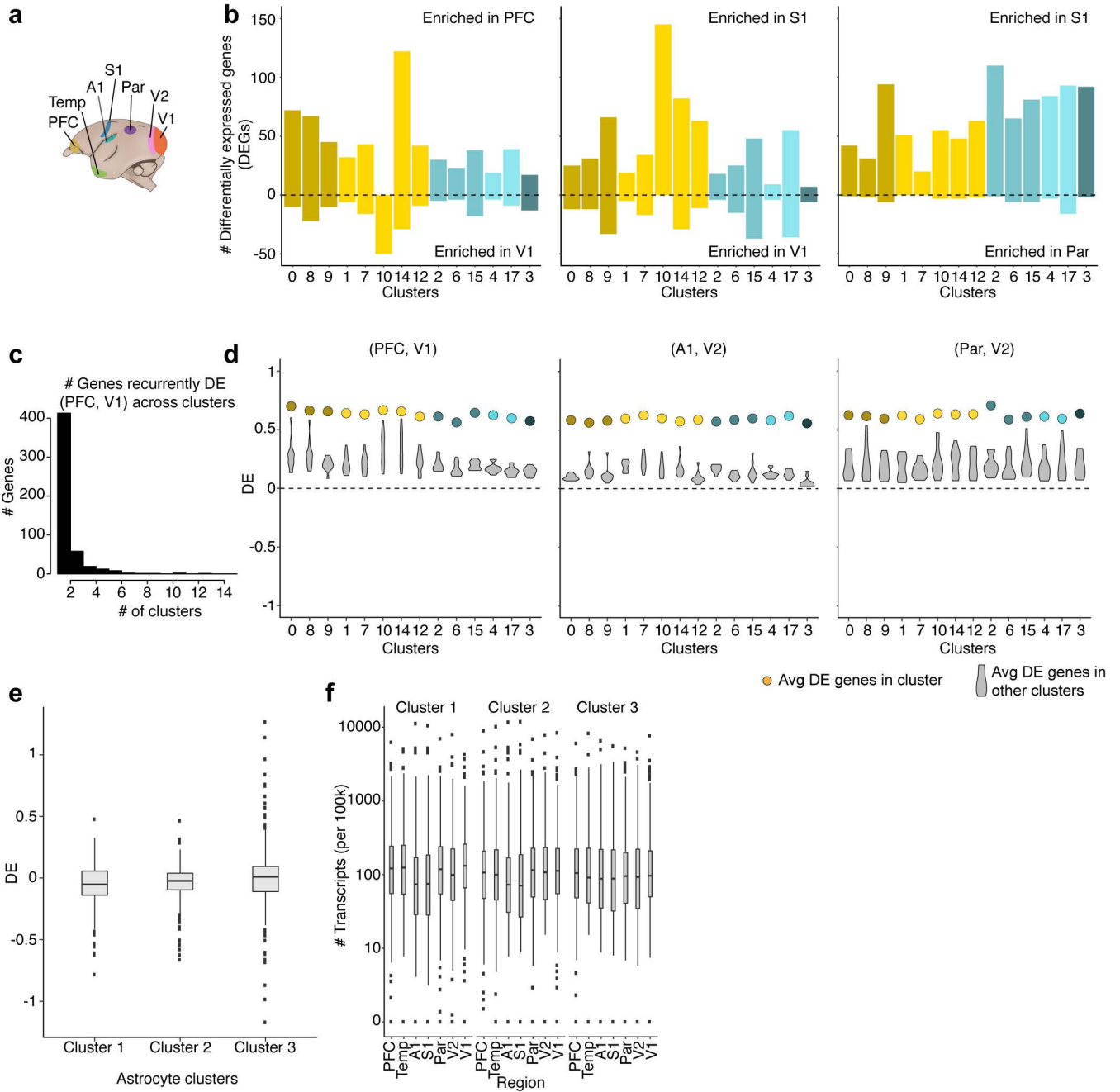
690    **Extended Data Figure 3. Regional gene expression variation in neocortex. a,** Schematic of

691    neocortical region locations in marmoset. **b,** Histogram of the number of regionally differentially

692    expressed genes (rDEGs) (>3-fold expression difference) between three representative pairs of

693    regions, in each cell type (cluster) for which there were at least 50 cells per region. **c,** Histogram

694    of the number of interneuron clusters (cell types) in which a given gene is differentially expressed.

695    At a threshold of >3-fold, most genes are only differentially expressed in a single cell type (cluster).

696    **d,** Colored dots represent average fold difference of DEGs in each cluster in marmoset

697    interneurons. Violin plots represent the distribution of average fold differences in each cluster (cell

698    type) when using rDEGs from other clusters. Three representative region pairs are shown. **d,** Fold

699    ratios ($\log_{10}$) between PFC and V1 for three astrocyte subtypes (n = 32,600 nuclei) in marmosets

700    using rDEGs identified between PFC and V1 in interneurons.  **e,** Expression in astrocytes of genes

701    that exhibited an anterior–posterior expression gradient in interneurons. Regions are arranged in

702    anterior–posterior order on the x axis.

703

704

28

# Extended Data Figure 4

## a



**Conserved gene expression (marmoset)**

Expression per 100k transcripts (max)

**Conserved gene expression (mouse)**

Expression per 100k transcripts (max)

| Marmoset max | Gene | Mouse max |
|---|---|---|
| 191 | CCK | 279 |
| 229 | ID2 | 135 |
| 60 | HAPLN1 | 90 |
| 96 | CACNA2D1 | 93 |
| 437 | KIT | 155 |
| 90 | LAMP5 | 75 |
| 940 | CXCL14 | 320 |
| 96 | SEMA5A | 63 |
| 480 | VIP | 1947 |
| 131 | PCP4 | 119 |
| 216 | CNR1 | 445 |
| 11 | EGLN3 | 18 |
| 10 | PAX6 | 18 |
| 100 | NDNF | 118 |
| 11 | SHISA8 | 14 |
| 5 | TH | 10 |
| 149 | PVALB | 161 |
| 633 | SST | 1229 |
| 75 | TRHDE | 44 |
| 67 | CORT | 18 |
| 29 | PENK | 171 |
| 276 | RELN | 145 |

## b

**Divergent gene expression (marmoset)**

**Divergent expression (mouse)**

| Marmoset max | Gene | Mouse max |
|---|---|---|
| 2763 | CRHBP | 1194 |
| 3692 | TAC1 | 911 |
| 1213 | PNOC | 999 |
| 353 | CADPS2 | 2116 |
| 274 | SNCG | 694 |
| 789 | IGFBP4 | 377 |
| 3720 | CPLX1 | 4234 |
| 973 | EPHA6 | 55 |
| 6313 | RGS5 | 198 |
| 7412 | CALB2 | 850 |
| 6586 | CRH | 2995 |
| 469 | CA10 | 0 |
| 845 | DIRAS3 | 0 |
| 341 | ARHGAP18 | 70 |
| 508 | KCNA1 | 3120 |
| 457 | SULF1 | 3 |
| 550 | TPD52L1 | 51 |
| 73 | TRPM8 | 0 |
| 39 | PDYN | 601 |
| 197 | LYPD6 | 815 |
| 9889 | NPY | 35299 |
| 149 | NOS1 | 1278 |
| 6 | HTR3A | 776 |
| 0 | CRSPLD2 | 373 |

705 **Extended Data Figure 4. Conserved and divergent gene expression across neocortical**

706 **types. a,** Heat map of exemplar genes that had consistent patterns of expression in LIGER

707 integrated marmoset–mouse clusters from Extended Data Fig. 2b. Each gene (row) is scaled to

708 the scaled max (black) expression (values given outside plots) for each species separately. **b,**
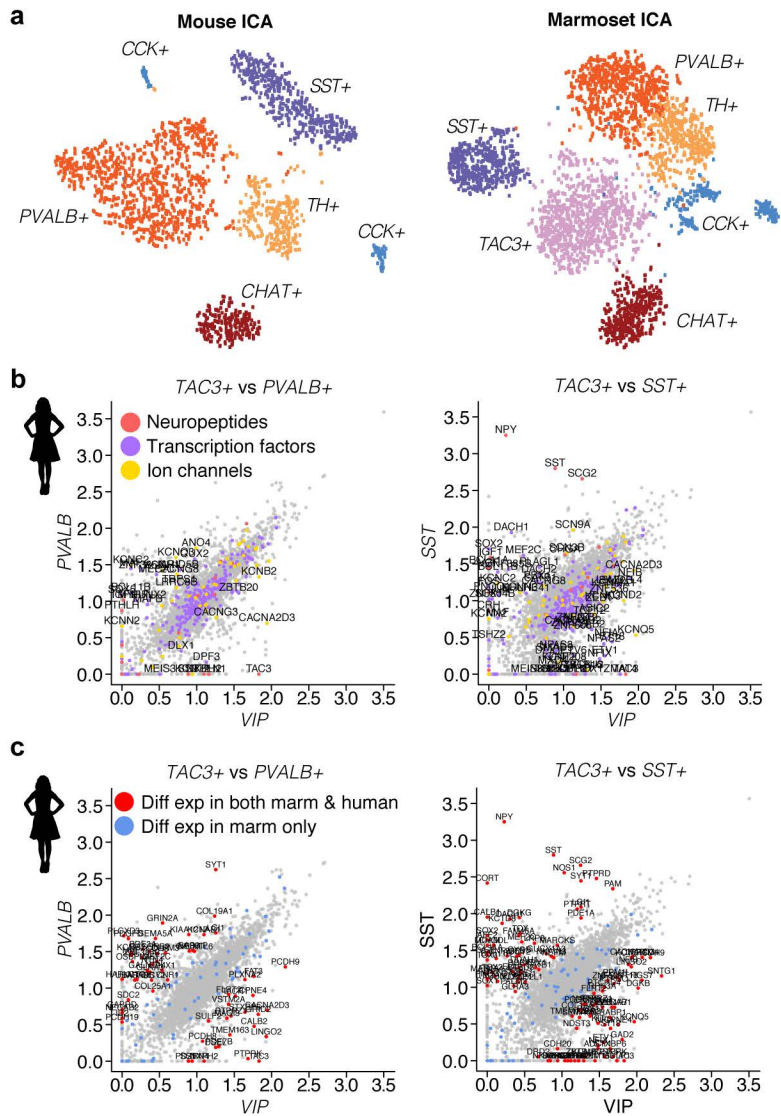
709 Heatmap of exemplar genes that have divergent expression patterns in LIGER-integrated

710 marmoset–mouse clusters from Extended Data Fig. 2b. Each gene (row) is scaled to the scaled

711 max (black) expression (values given outside plots) for each species separately.

712

713

# Extended Data Figure 5

**a**

Marmoset hippocampus



**b**



- GAD1+/LAMP5+/LHX6-
- GAD1+/LAMP5+/LHX6+

**c**
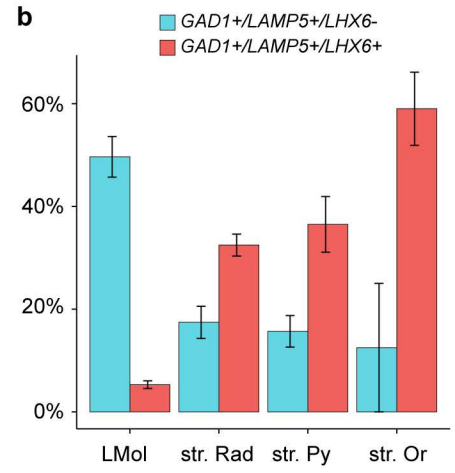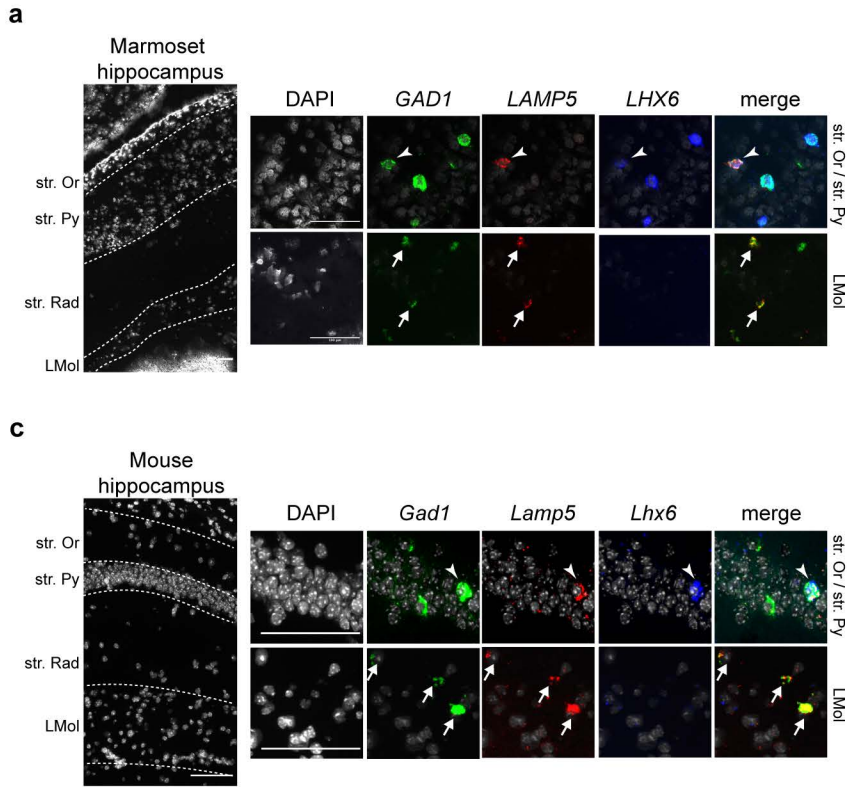
Mouse hippocampus

714 **Extended Data Figure 5. LAMP5+ interneuron types in hippocampus. a,** Single-molecule

715 fluorescence in situ hybridization (smFISH) for *GAD1*, *LAMP5,* and *LHX6* in marmoset

716 hippocampal layers (CA1/CA2 subfields). Arrowhead indicates triple positive cells; arrow

717 indicates the *LHX6-* population. (*top row)* Strata oriens (Str. Or) and strata pyramidale (Str. Py).

718 (*bottom row)* strata lacunosum moleculare (LMol). Scale bars = 100 um. **b,** Quantification of

719 *GAD1*/*LAMP5/LHX6+* (green) and *GAD1*/*LAMP5/LHX6-* cells as percentage of all *GAD1+* cells

720 in marmoset hippocampus (compare to mouse data in[38]). **c,** smFISH for *Gad1*, *Lamp5*, and *Lhx6*

721 in mouse hippocampal layers (CA1).

722

723

724    **Extended Data Figure 6. Interneuron types in striatum. a,** t-SNE representations of ICA-based

725    clustering for mouse and marmoset striatal interneurons. **b,** Gene expression differences in

726    human caudate between *TAC3*+ and *PVALB*+ (left) or *TAC3*+ and *SST*+ (right) populations.

727    Neuropeptides (red) and transcription factors (blue) are labeled. **c,** Same data as in **b**, but instead

728    highlighting genes that were differentially expressed in both marmoset and human (red), or only

729    in marmoset (blue).

730

731  **METHODS**
732

733  Specimen information is available in Extended Data Table 1.

734  Reagent information is available in Extended Data Table 2.

735

736  **Specimens and Donors for Nuclei Drop-seq**

737

738  *Mouse.* Mouse experiments were approved by and in accordance with Harvard Medical School

739  IACUC protocol number IS00000055-3. Sections of frontal and visual cortex were prepared from

740  male and female adult mice (60–70 days old; C57Blk6/N, Charles River Labs Stock #027). Mice

741  were deeply sedated with isoflurane and transcardially perfused with ice-cold Sucrose-HEPES

742  buffer described in[1], which contains (in mM) 110 NaCl, 2.5 KCl, 10 HEPES, 7.5 $MgCl_2$, 25 glucose,

743  75 sucrose (~350 mOsm/kg$^{-1}$); sectioned; and flash-frozen on liquid nitrogen.

744

745  *Marmoset*. Marmoset experiments were approved by and in accordance with Massachusetts

746  Institute of Technology IACUC protocol number 051705020. Three adult marmosets (1.5–2 years

747  old; one male, 2 females) were deeply sedated by intramuscular injection of ketamine (20-40

748  mg/kg) or alfaxalone (5-10 mg/kg), followed by intravenous injection of sodium pentobarbital (10–

749  30 mg/kg). When pedal withdrawal reflex was eliminated and/or respiratory rate was diminished,

750  animals were transcardially perfused with ice-cold Sucrose-HEPES buffer. Whole brains were

751  rapidly extracted into fresh buffer on ice. Sixteen 2-mm coronal blocking cuts were rapidly made

752  using a custom-designed marmoset brain matrix. Slabs were transferred to a dish with ice-cold

753  Dissection Buffer[1], and regions of interest were dissected using a marmoset atlas as reference[2].

754  Regions were snap-frozen in liquid nitrogen and stored in individual microcentrifuge tubes at -

755  80°C.

756

757    *Macaque.* Whole brains from two healthy, immunologically and treatment-naive adult macaques

758    (2 males; 10–11 years old) were obtained from terminal experiments (IACUC 4315-02). Animals

759    were deeply sedated with ketamine and euthanized by pentobarbital overdose, and transcardially

760    perfused with ice-cold Sucrose-HEPES buffer. Brains were rapidly blocked in ~5-mm coronal

761    slabs and frozen in liquid nitrogen or isopentane on dry ice.

762

763    *Human.* Frozen tissue was obtained from the Harvard Brain Tissue Resource Center (HBTRC;

764    McLean Hospital). Four donors were used for analysis of striatal interneurons, and two for

765    analysis of neocortical interneurons. History of psychiatric or neurological disorders was ruled out

766    by consensus diagnosis carried out by retrospective review of medical records and extensive

767    questionnaires concerning social and medical history provided by family members. Several

768    regions from each brain were examined by a neuropathologist. The cohort used for this study did

769    not include subjects with evidence of gross and/or macroscopic brain changes, or clinical history,

770    consistent with cerebrovascular accident or other neurological disorders. Subjects with Braak

771    stages III or higher (modified Bielchowsky stain) were not included. None of the subjects had

772    significant history of substance dependence within 10 or more years of death, as further

773    corroborated by negative toxicology reports.

774

775    **Nuclei Drop-seq library preparation and sequencing**

776    Nuclei suspensions were prepared from frozen tissue and used for Nuclei Drop-seq following the

777    protocol described at https://protocols.io/view/extraction-of-nuclei-from-brain-tissue-2srged6.

778    Drop-seq libraries were prepared as previously described[3] with modifications, quantification, and

779    QC as described in[1], as well as the following modifications optimized for nuclei: in the Drop-seq

780    lysis buffer, 8 M guanidine hydrochloride (pH 8.5) was substituted for water, nuclei were loaded

781    into the syringe at a concentration of 176 nuclei/μL, and cDNA amplification was performed using

782    around 6000 beads per reaction (15 PCR cycles were used for marmoset nuclei, and 16 for

33

783 macaque and human nuclei). Raw sequencing reads were aligned to the following genome

784 assemblies: GRCm38.81 (mouse), calJac3 (marmoset), Mmul8.0.1 (macaque), and hg19

785 (human). Reads that mapped to exons or introns of each assembly were assigned to annotated

786 genes.

787

**Mouse single-cell dataset**

789 Interneurons were curated *in silico* from the single-cell datasets available in [1] from available

790 structures: frontal and posterior neocortex, striatum, cerebellum, thalamus, hippocampus,

791 substantia nigra, and entopeduncular nucleus.

792

**Single species independent component analysis (ICA)**

794 Initial analyses to identify interneurons based on marker expression were conducted on each

795 species separately. Nuclei with fewer than 300 detected genes were removed from analysis.

796 Briefly, independent component analysis (ICA, using the fastICA package in R) was performed

797 on each species and each region's digital gene expression (DGE) matrix separately after

798 normalization and variable gene selection as in [1]. These first-round individual-species analyses

799 produced clustering solutions with ~8–11 clusters of major cell types (neurons, glia, vasculature),

800 from which interneuron clusters could be identified based on canonical markers (e.g. *GAD1*,

801 *GAD2*). The raw DGEs were subsetted to include only cells from these clusters to form new,

802 interneuron-only DGEs. Normalization, variable gene selection, and ICA was repeated on these

803 interneuron-only DGEs, but this time the full ICA curation pipeline described in [1] was used to

804 identify doublets, outliers, artifactual signals, and biological components of interest. Cells

805 identified by this procedure as doublets or outliers were removed from the DGEs, and these

806 filtered DGEs were then carried forward for integrated analyses across regions and/or species

807 using LIGER[4].

808

809 **Interneuron abundances and local specialization in neocortex**

810 Proportions of *PVALB+* or *SST+* (MGE-derived) and *VIP+* or *LAMP5+* (non-MGE derived) were

811 calculated for each species separately for frontal association areas (FC, mouse) or prefrontal

812 cortex (PFC, primates) and visual cortex (V1). Cells were allocated to MGE or non-MGE pools

813 based on their cluster assignment in individual-species ICA clustering. Error bars represent 95%

814 confidence intervals for binomial probability, computed with the R package Hmisc.

815
816
817
818 **Identification of conserved and divergent genetic programs within conserved interneuron**

819 **types**

820 Interneurons from the neocortex of each species were partitioned into four main classes based

821 on marker expression (*VIP, LAMP5, PVALB, SST*). (We also detected a rare and distinct, fifth

822 category of cortical GABAergic cell – *MEIS2+* cells – which in mouse reside in deep layer white

823 matter and make long-range projections. However, consistent with other reports[5], this type was

824 inconsistently observed across individual animals and regions, likely due to its laminar location

825 and low abundance, and was not analyzed further.)

826

827 For each species and each of the four main classes, transcripts were pooled across cells,

828 normalized by total number of transcripts, and scaled to 100k transcripts, which yielded four

829 vectors of representative gene expression for each class. We then applied a series of filters to

830 search all expressed genes for those that were selectively expressed by at least one of the four

831 cell types in at least one species. Genes with low expression (<10 transcripts per 100k in any

832 species) were removed. At least one species had to show a >3-fold difference between the

833 maximum and minimum expression level across the four types. These filters identified an initial

834 set of putative markers in one or more species. To search for genes that were consistent or

835 differed across species, for each gene (only one-to-one orthologues were considered), Pearson

35

836    correlations between pairs of species were computed, yielding six correlation values. If a gene

837    was not detected in a given species, values were set to 0 for pairs that included that species.

838

839    **Quantitative expression level comparisons across neocortical classes**

840    To examine the extent to which evolution has constrained a gene's quantitative expression level

841    across the four main neocortical cell classes, we focused on meaningfully expressed genes (4051

842    genes that exhibited at least 1.5-fold expression variation across the four classes in at least one

843    species, and were also present with an abundance of at least 10 transcripts per 100k in at least

844    one cell class). This gene list was intersected with a set of genes predicted to be intolerant of loss

845    of function (pLI > 0.9) across 60,706 humans[6], yielding a set of 1286 genes that were meaningfully

846    expressed in interneurons and showed evidence of intolerance to protein-truncating variants.

847    Pearson correlations were computed on the vector of expression values for each of these genes

848    in all possible pairs of species.

849

850    **Species comparisons of differential expression amongst pairs of cell types**

851    Integrated species analyses were performed using LIGER[4] between species pairs, which had the

852    advantage of allowing each pair of species to jointly determine cluster definition. Parameter values

853    were explored over a range using LIGER functions to suggest optimal values; the resultant

854    clusterings (e.g. Extended Data Fig. 2) used the following parameters: variance threshold = 0.15

855    (for inclusion of genes into LIGER alignment), k = 25 (number of factors), lambda = 5

856    (regularization parameter to penalize dataset-specific influence on alignment), resolution = 0.8

857    (controls resolution of clusters in community detection). For each cluster, the expression values

858    of meaningfully expressed genes (at least 10 transcripts per 100k in both species) were extracted

859    and fold differences for each gene was computed relative to each other cluster for each of the

860    species independently. These fold differences were then correlated across pairs of species.

861

862

**Neocortical regionally differentially expressed genes (rDEGs)**

864

To examine gene expression variation across neocortical regions in marmoset, region (n=7) datasets were pooled into a region-integrated LIGER analysis (variance threshold = 0.15, k = 25, lambda = 5, resolution = 0.8). These parameters produced 17 clusters; two clusters were removed for having fewer than 50 cells from one or more regions, yielding in a final set of 15 clusters for cross-region comparisons. For each cluster, differential expression was computed between all region pairs using a fold-difference threshold of 3.

*Interneuron rDEGs in astrocytes*

Marmoset neocortical astrocytes were analyzed by identifying the cluster(s) that expressed known astrocyte markers (e.g. *AQP4, GFAP, GJA1, GLUL*) from the same individuals used in interneuron analyses. Cells in these clusters were isolated from raw data and clustered using the ICA pipeline described above, which resulted in three astrocyte subtypes. For each astrocyte subcluster, fold differences of rDEGs identified in interneurons (in comparisons of PFC to V1) were computed for PFC and V1 in astrocytes.

*Marmoset interneuron rDEGs profiled in macaque, human and mouse*

Marmoset interneuron rDEGs (identified in comparisons of PFC to V1) were profiled in other species: for each rDEG, fold differences between frontal/prefrontal cortex and V1 cells were computed for each cluster identified by each species' ICA-based clustering. The percentage and median differential expression ($\log_{10}$-transformed fold differences) of genes that were rDEGs in marmoset and were also rDEG was calculated in each species.

*Spatial correlations*

888    For each cluster, expression levels of rDEGs identified in comparisons between PFC and V1 were

889    examined in the other neocortical regions (n = 5). To quantify the existence of spatial gradients,

890    for each gene the Pearson correlation between expression and spatial order along the anterior–

891    posterior axis was computed in the five remaining regions in order (Temp, S1, A1, Par, V2). Mean

892    correlation values were compared to a null distribution obtained by permuting the ordering (120

893    possible orderings).

894

895    **Hierarchical clustering**

896    Dendrograms of cell-type relationships were produced using hierarchical clustering (using the

897    hclust function, method = complete, in R) of genes normalized (to 100k transcripts), using $\log_{10}$

898    transformed values of all expressed genes (genes with at least 10 transcripts per 100k

899    transcripts).

900

901
902    **Single-molecule fluorescent *in situ* hybridization (smFISH)**

903

904    *Neocortex and hippocampus*

905    Frozen, unfixed tissue sections (12 µm) of mouse (P60-P70; Charles River, C57BL/6; n = 2),

906    marmoset (n = 2), and ferret (n = 1) brain tissue were cut on a cryostat (Leica CM 1950), adhered

907    to SuperFrost Plus microscope slides (Fisher Scientific, 12-550-15) and processed for three-color

908    smFISH using the ACD v2 RNAscope multiplexed fluorescence protocol for fresh frozen tissue.

909    Probes are listed in Extended Data Table 2. The ferret (*Mustela putorius furo*) was sourced from

910    Marshall Bioresources, and was used according to protocols approved by IACUC of Boston

911    Children's Hospital.

912

913    *In situ–based quantification of neocortical & hippocampal LAMP5+ subtypes*

914 For hippocampus and neocortical area S1, laminar boundaries were identified with DAPI stains

915 in mouse and marmoset tissue. Neocortical laminae were separated into five bins (layer 1, layer

916 2/3, layer 4, layer 5, layer 6); hippocampal laminae within CA1 and CA2 regions were separated

917 into four bins. Within each bin, *GAD1+*, *GAD1+/LAMP5+*, and *GAD1+/LAMP5+/LHX6+* cells were

918 counted in two sections of each replicate. In total, 3998 cells were counted.

919

920 *Single-molecule FISH (smFISH) in marmoset striatum*

921 One male marmoset (age = 6 years) was euthanized and perfused with ice-cold saline. The whole

922 brain was immediately removed, embedded in Optimal Cutting Temperature (OCT) freezing

923 medium, and flash-frozen in an isopropyl ethanol-dry ice bath. Samples were cut into 16 μm

924 sections on a cryostat (Leica CM 1850), adhered to SuperFrost Plus microscope slides (Fisher

925 Scientific, 12-550-15), and stored at -80°C until use.  Samples were immediately fixed in 4%

926 paraformaldehyde and stained on the slide according to the Advanced Cell Diagnostics

927 RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (ACD, 323100) protocol.  Samples were

928 stained for *VIP* (ACD, 554571-C2) and *NKX2-1* (ACD, 532751-C3) with antisense probes, and

929 coverslipped with Vectashield HardSet Antifade mounting medium with DAPI (Vector

930 Laboratories, H-1500).  Z-stack serial images were taken through the whole depth on a Nikon Ti

931 Eclipse inverted microscope with an Andor CSU-W1 confocal spinning disc unit and an Andor

932 DU-888 EMCCD using a 20×, 0.75 NA air objective, and later max-projected in FIJI (ImageJ,

933 NIH). Fields of view were randomly chosen across the whole striatal sample. Probes listed in

934 Extended Data Table 2.

935

936 **Fate mapping of Lamp5+/Lhx6+ cells in mouse hippocampus and neocortex**

937 To label *Lamp5+/Lhx6+* cells in the mouse neocortex and hippocampus, we utilized an

938 intersectional genetics approach. In mouse mature cortical interneurons, *Id2* and *Lamp5* are

939 expressed in nearly identical populations[7], and so we utilized an Id2-CreER driver line [8] (Jax

940     stock# 016222) in combination with an Nkx2.1-Flpo driver[9] (Jax stock# 028577) and the Cre/Flp

941     dependent tdTomato reporter Ai65[10] (Jax stock# 021875) to obtain selective labeling of

942     Lamp5/Lhx6 cells. Tamoxifen (20 mg/ml in corn oil) was administered to *Id2-CreER; Nkx2.1-Flpo;*

943     *Ai65* animals (3 × 5 mg by oral gavage over 5 days) between P30–P40 to activate the CreER,

944     after which animals were either perfused with 4% PFA/PBS and their brains processed for

945     immunohistochemistry (20-μm cryosections; tdTomato signal was enhanced using rabbit anti-

946     RFP from Rockland Immunochemicals; cat# 600-401-379), or acute brain slices were prepared

947     for morphological fills as described previously[11]. Fluorescence images were acquired on a Zeiss

948     Axio Imager.A1 and levels/contrast adjustments performed using Photoshop (Adobe).

949

950     **Ferret 10X Chromium Single Cell 3' v3**

951     A dataset of 801 interneurons was generated from ferret (P42) striatum. Single-nuclei

952     suspensions from frozen tissue was generated as for Drop-seq; GEM generation and library

953     preparation followed protocol #CG000183_ChromiumSingleCell3'_v3_UG_Rev-A. Sequencing,

954     alignment and clustering (using the ICA pipeline described above) proceeded as for the Drop-seq

955     datasets above.

956

**METHODS REFERENCES**

1.  Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174,** 1015–1030 (2018).
2.  Paxinos, G., Watson, C., Petrides, M., Rosa, M. & Tokuno, H. *The marmoset brain in stereotaxic coordinates*. (2012).
3.  Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
4.  Welch, J., Kozareva, V., Ferreira, A., bioRxiv, C. V.2018. Integrative inference of brain cell similarities and differences from single-cell genomics. *biorxiv.org* doi:10.1101/459891
5.  Hodge, R. D. *et al.* Conserved cell types with divergent features between human and mouse cortex. *biorxiv.org* (2018). doi:10.1101/384826
6.  Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
7.  Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature* **555,** 457–462 (2018).
8.  Rawlins, E. L., Clark, C. P., Xue, Y. & Hogan, B. L. M. The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development* **136,** 3741–3745 (2009).
9.  He, M. *et al.* Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. *Neuron* **91,** 1228–1243 (2016).
10. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85,** 942–958 (2015).
11. Schuman, B. *et al.* Four Unique Interneuron Populations Reside in Neocortical Layer 1. *J. Neurosci.* **39,** 125–139 (2019).