Brain-wide single neuron reconstruction reveals morphological diversity in molecularly defined striatal, thalamic, cortical and claustral neuron types

Hanchuan Peng[1,2,*], Peng Xie[1,2,#], Lijuan Liu[2,#], Xiuli Kuang[3], Yimin Wang[2,4], Lei Qu[2,5], Hui Gong[6,7], Shengdian Jiang[2], Anan Li[6,7], Zongcai Ruan[2], Liya Ding[2], Chao Chen[3], Mengya Chen[4], Tanya L. Daigle[1], Zhangcan Ding[2], Yanjun Duan[2], Aaron Feiner[1], Ping He[4], Chris Hill[1], Karla E. Hirokawa[1], Guodong Hong[2], Lei Huang[2], Sara Kebede[1], Hsien-Chi Kuo[1], Rachael Larsen[1], Phil Lesnar[1], Longfei Li[5], Qi Li[4], Xiangning Li[6,7], Yaoyao Li[3], Yuanyuan Li[5], An Liu[2], Donghuan Lu[8], Stephanie Mok[1], Lydia Ng[1], Thuc Nghi Nguyen[1], Qiang Ouyang[2], Jintao Pan[2], Elise Shen[1], Yuanyuan Song[2], Susan M. Sunkin[1], Bosiljka Tasic[1], Matthew B. Veldman[9], Wayne Wakeman[1], Wan Wan[5], Peng Wang[4], Quanxin Wang[1], Tao Wang[5], Yaping Wang[2], Feng Xiong[2], Wei Xiong[3], Wenjie Xu[1], Zizhen Yao[1], Min Ye[3], Lulu Yin[2], Yang Yu[1], Jia Yuan[2], Jing Yuan[6,7], Zhixi Yun[2], Shaoqun Zeng[6], Shichen Zhang[2], Sujun Zhao[2], Zijun Zhao[2], Zhi Zhou[1], Z. Josh Huang[10], Luke Esposito[1], Michael J. Hawrylycz[1], Staci A. Sorensen[1], X. William Yang[9], Yefeng Zheng[8], Zhongze Gu[2], Wei Xie[2], Christof Koch[1], Qingming Luo[6,11], Julie A. Harris[1], Yun Wang[1], Hongkui Zeng[1,*]

[1]Allen Institute for Brain Science, Seattle, WA 98109, USA

[2]SEU-ALLEN Joint Center, Institute for Brain and Intelligence, Southeast University, Nanjing, Jiangsu, China

[3]School of Optometry and Ophthalmology, Wenzhou Medical University, Wenzhou, Zhejiang, China

[4]School of Computer Engineering and Science, Shanghai University, Shanghai, China

[5]Key Laboratory of Intelligent Computation & Signal Processing, Ministry of Education, Anhui University, Hefei, Anhui, China

[6]Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, MoE Key Laboratory for Biomedical Photonics, Huazhong University of Science and Technology, Wuhan, Hubei, China

[7]HUST-Suzhou Institute for Brainsmatics, JITRI Institute for Brainsmatics, Suzhou, Jiangsu, China

[8]Tencent Jarvis Lab, Shenzhen, Guangdong, China

[9]Center for Neurobehavioral Genetics, Jane and Terry Semel Institute for Neuroscience and Human Behavior; Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[10]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[11]School of Biomedical Engineering, Hainan University, Haikou, Hainan, China

[#]Equal contribution

*Correspondence: Hanchuan Peng (h@braintell.org) or Hongkui Zeng (hongkuiz@alleninstitute.org)

ABSTRACT

Ever since the seminal findings of Ramon y Cajal, dendritic and axonal morphology has been recognized as a defining feature of neuronal types. Yet our knowledge concerning the diversity of neuronal morphologies, in particular distal axonal projection patterns, is extremely limited. To systematically obtain single neuron full morphology on a brain-wide scale, we established a platform with five major components: sparse labeling, whole-brain imaging, reconstruction, registration, and classification. We achieved sparse, robust and consistent fluorescent labeling of a wide range of neuronal types by combining transgenic or viral Cre delivery with novel transgenic reporter lines. We acquired high-resolution whole-brain fluorescent images from a large set of sparsely labeled brains using fluorescence micro-optical sectioning tomography (fMOST). We developed a set of software tools for efficient large-volume image data processing, registration to the Allen Mouse Brain Common Coordinate Framework (CCF), and computer-assisted morphological reconstruction. We reconstructed and analyzed the complete morphologies of 1,708 neurons from the striatum, thalamus, cortex and claustrum. Finally, we classified these cells into multiple morphological and projection types and identified a set of region-specific organizational rules of long-range axonal projections at the single cell level. Specifically, different neuron types from different regions follow highly distinct rules in convergent or divergent projection, feedforward or feedback axon termination patterns, and between-cell homogeneity or heterogeneity. Major molecularly defined classes or types of neurons have correspondingly distinct morphological and projection patterns, however, we also identify further remarkably extensive morphological and projection diversity at more fine-grained levels within the major types that cannot presently be accounted for by preexisting transcriptomic subtypes. These insights reinforce the importance of full morphological characterization of brain cell types and suggest a plethora of ways different cell types and individual neurons may contribute to the function of their respective circuits.

INTRODUCTION

Understanding the taxonomic organization of cell types in the brain will yield insight into its fundamental components and landscape, providing an atlas much like the periodic table of elements in chemistry and the taxonomy of living species in biology. As neurons exhibit extraordinary diversity across molecular, morphological, physiological, and connectional features, a complete and accurate classification and creation of a cell type atlas needs to consider and integrate these distinct cellular properties [1]. Recent advances in high-throughput single cell RNA-sequencing has enabled the systematic classification of cell types at the transcriptomic level [2-6]. This approach captures major cell types with known anatomical and functional properties, but also reveals many potential new cell types. Systematic classification of cortical neurons using a combination of local morphological and electrical properties has also been achieved [7-9], and new technologies have been developed to correlate morpho-electrical and transcriptomic cell types [10,11]. Although dendritic and long-range axonal morphologies have long been held as the central defining feature of neuronal types [12], so far very few tools are available and little systematic effort has been made to classify individual single neurons using brain-wide axonal projection patterns alone or with other cellular properties.

Brain-wide inter-areal connectivity has been mapped extensively using injections of anterograde and retrograde tracers to label populations of projection neurons [13-18]. However, it remains largely unknown how population-level projection patterns are reflected at the single neuron level. Triple retrograde tracing studies suggest that individual neurons within a brain region often have heterogeneous axonal projection patterns [19-21]. Thus, characterizing single neuron axonal projections through reconstruction of complete morphologies not only provides critical information related to classifying cell types, but also how neural signals are organized and transmitted to their target regions.

Despite its importance, data on single neuron axonal morphologies are currently lacking for most projection neuron types in mammals, in large part because axons often cover large distances and are severed in *ex vivo* brain slices. Previous efforts have been made in rodents to fully label single neurons with small molecules or fluorescent proteins through *in vivo* whole-cell patching, *in vivo* electroporation [22-24], sparse transgenic labeling [25], or sparse viral labeling with sindbis virus [26-28] or adeno-associated virus (AAV) [29-31]. Conventionally this is followed by serial sectioning, imaging of each section, and manual reconstruction of the labeled neurons across many consecutive sections. Although relatively few such studies exist due to the labor-intensive process, they reveal critical features of specific projection neuron types that likely have important functional implications [22,24,32-39]. The recent development of high-throughput and high-resolution fluorescent imaging platforms, such as fMOST [40] and MouseLight [29], coupled with more efficient sparse viral labeling strategies, now enable the potential for large-scale generation of neuronal morphology datasets. These studies also revealed a need for further improvements in tools for generating very sparse and strong labeling of single neurons at a brain-wide scale, as well as computational tools to expedite the laborious reconstruction process.

As part of the BRAIN Initiative Cell Census Network (BICCN) efforts to characterize brain cell types across multiple modalities, we have established a pipeline to label, image, reconstruct and classify single neurons at a brain-wide scale using complete morphology data. As a foundational step toward this goal, we report here the largest set of single neuron reconstructions to this date, 1,708 neurons from striatum, thalamus, cortex, claustrum and other brain regions in mice, as a multi-site collaborative effort. These neurons are labeled by cell class or type selective Cre driver lines, enabling correlation of their morphologies and projection patterns with their molecular identities. We also provide a corresponding set of single-cell RNA-seq data from retrogradely labeled neurons (i.e. Retro-seq) to corroborate our findings. Overall, our study reveals substantial morphological and projection diversity of individual neurons; this diversity is governed by underlying rules that are region- and cell type-specific. The results demonstrate that full morphology reconstruction is an essential component of cell type characterization, highly complementary to other molecular, anatomical and physiological approaches in order to gain a full understanding of cell type diversity. To facilitate and encourage further work in this key area all imaging data and computational reconstruction tools are made publicly available. Our ultimate goal is to enable and encourage a community-based effort to generate a sufficiently large set of full morphology reconstructions, potentially tens to hundreds of thousands, to facilitate cell type classification, comparison with molecular, physiological and other cellular properties and understanding the functional implications of single neurons.

RESULTS

128 **Sparse, robust and consistent neuronal labeling**

129 Full reconstruction of a single neuron requires sparse, robust and consistent neuronal labeling.
130 Previous genetic approaches applied to produce sparse or single neuron labeling, including viral
131 delivery (*e.g.*, using sindbis virus or AAV) and *in vivo* electroporation [22-24,26-31,37], often resulted
132 in substantial cell-to-cell and animal-to-animal variations, and were usually restricted to few
133 brain regions. In rare cases, extremely sparse labeling was obtained utilizing the low
134 recombination efficiency of CreER driver lines and a sensitive, alkaline phosphatase-based
135 histochemical reporter [25,41]. To achieve more efficient, widespread, consistently sparse yet strong
136 labeling, we utilized TIGRE2.0 transgenic reporter lines that exhibit viral-like transgene
137 expression levels [42,43], coupling them with Cre expression from either driver lines or viral
138 delivery. We employed two general approaches.

139 The first was to use the GFP expressing Ai139 or Ai140 TIGRE2.0 reporter line in conjunction
140 with sparse Cre-mediated recombination (**Fig. 1a**). We used CreERT2 driver lines and titrated
141 the level of CreERT2-mediated recombination using low-dose tamoxifen induction
142 (**Supplementary Table 1**). We found optimal tamoxifen doses for sparse labeling in each case
143 using serial two photon tomography (STPT) to quickly screen for brain-wide transgene
144 expression [16] (**Extended Data Fig. 1a-c**). We also tested a dual reporter combination with
145 TIGRE2.0 (Ai140) and TIGRE1.0 (Ai82) [44] reporter lines, using the tTA2 from the single
146 TIGRE2.0 allele to drive two copies of the TRE promoter driven GFP expression cassettes (**Fig.**
147 **1a,** "optional" cross). We found this strategy generated an even higher level of GFP expression
148 than Ai139 or Ai140 alone, well suited for fMOST imaging (see below).

149 The second approach was to use a new TIGRE2.0 Cre reporter line: TIGRE-MORF (also called
150 Ai166) [43], which is sparsely activated in conjunction with Cre delivery (**Fig. 1b**). TIGRE-
151 MORF/Ai166 expresses the MORF gene, which is composed of a farnesylated EGFP (GFPf)
152 preceded by a mononucleotide repeat of 22 guanines (G_{22}-GFPf). The GFPf transgene is not
153 translated at the baseline due to the out-of-frame G_{22} repeat relative to the open reading frame of
154 GFPf, which lacks its own translation start codon. However, during DNA replication or repair,
155 rare events of stochastic frameshift of the mononucleotide repeat result in correction of the
156 translation frame (*e.g.*, G_{22} to G_{21}) and produce expression of the GFPf protein in a small subset
157 of cells. TIGRE-MORF/Ai166 (below simplified as Ai166) exhibits a labeling frequency of 1-
158 5% when crossed to different Cre driver mouse lines [43]. Even with this frequency, we found that
159 combining Ai166 with many Cre driver lines densely expressing the Cre transgene did not
160 produce sufficient sparsity to readily untangle the axonal ramifications, whereas combining it
161 with Cre lines that are already relatively sparse to begin with, or with CreERT2 lines with
162 intermediate dosing level of tamoxifen (**Supplementary Table 1**), leads to extremely sparse
163 labeling well suited for reconstruction of elaborate axonal arborizations of many neuronal types
164 (**Extended Data Fig. 1d-j**). The use of membrane associated GFPf also enabled robust labeling
165 of very thin axon fibers. Leaky background expression of GFP reported in other TIGRE2.0 lines
166 [42] is not present in Ai166 mice due to the strict dependency of translational frameshift for the
167 expression of GFPf reporter, making TIGRE-MORF/Ai166 an ideal reporter line for sparse and
168 strong labeling of various neuronal types across the brain.

169

170 **High quality imaging data reveals diverse neuronal morphologies**

171 For this study, we generated 53 high-quality (*i.e.*, strong, even and not-too-dense labeling)
172 fMOST-imaged brain datasets with sparsely labeled cells in cortical, thalamic, claustral, and
173 striatal regions, and for cholinergic, noradrenergic and serotonergic neuronal types (**Fig. 1c-h**,
174 **Extended Data Fig. 2, Supplementary Table 1**). Critically, our approach can be extended to
175 any cell type for which appropriate Cre-dependent labeling methods are available. In the cortex,
176 we imaged from different excitatory projection classes using selective CreERT2 driver lines
177 [4,42,45]. For example, Cux2-CreERT2;Ai166 labeled the cortical layer (L) 2/3/4 intratelencephalic
178 (IT) subclasses of excitatory neurons (**Fig. 1c**). Plxnd1-CreER;Ai166 labeled cortical L2/3 and
179 L5 IT subclasses, as well as striatal medium spiny neurons (MSN, **Fig. 1e**). Fezf2-CreER;Ai166
180 labeled cortical L5 extratelencephalic (ET, also known as pyramidal tract, PT) subclass (**Fig.**
181 **1d**). Tle4-CreER;Ai166 labeled L6 corticothalamic (CT) subclass (**Extended Data Fig. 2a**).
182 Nxph4-T2A-CreERT2;Ai166 labeled cortical L6b subplate neurons (**Extended Data Fig. 2b**). In
183 the cortex, we also labeled and imaged *Pvalb*+ cells, which includes a subclass of inhibitory
184 interneurons, *e.g.*, basket cells (BC), and a subset of L5 ET excitatory neurons, using Pvalb-T2A-
185 CreERT2;Ai166 (**Fig. 1f**) and *Sst*+ interneurons using Sst-Cre;Ai166 (**Extended Data Fig. 2c**).
186 In the thalamus, we used Tnnt1-IRES2-CreERT2;Ai140;Ai82 and Vipr2-IRES2-Cre-neo;Ai166
187 to label excitatory projection neurons as well as striatal MSNs (**Fig. 1g,h**). Vipr2-IRES2-Cre-
188 neo;Ai166 also labeled axons consistent with projections from retinal ganglion cells [46], as well as
189 cortical chandelier cells (ChC) (**Fig. 1h**). Gnb4-IRES2-CreERT2;Ai140;Ai82 labeled the *Car3*+
190 IT subclass of L6 excitatory neurons in cortex and claustrum (CLA, **Extended Data Fig. 2d**).
191 Cell types containing neuromodulators were also labeled with selective Cre driver lines [47],
192 including noradrenergic neurons in the locus ceruleus (LC) using Dbh-Cre_KH212;Ai166
193 (**Extended Data Fig. 2e**) and serotonergic neurons in the dorsal raphe (DR) and other brainstem
194 regions using Slc6a4-CreERT2;Ai166 (**Extended Data Fig. 2f**).

195 It is apparent that these neurons display a remarkable array of dendritic and axonal
196 morphologies. Specifically, in these sparsely labeled brains, cortical IT and ET neurons not only
197 have primary long-range projections but also local axonal branches that are well segregated and
198 clearly identifiable, enabling truly complete reconstruction of the entire local and long-range,
199 cortical and subcortical axonal arborization (**Fig. 1c-e**). L5 ET neurons form the 'driving' type of
200 synapses in the thalamus [48,49], which have enlarged and intensely fluorescent boutons (**Fig. 1f**).
201 L6b subplate neurons extend their local axon collaterals upwards into layer 1 (**Extended Data**
202 **Fig. 2b**). The axons of thalamic projection neurons form either dense or dispersed clusters in the
203 cortex (**Fig. 1g-h**). On the other hand, claustral, noradrenergic and serotonergic neurons have
204 widely dispersed, thin axons that are nonetheless well labeled (**Extended Data Fig. 2d-f**). One
205 can also clearly see individual axons in the substantia nigra from striatal medium spiny neurons
206 (**Fig. 1e,g**), individual axon terminal clusters in the superficial layers of the superior colliculus
207 likely coming from retinal ganglion cells (**Fig. 1h**), as well as the dense and fine local axonal
208 branches of a variety of cortical and striatal interneurons (*e.g.*, basket cells, chandelier cells, and
209 Martinotti cells) (**Fig. 1f,h** and **Extended Data Fig. 2c**).

210 Of note, sparsely labeled neurons were frequently observed in other regions of the brain for all of
211 these crosses but are not described in detail here. Each of these brains contains ~100-1,000
212 labeled neurons (**Supplementary Table 1**). Thus, tens of thousands of neurons could be
213 reconstructed from these and newly generated datasets in the coming years. The whole brain
214 image series are publicly available through the BICCN web portal (https://biccn.org/) as a unique
215 resource for the community.

Pipeline for image data processing, morphology reconstruction and registration

We acquired whole brain images with sufficient resolution (~0.3 x 0.3 x 1 μm XYZ) for reconstructing fine-caliber axons using fluorescence micro-optical sectioning tomography (fMOST), a high-throughput, high-resolution, brain-wide fluorescent imaging platform [40]. To handle the large imaging datasets generated, we established a standardized image data processing and informatics workflow (**Fig. 2a**) for efficient whole brain morphology reconstruction utilizing Vaa3D, an open-source, cross-platform visualization and analysis system [50,51]. Each fMOST dataset is first converted to a multi-level navigable dataset using the Vaa3D-TeraFly program [52], which allows smooth handling of terabyte-scale datasets. Neuron visualization and reconstruction is then carried out on the TeraFly files. A series of tools, especially those based on the "Virtual Finger" method [53], were developed within Vaa3D to facilitate semi-automated and manual reconstruction. Further, a virtual reality (VR) environment created within Vaa3D, TeraVR, significantly enhances a user's ability to see the 3D relationships among intertwined axonal segments, improving precision and efficiency of reconstruction [54]. After quality control (QC) and manual correction, we used Vaa3D's deformable model to automatically fit the tracing to the center of fluorescent signals. The final reconstructed morphology was completed as a single tree without breaks, loops, or trifurcations. All these data processing, reconstruction, and workflow control processes were managed using a newly designed software system for massive scale data production (Jiang et al, manuscript in preparation).

In parallel, each fMOST dataset was registered to the 3D Allen mouse brain Common Coordinate Framework (CCFv3, http://atlas.brain-map.org/) [55], using a newly developed mBrainAligner program (Methods; Qu et al, manuscript in preparation) specifically designed for fMOST datasets to handle the challenges of brain shrinkage and deformation related to modality-specific technical protocols (**Fig. 2b**, **Extended Data Fig. 3**). Following registration of the whole-brain image dataset, all individual neuron reconstructions were also registered to the CCFv3 using the source brain's transformation parameters (**Fig. 2b**). Registration to CCFv3 enables digital anatomical delineation and spatial quantification of each reconstructed morphology and its compartments (*e.g.*, soma, dendrites, axon arbors). Since neurons are reconstructed from different brains, co-registration to the CCFv3 allows them to be compared and analyzed using a unified framework, mBrainAnalyzer (**Fig. 2b, Extended Data Fig. 3**), which automatically detects the arbors of each neuron followed by mapping of these dendritic and axonal arbors onto the standardized CCFv3 space. Using these tools and following this workflow, we reconstructed the full 3D morphology of 1,708 neurons from cortex, claustrum, thalamus, striatum and other regions (**Supplementary Table 2**). Quantification of all detected axon arbors, shown with respect to the anatomical locations of their somas, provides an informative global view of the brain-wide projection patterns for these single neurons (**Fig. 2c**). Detailed analysis of this brain-wide projection arbor map to understand the morphological diversity of cell types sheds light on how to perform more comprehensive analysis of cell types with finer resolution and precision.

We established a stringent QC process that includes ensuring the completeness of reconstructed morphologies (**Extended Data Fig. 4**). A conventional way to assess the completeness of axon labeling and reconstruction is whether an axon ends at a bouton (indicated by an enlargement with more intense signal (see arrowheads in **Fig. 1**) or gradually tapers off, the former suggesting a complete labeling [36]. We implemented this assessment in our reconstruction refinement process

261 to identify potential inaccuracies (**Extended Data Fig. 4c-e)**. In our final QC-passed
262 reconstructions we found that the ratio between terminal axon branches with and without a
263 terminal bouton was about 10:1, indicating a high degree of completeness of our reconstructed
264 morphologies.

265 We calculated a set of morphological features (Methods) for every neuron, focused on soma
266 location, dendrite and axonal properties. We also used a graph-partition approach to
267 automatically detect the major domains of arborization for each neuron (Methods).
268 Morphological features such as length, depth, area, etc., at the whole neuron level were also
269 computed for each arbor-domain for analysis. In the following sections, we use a combination of
270 these metrics to classify and analyze single neurons in several major brain divisions.

271

272 **Striatal medium spiny neurons exhibit converging projections while retaining topography**

273 We reconstructed 305 neurons in the dorsal striatum (caudate putamen, CP) from 4 Cre driver
274 lines: Tnnt1, Plxnd1, Vipr2 and Pvalb (**Fig. 3a, Supplementary Table 2**). These neurons can be
275 divided into 3 groups based on their projection targets: (1) those with main axon projections
276 terminating in the external segment of the globus pallidus (GPe, n=179 cells), (2) those
277 terminating in the reticular part of the substantia nigra (SNr, n=96 cells) and (3) those
278 terminating within striatum itself (others, n=30 cells). These groups of neurons are distributed
279 throughout CP and largely intermingled. Axonal projections from groups 1 and 2 correspond to
280 the two well-known types of striatal medium spiny neurons, dopamine receptor D1 (*Drd1*)
281 neurons projecting to SNr and dopamine receptor D2 (*Drd2*) neurons projecting to GPe [56].

282 Individual striatal neurons projecting to GPe or SNr exhibit a simple point-to-point morphology:
283 GPe-projecting neurons have one major axon arbor, targeting GPe; SNr-projecting neurons have
284 one major axon arbor targeting SNr. The GPe type has more elaborate axon arborization near the
285 soma. Sholl analysis shows that the number of local crossings (<1 mm to soma) of the GPe type
286 is 2.8 times of that of the SNr type. The radius of local axons of the GPe type is also 2.5 times as
287 large. Most SNr-projecting neurons send minor branches targeting the internal segment of globus
288 pallidus (GPi) or GPe.

289 The dominant feature of both types of striatal neurons is convergent projection within the main
290 target region, GPe or SNr, consistent with the ~20-fold smaller sizes/volumes of these regions
291 compared to the dorsal striatum (**Fig. 3b,c**). We find that both the center-to-center distance and
292 the degree of overlap of axon arbors between each pair of neurons are proportional to that pair's
293 soma-to-soma distance (**Fig. 3c**), indicating a regular spatial organization of these neurons' axon
294 projections. The axon arbor distances between striatal neurons within the same type are
295 substantially smaller and their overlapping scores are substantially greater in comparison to those
296 of neurons from each of the various thalamic nuclei (see next section) (**Fig. 3d**). Furthermore,
297 axon arbors in GPe or SNr can be grouped into domains based on the degree of overlap; these
298 domains are arranged topographically and correspond to the topographic localization of the
299 somas in striatum (**Fig. 3e**).

300

301 **Thalamic neurons exhibit diverse and nucleus-specific thalamocortical projection patterns**

302 We reconstructed 735 thalamic neurons from 3 Cre lines: Tnnt1, Vipr2 and Pvalb
303 (**Supplementary Table 2**). Those (n=17 neurons) from Pvalb-Cre are mostly GABAergic

304 neurons and in the reticular nucleus (RT) projecting back to other thalamic nuclei. Tnnt1 and
305 Vipr2 Cre lines contained mostly thalamocortical projection neurons from many (but not all)
306 nuclei, including sensory-motor relay nuclei (n=658 cells) and higher-order or associational
307 nuclei (n=60 cells). In this dataset, the reconstructed cells covered 22 of the 44 thalamic regions
308 in CCFv3, which can be broadly divided into two major groups [49,57,58]: (1) primary sensory or
309 motor relay nuclei, also known as "core" or "driver" nuclei, which include the ventral
310 posteromedial nucleus (VPM), ventral posterolateral nucleus (VPL), ventral posteromedial
311 nucleus parvicellular part (VPMpc), ventral posterolateral nucleus parvicellular part (VPLpc),
312 dorsal part of the lateral geniculate complex (LGd), medial geniculate complex (MG) and ventral
313 anterior-lateral complex (VAL), as well as anterior and medial nuclei such as anteromedial
314 nucleus (AM), submedial nucleus (SMT) and the posterior part of the mediodorsal nucleus (MD)
315 (**Extended Data Fig. 5**); (2) associational or higher-order thalamic nuclei, also known as
316 "matrix" or "modulatory" nuclei, which include the lateral posterior nucleus (LP), posterior
317 complex (PO), lateral dorsal nucleus (LD), ventral medial nucleus (VM), nucleus of reuniens
318 (RE), central medial nucleus (CM), interanterodorsal nucleus (IAD), anterior part of MD, and
319 paraventricular nucleus (PVT) (**Extended Data Fig. 6**). In general, "core-type" thalamocortical
320 projections target one or a small number of cortical regions and predominantly terminate in L4.
321 In contrast, the "matrix-type" thalamocortical projections generally target multiple cortical areas
322 with dense axonal terminals in L1.

323 As different thalamic regions have unique cortical projection patterns [16,57,59], we first compared
324 projections of individual thalamic neurons with projections of anterograde bulk tracing from the
325 Allen Mouse Connectivity Atlas (**Supplementary Table 3**). We find that single cell projections
326 are basically consistent with the bulk anterograde tracing results for the thalamic nuclei their
327 somas reside in (**Fig. 4a-h**, left two columns).

328 Also consistent with prior knowledge, we find that the single sensory-motor relay neurons
329 usually have one major axon arbor targeting the primary sensory or motor cortex of the
330 corresponding modality, i.e. VPM and VPL to primary somatosensory area (SSp), VPMpc to
331 gustatory areas (GU), VPLpc to visceral area (VISC), LGd to primary visual area (VISp), MG to
332 auditory areas (AUD), and VAL to primary motor area (MOp). Axons from these nuclei
333 terminate predominantly in L4 and L6, consistent with the "core-type" classification (**Fig.
334 4a,b,d, Extended Data Fig. 5**). Reconstructed neurons from AM, SMT and posterior MD also
335 send a single major axon arbor to various parts of orbital area (ORB), with a similar mid-layer
336 termination pattern, suggesting these neurons also belong to the "core" projection type (see also
337 Harris et al. 2019) (**Fig. 4c,e, Extended Data Fig. 5**).

338 We quantitatively analyzed morphometric features of 1,103 axon arbors from 624 neurons
339 located in the primary sensory thalamic nuclei: VPM, VPL, LGd and MG (**Extended Data Fig.
340 7**). Our results suggest there are two major types of axon arbors targeting cortical layer 4
341 (spanning L4 and lower L2/3); a smaller "type 1" arbor and a larger "type 2" arbor (cortical
342 area >0.3 mm^2). We also identified a "type 3" arbor terminating in cortical L6; this type was
343 most often a minor branch originating from the type 1 or type 2 arbors, so we did not use it to
344 classify neurons. Single neurons were thus assigned to either small-arbor or large-arbor type
345 (**Fig. 4a,b,d**). The ratio of large- to small-arbor neurons varies by thalamic region. Large-arbor
346 neurons account for 28.3% of the total reconstructions from VPM, 28.0% from VPL, 46.8%
347 from LGd and 58.3% from MG. Notably, somas with small and large arbors are spatially
348 intermingled in each nucleus (**Extended Data Fig. 7e**). Neurons located in SMT are also

separable into small- and large-arbor types like the primary relay nuclei (**Fig. 4c**), whereas the current set of AM neurons all have small arbors and the posterior MD neurons all have large arbors (**Fig. 4e, Extended Data Fig. 5**).

A small fraction of the core-type thalamic projection neurons have more than one axon arbor targeting different cortical areas (6.77% for VPM, 14.44% for VPL, 12.5% for MG, but 0% for LGd). In the case of these cells in VPM and VPL, usually they have a larger main arbor targeting SSp, and a smaller secondary arbor targeting the supplemental somatosensory area (SSs) (**Fig. 4a,b**). MG neurons with two or more cortical targets are mostly of the large-arbor type (**Fig. 4d**). These multi-target MG neurons are more like the matrix type thalamocortical neurons described below, showing stronger projections to L1 and L5, located in the associational parts of MG (*e.g.*, MGm) medial to the core relay auditory nucleus, MGv.

Outside the sensory-motor relay thalamic nuclei, nearly all reconstructed neurons have a large diffusely branched axon arbor and/or several arbors projecting to multiple cortical areas, often with columnar or L5-dominant axon termination patterns (**Fig. 4e-h, Extended Data Fig. 6**). Many (81%) of these cells also have axon branches >1 mm long in L1, consistent with the "matrix" type, but they also exhibit a diverse range of projection and morphological patterns. For example, LP neurons preferentially project to two or more higher visual cortical areas. They do not directly project into VISp, only one out of 16 reconstructed LP neurons has axon fibers in L1 extending into VISp. LP neurons can be roughly divided into an anterior and a posterior group, consistent with previous functional studies [60]. Posterior LP neurons mainly project to lateral and posterior higher visual areas, whereas anterior LP neurons mainly project to medial and anterior higher visual areas with some extending an axon projection into anterior cingulate area (ACA) (**Fig. 4f, Extended Data Fig. 6**). PO neurons project to both SSp and MOp/MOs (secondary motor area). Their axon arbors in these target regions terminate broadly across layers with an apparent preference in lower L2/3, with 5 out of 7 sending rich axon arbors (>1 mm) to L1 (**Fig. 4g, Extended Data Fig. 6**). Neurons in anterior MD appear very different from those in posterior MD and are more similar to those in neighboring nuclei such as IAD and CM. They have multiple axon arbors that target multiple medial and lateral prefrontal cortical areas including prelimbic area (PL), ORB and agranular insular area (AI) (**Fig. 4e, Extended Data Fig. 6**). VM neurons have multiple axon arbors, heavily targeting MOp/MOs with additional branches targeting various somatosensory areas (**Fig. 4h, Extended Data Fig. 6**).

Quantitative interareal projection matrix (**Extended Data Fig. 8, Supplementary Tables 2 and 3**) confirms the above observations and further demonstrates the distinction between core- and matrix-type neurons, with the former (from VPM, VPL, VPMpc, VPLpc, LGd, VAL, AM, and SMT) predominantly targeting a single cortical area (sometimes with a secondary area) and the latter (MG, MD, LP, PO, LD, VM, RE, and CM) targeting multiple cortical areas. Note that MG contains a mixture of core and matrix cells. Within each nucleus, on the other hand, individual neurons show a high degree of consistency with each other in projection patterns, and each neuron's projection pattern is also similar to the population projection pattern for that nucleus.

Cortical L2/3/4 intratelencephalic (IT) neurons have variable intracortical projection targets

We reconstructed 160 cortical neurons in total from the Cux2-CreERT2 line, and further analyzed 93 of these neurons in somatosensory and motor regions: SSp (n=40), SSs (n=15),

MOp (n=20) and MOs (n=18) (**Fig. 5a, Supplementary Table 2**). *Cux2*+ neurons are located in L2/3 or L4 and their long-range projections are confined within cortex and striatum, consistent with them belonging to the corticocortical projecting IT subclass [61]. Clustering analysis using local dendritic and axonal morphological features grouped these cells largely by region and layer (**Fig. 5b**).

We identified 25 cells from SSp and 3 cells from SSs to be in L4. These cells were distinguished from L2/3 neurons using morphological features; they have either no apical dendrites (*i.e.*, spiny stellate cells) or a simple apical dendrite that does not branch in L1 (*i.e.*, untufted or star pyramid cells), in contrast to the pyramidal L2/3 cells which have tufted apical dendrites in L1 [61] (**Fig. 5h-i, Extended Data Fig. 9**). L2/3 cells also have local axons in L2/3, which also project downward into L5, whereas L4 cells have local axons mainly projecting up to L2/3, another differentiating feature [62]. Interestingly, we also found 4 neurons from MOp/MOs with these L4-like features – minimal apical dendrites and upward-projecting local axons (**Fig. 5f-g**), and they are located between L2/3 and L5 since L4 is not delineated in CCFv3, suggesting that these are the L4-like cells located in motor cortex [63] that can also be identified transcriptomically (see flagship paper).

Several regional differences in axon projections of these L2/3/4 IT neurons are observed. The proportion of cells with intracortical long-range projections (defined as having >5 mm axon branches that are >1 mm away from the soma) is larger in SSs, MOp and MOs than SSp (**Fig. 5c**). Consistent with prior notion [61], all but one SSp L4 cells have only local axons but no long-range projections (**Extended Data Fig. 9**). However, nearly all L4 and L4-like cells in SSs, MOp and MOs do have axon projections outside of their local area (**Fig. 5f-i**), as we reported before [57]. For the cells with long-range projections, the average total length of long-projecting axons and the number of target regions per cell are also greater in MOp/MOs than in SSp/SSs ($p < 0.003$ for axon length and <0.002 for target regions, Mann-Whitney rank test).

Several recent studies have shown that transcriptomically defined cortical IT neurons are organized by layer, but also exhibit a gradual transition in transcriptomic type spatially along the cortical depth, by scRNA-seq [64], MERFISH [65] and Patch-seq [66]. Here we arrange the L2/3/4 IT cells according to the depth of their soma from the pial surface, region by region, and examine their local dendritic and axonal morphologies and long-range projections (**Fig. 5f-i**). We find that within each region, across depths individual neurons exhibit variable long-range projection patterns with no obvious depth correspondence, suggesting that long-range projection patterns of these L2/3/4 IT neurons may not correlate with their transcriptomically defined subtypes.

Clustering based on long-range projection patterns grouped the cells into 4 clusters (T1-T4), although variation appears largely continuous (**Fig. 5d**). We did not observe a clear one-to-one correspondence between these long-projecting clusters and either local morphology clusters (S1-S4) or soma areal origin (**Fig. 5e**). Comparing single neuron and population projection patterns (**Supplementary Table 3**) shows that all neurons together recapitulate the population projection pattern, but each neuron selects a subset of projection targets (**Fig. 6a-b, d-e, g-h**), in contrast with the thalamic neurons reported above. Thus, the selection of a subset of projection targets by each neuron appears random without specific correlation to soma depths or dendritic morphologies.

We next asked if axon termination patterns of single cells recapitulate the overall feedforward or feedback projection patterns apparent at the population level [57]. Given that each L2/3/4 IT cell

437 projects to only a subset of their intracortical targets, we pooled all the terminal axon arbors to a
438 specific target from all the cells within a source region (**Fig. 6c,f,i**). Interestingly, we find that
439 nearly all axon terminals of SSp cells are concentrated in middle layers (L2/3-5) of all target
440 regions (MOs, MOp, SSs and VISC, **Fig. 6i)**, suggesting all outward projections from SSp are
441 feedforward, consistent with it being at the bottom of the cortical hierarchy [57]. On the other hand,
442 axons of MOp cells mainly terminate in L2/3-5 of target areas MOs and contralateral MOp, but
443 have prominent termination in L1 of target areas SSp and SSs (**Fig. 6f**), consistent with the
444 notion that the former projections are feedforward and the latter feedback. The MOs cells do not
445 show such a clear division with the available number of terminal axon arbors to different targets
446 (**Fig. 6c**).

447 Finally, we investigated the projection target specificity of transcriptomic cell types using Retro-
448 seq [4], in which the transcriptomes of 822 retrogradely labeled neurons from SSp, SSs, MOp and
449 MOs were mapped to our transcriptomic taxonomy [64] to identify the transcriptomic type of each
450 retrogradely labeled neuron (**Extended Data Fig. 10, Supplementary Table 4**). We found that
451 for each source region, L2/3 and L4/5 IT neurons labeled from different injection targets were
452 mostly mapped to a few common transcriptomic types. Taken together, the above results suggest
453 that within the L2/3/4 IT subclasses, the projection patterns at a single cell level do not correlate
454 one-to-one with the cell's transcriptomic type in the adult animal.

455

456 **Cortical L5 extratelencephalic (ET) neurons show distinct subcortical projection specificity**

457 We reconstructed 251 L5 ET neurons in total from two Cre lines: Fezf2 and Pvalb, and analyzed
458 more than half of these neurons from the SSp (n=130), SSs (n=19), MOp (n=13) and MOs (n=4)
459 (**Extended Data Fig. 11, Supplementary Table 2**). We first clustered the cells into four groups
460 based on their local dendritic and axonal features (**Fig. 7a-b**). Cluster S1 is the most unique;
461 these neurons have larger dendritic and local axonal arbors and nearly all MOp and MOs neurons
462 belong to this cluster. Clusters S2-S4 are separable by features that capture the complexity of
463 local axons (*e.g.,* length, total area containing axon, and axon lamination).

464 The L5 ET neurons have 11.3 projection targets (defined using a stringent threshold of axon
465 length >1 mm in the target region) on average, significantly more than the L2/3/4 IT neurons
466 which have 8.0 projection targets on average ($p<2.0e-6$, Mann-Whitney rank test). Like L2/3/4
467 IT neurons, L5 ET neurons exhibit extensive diversity and heterogeneity in each of their selected
468 subset of projection targets. Nonetheless, we clustered these cells based on their long-range
469 projections within each region (MOp/MOs combined, SSp, and SSs) and identified subtypes with
470 specific projection targets (**Fig. 7c-f**).

471 L5 ET neurons in MOp/MOs form 4 distinct projection types (**Fig. 7d,g**). Notably, MO3 and
472 MO4 types project to regions in the medulla whereas MO1 and MO2 types do not, a distinction
473 consistent with recent findings from combined transcriptomic/epigenomic and projection studies
474 [32,67]. Note that few SSp neurons and no SSs neurons project to the medulla (**Fig. 7e,f**), suggesting
475 medulla projection may be primarily a feature of MOp/MOs neurons. Within the medulla-
476 projecting types, MO4 neurons have few projections to other parts of the brain such as thalamus
477 and basal ganglia compared to MO3 neurons. MO4 neurons also have stronger preference to
478 project to the contralateral side of medulla. Within the non-medulla-projecting types, MO1 and
479 MO2 neurons again display differential projection patterns in the thalamus, with MO1
480 preferentially targeting VM/VAL/MD whereas MO2 preferentially target PO.

481 L5 ET neurons in SSp can also be divided into 4 clusters based on their projections (**Fig. 7e,h**).
482 The 4 clusters are most distinguishable by their differential targeting of various thalamic nuclei.
483 For example, SSp3 neurons have more projections to VAL and VM than neurons in other
484 clusters. In addition, cluster SSp1 contains most of the cells projecting to medulla. Cells in
485 cluster SSp4 have the fewest subcortical projections. In SSs, L5 ET neurons can be divided into
486 3 clusters (**Fig. 7f,i**). Cells in clusters SSs1 and SSs2 have fewer projections to the midbrain
487 compared to those in SSs3. On the other hand, SSs1 and SSs2 cells project to CEA and MEA but
488 SSs3 cells do not.

489 Similar to the IT neurons, there is no clear one-to-one correspondence between local morphology
490 and long-range projections of L5 ET neurons, except that neurons in the medulla-projecting
491 SSp1 cluster are mostly present in the S1 and S3 clusters, which have deeper soma depth and
492 larger cortical areas (**Fig. 7j**). This is consistent with previous finding that these neurons are
493 located in the deeper part of layer 5 [32]. We compared the local morphologies of MOp/MOs
494 neurons with or without medulla projection and found that medulla-projecting neurons have a
495 weak tendency for more extensive and complex dendrites (**Extended Data Fig. 12**).

496

497 **Combination of single neuron projections recapitulates population-level mesoscale**
498 **projection pattern**

499 Mapping axonal projections at the mesoscale has revealed region-specific patterns of
500 connectivity across the mouse brain [14,16,57], yet how projection motifs at the single cell level
501 compose the population-level patterns are still unclear for most of the brain regions. To directly
502 compare the mesoscale and single cell projection patterns, we identified 1,271 single cell
503 morphologies and 141 mesoscale experiments from the Allen Mouse Brain Connectivity Atlas,
504 matched based on soma or injection site being within the same CCFv3 structure
505 (**Supplementary Table 3**). This dataset covers 14 cortical areas and layers combined, 14
506 thalamic nuclei and one striatal structure (CP). Using this location-matched dataset, we
507 determined the targets of single cells and mesoscale groups by thresholding at both individual
508 and group levels and constructed a comparative map (**Fig. 8a**, see Methods). Overall, the
509 combined single cell projection pattern from a region (and cortical layer) is highly concordant
510 with that of the mesoscale experiments.

511 We also observe a few exceptions to this general trend. The combined patterns from single L5
512 ET neurons across several cortical areas collectively project to more subcortical targets than
513 mesoscale experiments (indicated with more red colored dots), likely due to broader distribution
514 of the single neurons in the source areas. On the other hand, for several thalamic nuclei (i.e.,
515 SMT, VAL, AM, LD, RE and CM), single neurons collectively have not captured the full
516 projection patterns from mesoscale experiments (indicated with more blue dots). This difference
517 could be due to several reasons: (1) since some of these nuclei are small, the mesoscale
518 experiments may include projections labeled from neighboring nuclei so the single cell data may
519 more accurately represent the true output pattern; (2) the number of reconstructed single neurons
520 is still relatively small and may not fully represent all projection types in a given nucleus; (3) the
521 reconstructed neurons may represent only a subset of the cell types located in these nuclei given
522 the Cre driver labeling method employed, and there may be other types of projection neurons not
523 labeled in these specific Cre lines.

To quantitatively compare the single cell and mesoscale tracer experiments, we calculated the correlation coefficient of each single cell's brain-wide projection weights with the average projection weights from the location-matched mesoscale experiments (**Fig. 8b**). The correlation coefficient ranges from -0.11 (e.g. AM) to 0.99 (e.g. LGd), with a median of 0.82. High correlation coefficients may indicate simple compositions of projecting patterns, e.g. LGd with almost pure VISp projections and CP projecting to either GPe or SNr. Low correlation coefficients may indicate complex composition of projecting patterns, e.g. CM, AM and RE for reasons mentioned above. To compare single cell projection strength relative to mesoscale data, we developed a 'Neuron-beta' metric, as the covariance of a single cell and the average of mesoscale samples, relative to the mesoscale variance (**Fig 8b**, see Methods). Single cells with Neuron-beta values >1.5 correlate well with mesoscale data but fluctuate more variably. For example, individual VM neurons are highly diverse but positively correlated with mesoscale data. Small (<0.5) positive Neuron-beta values result from low correlation (<0.13) of single and mesoscale data. For cell types with Neuron-beta values around 1, single cell and mesoscale data appear to be comparable.

To study how well the mesoscale projection pattern could be broken down to our set of single cells, we performed linear regression with Lasso regularization. This approach selects a minimal set of single cells and uses weighted summation of single cell axon length to approximate the cell type specific mesoscale axonal weights. The overall coefficient of determination (R^2) is 0.86, indicating that mesoscale connectivity is recapitulated well (**Fig. 8c**). R^2 is >0.8 for most groups except for the above-mentioned thalamic nuclei and L4 IT neurons from SSp (**Fig. 8a** lower panel). The discrepancy in the SSp L4 IT group is likely due to the fact that the Cre lines of L4 mesoscale experiments (Nr5a1-Cre, Scnn1a-Tg3-Cre and Rorb-IRES2-Cre) are not entirely L4 specific and also label some L5 cells [4], thus the single cell projection pattern is more accurate in this case. Only 165 out of 1271 single cells with non-zero values contribute to the regression. These cells represent a minimal set of stereotypes to make up the population level connectivity (**Fig. 8d**). Averaging across all single cells shows a low level of approximation (R^2=-0.03), suggesting highly diverse morphologies and projection patterns among the single cells.

Extensive projection diversity in transcriptomically homogeneous Car3 cortical and claustral neurons

Finally, we investigated a special type of cortical excitatory neurons, the L6 Car3 IT transcriptomic subclass [4,64], whose morphology and projection patterns have been unknown. This subclass of neurons is selectively labeled by a unique set of marker genes including *Gnb4*, which is located in the deep layer (mostly L6) of all lateral cortical areas and shares the same transcriptomic clusters with neurons from the claustrum (CLA, **Extended Data Fig. 13a,b**). Previous mesoscale anterograde tracing showed that claustrum neurons project widely into cortex, with particularly strong connections with prefrontal and retrohippocampal cortical areas [68]. We performed similar experiments with the same AAV tracer injected in CLA, SSs or SSp of the Gnb4-IRES2-CreERT2 mice, and found that L6 Car3 neurons in SSs and SSp also showed intracortical projections, but with a more restricted, distinct set of targets compared to CLA projections (**Extended Data Fig. 14**).

We reconstructed 96 neurons from the cortex and CLA of the Gnb4-IRES2-CreERT2 line (**Supplementary Table 1**), which contain 34 CLA neurons, 59 neurons from multiple lateral

568 cortical areas, and 3 neurons from the dorsal part of endopiriform nucleus (EPd) (**Fig. 9a,**
569 **Supplementary Table 2**). We performed clustering using four feature sets – projection pattern,
570 soma location, axon morphology and dendrite morphology, and identified 14 clusters (**Fig. 9b-**
571 **d**). We calculated the total number of projection targets using two different thresholds to label a
572 region as "targeted" (**Fig. 9e**). Using a minimum of 1 mm of axon length (identical threshold as
573 used in [22]) we found the median number of targets to be 29 for CLA neurons and 18 for cortical
574 L6 Car3 neurons. Alternatively, using the existence of at least one axon terminal as the threshold
575 we found the median number of targets to be 21 for CLA neurons and 12 for L6 Car3 neurons. In
576 both cases, the numbers of targets were substantially greater than the above L2/3/4 IT neurons as
577 well as that reported previously for L2/3 IT neurons of the primary visual cortex [22].

578 All CLA and cortical L6 Car3 neurons project almost exclusively into the cortex with none or
579 minimal axon projections into the striatum. This is another major difference between the L6 Car3
580 cortical and claustral neurons and other types of corticocortical-projecting IT neurons which
581 have substantial axon collaterals projecting to the striatum [61]. EPd neurons have distinct
582 projections to frontal cortex, piriform cortex and subcortical olfactory areas [69,70].

583 Cortical L6 Car3 cells are assigned to 8 clusters, with clusters 7, 10, 11, 12 and 13 projecting
584 ipsilaterally, and clusters 5, 6 and 14 bilaterally (**Fig. 9f**). The clusters are arranged
585 topographically from anterior to posterior cortex based on both soma location and projection
586 target specificity; each cluster contains a group of neurons that are located close to each other
587 and project to similar cortical target areas. Clusters 1, 2, 3 and 8 belong to the CLA-ipsilateral
588 group, in which cluster 1 is lateral-projecting and clusters 2, 3 and 8 are midline-projecting (**Fig.**
589 **9f**). Cluster 4 contains CLA-bilateral midline-projecting cells. Cluster 9 contains a mixture of
590 CLA, EPd and cortical cells, as they show similar soma locations and axon projection features
591 with preferential projection to the frontal pole (FRP) and ORBl areas on the ipsilateral side.

592 Interestingly, we identify a unique group of four L6 Car3 cells, all with their somas located in
593 temporal association areas (TEa) and ectorhinal area (ECT), three of which belonging to cluster
594 13 whereas the 4th one is an outlier, have substantial projections into amygdala areas including
595 lateral amygdalar nucleus (LA), basolateral amygdalar nucleus (BLA) and basomedial
596 amygdalar nucleus (BMA), in addition to their cortical targets (**Fig. 9b,d**). There are also several
597 CLA neurons, scattered in several clusters, with minor axon collaterals projecting into amygdala
598 areas [21,68] (**Fig. 9d**).

599 In our single-cell transcriptomic taxonomy of the entire mouse cortex and hippocampus [64], the
600 L6 Car3 subclass contains 1,997 cells, of which 799 are from CLA and 1,198 are from various
601 cortical regions (**Extended Data Fig. 13a,b, Supplementary Table 4**). Cells from different
602 regions are distributed similarly across 5 clusters within the subclass, suggesting a lack of
603 region-specificity for any cluster and that these cortical and claustral cells are highly related to
604 each other, possibly reflecting common or closely related developmental origins. In an attempt to
605 link molecular identities with the projection diversity described above, we performed Retro-seq
606 on cells isolated from CLA (238) and cortical areas SSs (11) and TEa (35) that were labeled by
607 retrograde tracers injected into far apart cortical areas, ACA (medial), MOp (central), ORBl or
608 VISpl (lateral) (**Extended Data Fig. 13c,d, Supplementary Table 4**). The CLA and cortical
609 Retro-seq cells projecting to different cortical areas share the same set of L6 Car3 clusters,
610 indicating no clear one-to-one correspondence between transcriptomic clusters and projection
611 target specificity.

612 Taken together, the remarkable morphological diversity of these fully reconstructed neurons
613 from claustrum, endopiriform nucleus and cortex, which all belong to a single, transcriptomically
614 defined L6 Car3 subclass, demonstrates that long-range axonal projections vary greatly
615 according to their cell body locations, indicative of a combined functional and topographic
616 organization of structural connectivity.

617

618 **DISCUSSION**

619

620 To fully understand the morphological and projection diversity and specificity of neurons across
621 the brain, a large number, likely in the range of hundreds of thousands of neurons will need to be
622 examined. Approaches such as MAPseq [22,71,72] can quickly survey projection specificity at the
623 regional level for many neurons in a high throughput manner. However, many essential details
624 can only be obtained through full morphological reconstructions. Collecting such ground truth
625 data provides an invaluable opportunity to uncover principles of neuronal diversity and circuit
626 organization, informing functional studies. To this end, we here report a large set of full neuronal
627 morphologies, generated from a standardized platform we established, which reveal region- and
628 cell type-specific diversities.

629 Our unique labeling strategy using stable and universal transgenic reporter mouse lines coupled
630 with a variety of sparse Cre delivery methods has several advantages. First, the TIGRE2.0-based
631 transgenic reporter lines, especially Ai166 which expresses a farnesylated GFP, produce very
632 bright GFP labeling of axon fibers under fMOST imaging, revealing numerous terminal boutons,
633 an essential requirement for obtaining truly complete morphologies. Second, this strategy
634 enables sparse labeling across multiple regions within the same brain, improving efficiency
635 compared to other methods (*e.g.*, *in vivo* electroporation or stereotaxic virus injection). Third, the
636 labeling is highly consistent from cell to cell, cell type to cell type, region to region, and brain to
637 brain, reducing variability and enhancing reproducibility. Finally, sparse Cre recombination can
638 be achieved through the use of transgenic Cre or CreERT2 driver lines labeling any neuronal
639 type, or low-dose Cre viral vectors delivered through either local or systemic (*e.g.*, retroorbital)
640 injections.

641 Development of novel and accessible software tools are essential for reconstruction efficiency.
642 Such powerful tools have been recently developed by the Janelia MouseLight project [36].
643 Similarly, our enhanced Vaa3D-based reconstruction toolkit streamlines large-volume fMOST
644 image data processing and computation-assisted manual reconstruction. The registration of the
645 fMOST whole-brain datasets to the CCF allows quantification of projection strength in each
646 target region across the entire brain for each neuron, and subsequent data-driven clustering to
647 identify similarities and differences across neurons and to group them into types. The
648 accumulation of an increasingly larger set of fully reconstructed neurons in the future can be
649 used as training datasets to develop machine learning-based automatic reconstruction algorithms
650 to further boost the throughput of reconstruction.

651 Our extensive and detailed analysis of this large collection of reconstructed neurons has yielded a
652 number of novel findings regarding neuronal projection diversity at multiple organizational
653 levels. First, at the regional level, neurons from different brain regions follow different
654 convergence or divergence rules in their long-range projections. Striatal neurons, both GPe-

projecting and SNr-projecting types, have highly convergent projections into their respective targets. The core-type thalamic neurons have largely point-to-point projections to their cortical targets. The matrix-type thalamic neurons, the claustral neurons and all classes of cortical neurons have divergent projections to a few (for thalamic neurons) or many (for claustral and cortical neurons) target regions. The degree of similarity between individual neurons within a given region also varies across regions. The striatal neurons and core-type thalamic neurons have one dominant axon branches that are highly consistent among individual neurons within the same type. The matrix-type thalamic neurons usually have a few axon branches but are also mostly consistent among the individual neurons within each nucleus. The cortical and claustral neurons, on the other hand, have highly variable axon branching and projection patterns with each neuron selecting only a subset of the targets.

Second, at the major cell class or type level, robust distinctions in morphological features between types have been discovered. For example, striatal GPe-projecting neurons have more extensive local axon arborization than SNr-projecting neurons. The core-type thalamic neurons can be divided into small-arbor and large-arbor types, whereas the matrix-type thalamic neurons all have much larger, diffusely distributed axon arbors. In the cortex, the L2/3/4 IT, L5 ET and L6 Car3 subclasses exhibit highly distinct projection patterns. L5 ET cells project predominantly to subcortical targets while having limited intracortical collateral projections. L2/3/4 IT and L6 Car3 cortical/claustral neurons both have predominantly intracortical projections, but are easily distinguishable by the much more extensive and diffused intracortical projection and the lack of striatal projection from the latter.

Third, further morphological diversities exist within a major cell class or type. Both core- and matrix-type thalamic neurons exhibit strong nucleus-specific projection patterns, sometimes even allowing further subdivision of a nucleus (e.g. LP and MD). In the cortex, we were able to cluster both L5 ET and L6 Car3 neurons into multiple subtypes with robustly distinct projection patterns that are often correlated with the regions their somas reside in. Furthermore, we identified cell populations with additional unique projection targets. For example, L5 ET cells in the motor cortex have a highly distinct subtype that situates in deep L5 and projects to the medulla; a unique set of L6 Car3 cells in TEa/ECT and claustrum send projections to amygdala areas. L2/3/4 IT neurons exhibit largely continuous variation in their projection patterns, but still show regional differences. L4 IT cells in SSp mostly do not have long projections, a dramatic difference from L4 IT cells we found in MOp, MOs and SSs which all have long intracortical projections. Among the L2/3/4 IT neurons with long projections, those from the motor cortex have longer total axon arborizations than those from the somatosensory cortex.

Fourth, hierarchical organization of thalamocortical and corticocortical connections has been shown with mesoscale dataset based on feedforward and feedback connections [57]. In this study, we further demonstrate hierarchical organization exists at single neuron level. The core-type thalamic neurons predominantly terminate in L4 and lower L2/3 of their cortical target areas, characteristic of feedforward projections. Most of our reconstructed matrix-type thalamic neurons have substantial axon branches in L1 in their cortical target areas, characteristic of feedback projections. In the cortex, we find that axon arbors of SSp neurons predominantly terminate in middle layers in all target areas, consistent with the low hierarchical position of SSp with all of its output pathways being feedforward. On the other hand, MOp neurons appear to show both feedforward, e.g. mid-layer projection to MOs, and feedback, e.g. L1 projection to SSp and SSs, patterns.

These wide range of projection patterns and rules revealed in our study suggest highly distinct functional roles between different major neuron classes and types, as well as a diverse range of potential ways for individual neurons to participate in circuit activities.

To bring all these findings into the general framework of cell type characterization and classification, a major remaining question is how the morphological and projectional diversities compare and correlate with the neurons' molecular identities. We attempted to address this question with two approaches, using validated driver lines to define the major cell class or type level identities of reconstructed neurons and Retro-seq to obtain transcriptomic profiles of neurons projecting to specific targets. Both approaches show that major cell classes or types of neurons have highly distinct morphological and projection patterns, however, at more fine-grained levels within these major types, especially for cortical and claustral neurons, we also observe extensive morphological and projection diversities that cannot be accounted for by preexisting transcriptomic subtypes or clusters. For example, Retro-seq cells with highly distinct targets (e.g. SSp L2/3/4 IT cells projecting to MOp or SSs, L6 Car3 and claustral neurons belonging to distinct morphological clusters) were mapped to the same transcriptomic clusters. Previous studies showed that L2/3 SSp pyramidal neurons projecting to MOp or SSs have distinct intrinsic and network physiological properties [20,73]. Even though they may not belong to distinct transcriptomic subtypes, it will be interesting to find out if there are gene expression differences that correspond to the differential connectional and physiological properties for these neurons, as did for primary visual cortical neurons projecting differentially to medial or lateral higher visual areas [74].

The large diversity of axonal morphologies and projection patterns observed from genetically well-defined neuronal populations is striking but also consistent with an increasing body of related work [22,36]. The large number of target regions each neuron innervates and their elaborate pattern differences (e.g. layer distributions) would make it extremely challenging if not impossible to accurately discern the difference of projection patterns between individual neurons using other approaches including multiplexed retrograde tracing or tissue dissection of barcode-labeled brains. It underscores the necessity of scaling up the full neuronal morphology characterization effort to gain a true understanding of the extent of such diversity. Such knowledge is foundational for the understanding of brain connectivity and function.

The apparent lack of correlation between transcriptomic and morphological types in many cases is intriguing. It is possible that the current unsupervised clustering approach is insufficient to uncover the genes specifically relevant to morphology from the thousands of genes expressed by the neurons. Alternatively, it is possible that morphological/connectional specificity is established during circuit development and that the associated gene signatures exist only at that time [75].

In either case, the result emphasizes the importance of performing single cell characterization within multiple modalities and taking an integrated approach to describe and classify cell types in an unbiased and comprehensive manner. In the future, it will be important to develop methods that allow full morphology reconstruction and gene expression profiling to be conducted in the same cell, and apply them to the study of single cells in both adult stage and during brain development, so that potential molecular correlates of morphological/connectional features can be identified. This and other approaches together will ultimately lead to an integrated understanding of the extraordinary cellular diversity of the brain that underlies its function.

ACKNOWLEDGMENTS

We are grateful to the In Vivo Sciences, Molecular Biology, Histology, and Imaging teams at the Allen Institute for their technical support. This work was supported by the Allen Institute for Brain Science and by multiple grant awards from institutes under the National Institutes of Health (NIH), including award number R01EY023173 from The National Eye Institute to H.Z., U01MH105982 from the National Institute of Mental Health and Eunice Kennedy Shriver National Institute Of Child Health & Human Development to H.Z., and U19MH114830 from the National Institute Of Mental Health to H.Z. and Q.L. Generation of TIGRE-MORF mouse line was partly funded by U01MH106008 and U01MH117079 from the National Institute Of Mental Health to X.W.Y. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH and its subsidiary institutes. This work is also funded by an initiative of Southeast University (SEU) to support Open Science oriented international collaboration. SEU supported the informatics data management and analysis pipeline at the SEU-Allen Joint Center. The work was also funded by the National Science Foundation of China (NSFC) grant 61721092 to Q.L., NSFC grant 61871411 to L.Q., the University Synergy Innovation Program of Anhui Province GXXT-2019-008 to L.Q., NSFC grant 91632201 to W.X., and by the Tiny Blue Dot Foundation to C.K. The Wenzhou Medical University reconstruction team received funding from the EPFL - Blue Brain Project. The authors wish to thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

AUTHOR CONTRIBUTIONS

H.Z. conceptualized the study. H.P. envisioned and led the development of the computational and data analysis platform. M.B.V., T.L.D., B.T. and X.W.Y. generated the TIGRE-MORF (Ai166) mouse line. Z.J.H. provided Fezf2-CreER, Plxnd1-CreER and Tle4-CreER mouse lines. K.E.H., R.L. T.L.D., B.T. and J.A.H. contributed to the generation and characterization of specific transgenic mouse lines. H.G., A.L., S.Z., X.L., J.Y. and Q.L. conducted fMOST imaging. W.W., S.J., Y.Y. and C.H. handled the imaging data. L.Q., L.N. and H.P. developed methods for registration of fMOST datasets to CCF. Zhi Z., S.J., Y.Y., Yimin W. and H.P. developed software tools for data conversion and morphology reconstruction. Yun W., X.K., Y.L., L.L., P.L., Y.S., L.Y., S.Z., A.F., E.S., J.P., J. Y., G.H., A.L., and Z.D. performed manual and semi-automatic morphology reconstruction. L.L. co-developed the neuronal reconstruction pipeline at SEU-ALLEN and contributed to analysis of striatal and thalamic cell types. Yun W., P.X., J.A.H. and H.P. performed manual or computational classification of morphological types. Zhi Z., S.K. Zi Z., and S.A.S. assisted with morphological analysis. L.D. contributed to single neuron and population level projection analysis. Y.Z., D.L., and Yun W. collaborated with H.P. on developing a quality control method for neuron reconstruction. H.P., W.X., Z.G. and H.Z. collaborated in setting up the SEU-ALLEN data-production team. K.E.H., Q.W. and J.A.H. conducted anterograde AAV tracing. T.N.N. performed retrograde tracing. Z.Y., T.N.N. and B.T. conducted scRNA-seq data generation and analysis. S.M. and S.M.S. provided project management. L.E., M.J.H., B.T., L.N., S.A.S., J.A.H., H.G., Q.L., H.P., H.Z. and C.K. provided scientific management. H.Z. led the writing of the manuscript in consultation with all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

METHODS

Animal care and use

Both male and female transgenic mice ⩾ P56 were utilized for all experiments. All animals were housed 3-5 per cage and maintained on a 12-hour light/dark cycle, in a humidity- and temperature-controlled room with water and food available *ad libitum*. All experimental procedures related to the use of mice were conducted with approved protocols in accordance with NIH guidelines, and were approved by the Institutional Animal Care and Use Committee (IACUC) of the Allen Institute for Brain Science.

Transgenic mice

All transgenic crosses are listed in **Supplementary Table 1**. Data for systematic characterization of the expression pattern of each transgenic mouse line can be found in the AIBS Transgenic Characterization database (http://connectivity.brain-map.org/transgenic/search/basic).

Induction of CreERT2 driver lines was done by administration via oral gavage (PO) of tamoxifen (50 mg/ml in corn oil) at original (0.2 mg/g body weight) or reduced dose for one day in an adult mouse. The dosage for mice age P7-P15 is 0.04 ml. Mice can be used for experiments at 2 or more weeks after tamoxifen dosing. Specific dose of tamoxifen to induce sparse labeling in each CreERT2 driver line is shown in **Supplementary Table 1**.

TissueCyte STPT imaging

Imaging by serial two-photon (STP) tomography (TissueCyte 1000, TissueVision Inc. Somerville, MA) has been described in earlier published studies [16,76].

Mice were deeply anesthetized with 5% isoflurane and intracardially perfused with 10 ml of saline (0.9% NaCl) followed by 50 ml of freshly prepared 4% paraformaldehyde (PFA) at a flow rate of 9 ml/min. Brains were dissected and post-fixed in 4% PFA at room temperature for 3–6 h and then overnight at 4 °C. Brains were rinsed briefly with PBS and stored in PBS with 0.1% sodium azide until imaging.

Prior to imaging, the brain was embedded in a 4.5% oxidized (10 mM NaIO4) agarose solution in a grid-lined embedding mold to standardize its placement in an aligned coordinate space. The agarose block was then left at room temperature for 20 min to allow solidification. Cross-linking between brain tissue and agarose was promoted by placing the solidified block in 0.5% sodium borohydride in 0.5 M sodium borate buffer (pH 9.0) overnight at 4 °C. The agarose block was then mounted on a 1 × 3 glass slide using Loctite 404 glue and prepared immediately for serial imaging.

825 Image acquisition was accomplished using TissueCyte 1000 systems (TissueVision, Cambridge,
826 MA) coupled with Mai Tai HP DeepSee lasers (Spectra Physics, Santa Clara, CA). The mounted
827 specimen was fixed through a magnet to the metal plate in the centre of the cutting bath filled
828 with degassed, room-temperature PBS with 0.1% sodium azide. A new blade was used for each
829 brain on the vibratome and aligned to be parallel to the leading edge of the specimen block.
830 Brains were imaged from the caudal end. The specimen was illuminated with 925 nm
831 wavelength light through a Zeiss 320 water immersion objective (NA = 1.0), with 250 mW light
832 power at objective. The two-photon images for red, green and blue channels were taken at 75 µm
833 below the cutting surface. To scan a full tissue section, individual tile images were acquired, and
834 the entire stage was moved between each tile. After an entire section was imaged, the x and y
835 stages moved the specimen to the vibratome, which cut a 100-µm section, and returned the
836 specimen to the objective for imaging of the next plane. The blade vibrated at 60 Hz and the
837 stage moved towards the blade at 0.5 mm per sec during cutting. Images from 140 sections were
838 collected to cover the full range of mouse brain at an x-y resolution of 0.35 µm per pixel. Upon
839 completion of imaging, sections were retrieved from the cutting bath and stored in PBS with
840 0.1% sodium azide at 4°C.

841

842 **fMOST imaging**

843 In summary, a GFP-labeled brain is first embedded in resin. The resin-embedded GFP
844 fluorescence can be recovered through chemical reactivation [77] provided by adding Na_2CO_3 in
845 the imaging water bath. Thus, a line-scanning block-face imaging system can be employed to
846 maximize imaging speed. Following imaging of the entire block-face, the top 1-µm tissue is
847 sliced off by a diamond knife, exposing the next face of the block for imaging. For the entire
848 mouse brain, a 15-20 TB dataset containing ~10,000 coronal planes of 0.2-0.3 µm X-Y
849 resolution and 1 µm Z sampling rate is generated within 2 weeks.

850 All tissue preparation has been described previously [78]. Following fixation, each intact brain was
851 rinsed three times (6 h for two washes and 12 h for the third wash) at 4°C in a 0.01 M PBS
852 solution (Sigma-Aldrich Inc., St. Louis, US). Then the brain was subsequently dehydrated via
853 immersion in a graded series of ethanol mixtures (50%, 70%, and 95% (vol/vol) ethanol
854 solutions in distilled water) and the absolute ethanol solution three times for 2 h each at 4°C.
855 After dehydration, the whole brain was impregnated with Lowicryl HM20 Resin Kits (Electron
856 Microscopy Sciences, cat.no. 14340) by sequential immersions in 50, 75, 100 and 100%
857 embedding medium in ethanol, 2 h each for the first three solutions and 72 h for the final
858 solution. Finally, each whole brain was embedded in a gelatin capsule that had been filled with
859 HM20 and polymerized at 50°C for 24 h.

860 The whole brain imaging is realized using a fluorescence micro-optical sectioning tomography
861 (fMOST) system. The basic structure of the imaging system is the combination of a line-
862 scanning upright epi-fluorescence microscopy with a mechanic sectioning system. This system
863 runs in a line-scanning block-face mode but updated with a new principle to get better image
864 contrast and speed and thus enables high throughput imaging of the fluorescence protein labeled
865 sample (manuscript in preparation). Each time we do a block-face fluorescence imaging across
866 the whole coronal plane (X-Y axes), then remove the top layer (Z axis) by a diamond knife, and
867 then expose next layer, and image again. The thickness of each layer is 1.0 µm. In each layer
868 imaging, we used a strip scanning (X axis) model combined with a montage in Y axis to cover

869 the whole coronal plane [79]. The fluorescence, collected using a microscope objective, passes a
870 bandpass filter and is recorded with a TDI-CCD camera. We repeat these procedures across the
871 whole sample volume to get the required dataset.

872 The objective used is 40X WI with numerical aperture (NA) 0.8 to provide a designed optical
873 resolution (at 520 nm) of 0.35 μm in XY axes. The imaging gives a sample voxel of 0.35 x 0.35
874 x 1.0 μm to provide proper resolution to trace the neural process. The voxel size may vary for
875 difference objectives. Other imaging parameters for GFP imaging include an excitation
876 wavelength of 488 nm, and emission filter with passing band 510-550 nm. The fMOST is a two-
877 color imaging system. The green channel is used to obtain the complete morphology of neurons,
878 and the red channel is used to obtain the cellular architecture information of propidium
879 iodide(PI) staining.

880

881 **Full neuronal morphology reconstruction system**

882 We usedVaa3D, an open-source, cross-platform visualization and analysis system, for the tasks
883 of reconstructing massive neuronal morphologies. To efficiently and effectively deal with the
884 whole-mouse brain imaging data, we incorporated several enabling modules into Vaa3D, such as
885 TeraFly, TeraVR, and a number of other supporting tools. TeraFly supports visualization and
886 annotation of multidimensional imaging data with virtually unlimited scales. A user can flexibly
887 choose to work at a specific region of interest (ROI) with desired level of detail (LoD). TeraVR
888 is an annotation tool for immersive neuron reconstruction that has been proved to be critical for
889 achieving precision and efficiency in morphology data production. It creates stereo visualization
890 for image volumes and reconstructions and offers an intuitive interface for the user to interact
891 with such data. Both TeraFly and TeraVR are seamlessly integrated in Vaa3D and can be used
892 combinedly and flexibly. From reconstructions (in SWC file format), morphological
893 quantification statistics is obtained to characterize neurons. QC process identifies errors based on
894 morphological indicators and does corrections in a feedback setting. QC process then refines the
895 skeleton location with Mean-Shift [80] and performs pruning focused on terminal location
896 refinement. When needed auto-refinement fits the tracing to the center of fluorescent signals.
897 The whole process ends with SWC resampling and registration. The final reconstruction of each
898 neuron is a valid single tree without breaks, loops, multiple branches from a single point, etc.

899

900 **Registration to CCF**

901 We used mBrainAligner based on BrainAligner [81] to perform 3D registration from fMOST
902 images (subject) to the average mouse brain template of CCFv3 (target) (**Extended Data Fig. 3**).
903 The main steps are: 1) fMOST images were first down-sampled by 64x64x16 (X, Y, Z) to
904 roughly match the size of target brain. 2) The stripe artifacts in fMOST images that raised from
905 diamond knife cutting and imaging process were eliminated by using log-space frequency notch
906 filter. 3) The dense outer-contour feature points of target and subject brain (about 1500 points per
907 brain) were uniformly sampled from the brains' outer-contour that obtained using adaptive
908 threshold, and then affine aligned using a reliable landmark points matching algorithm to ensure
909 the subject brain has the same position, orientation and scale as the target brain. 4) Intensity was
910 normalized by matching the local average intensity of subject image to that of target image in a
911 sliding window manner with patch size 41x41x41 and stride one. 5) For the target brain, 1,744

landmarks corresponding to the points of high curvature (corners or junction of different brain compartments) in CCFv3 annotation image were detected via 3D Harris corner detector. Based on a combination of texture, shape context and deep-learning-derived features, mBrainAligner established the correspondence between target and subject brain by iteratively deforming these target landmarks to fit the subject image, and accomplished the local alignment using the smooth-thin-plate-spline (STPS). 6) Finally, to ensure the accuracy of registration, automatic registration results were examined in the semi-automatic registration module of mBrainAligner, and if necessary, the boundaries of brain region were further optimized in a manual or semi-automatic way. Once images were aligned, the reconstructed neurons and somas were warped to CCF space using the generated deformation fields.

Processing single cell morphological data

Pre-processing of SWC files: SWC files were processed and examined with Vaa3D plugins to ensure topological correctness: sorted single tree with root node as soma. Terminal branches < 10 pixels were pruned to remove artifacts. SWC files were resampled with a step size of 64 (x), 64 (y) and 16(z) before registration.

Quantification of axon projection patterns: To analyze the distribution and amount of axon in brain-wide targets following registration to the CCFv3, we used a manually curated set of 316 non-overlapping structures at a mid-ontology level that are most closely matched in size or division. Ipsi- and contra-lateral sides of brain regions were calculated separately.

Morphological features: Axonal and dendritic morphological features, defined according to L-measurement (Scorcioni et al., 2008), were calculated using Vaa3D plugin "global_neuron_feature". Selected features include:

Axon global: 'Overall Width', 'Overall Height', 'Overall Depth', 'Total Length', 'Euclidean Distance', 'Max Path Distance', 'Number of Branches'.

Axon local: 'Total Length', 'Number of Branches'.

Dendrite: 'Overall Width', 'Overall Height', 'Overall Depth', 'Total Length', 'Max Euclidean Distance', 'Max Path Distance', 'Number of Branches', 'Max Branch Order'.

Local axons were defined as axon arbors within 200 microns from the somata. Local axons and dendrites were rotated based on principle component analysis (PCA) so dimensions were aligned with the largest to smallest spans. Then shifting was performed to localize somata at the origin of coordinates.

mBrainAnalyzer

Our mBrainAnalyzer toolbox, which was developed for analysis of full neuron morphology, includes multiple modules for feature quantification, arbor detection, statistical analysis and visualization. In addition to morphological features (e.g. total length, angle of branches etc.), this toolbox also quantifies projection intensities at branch length level and number of terminal levels. Using the arbor detection module, one can define sub-cellular components of a neuron as the granularity. Analysis and visualization can be performed at both whole-cell and arbor levels.

953 **Arbor detection and partition**

954 We detected and partitioned a series of neuronal arbors out of each neuron reconstruction using a
955 graph-partition clustering method. First, as a neuron consists of a number of topologically
956 connected reconstruction nodes, the neuron was viewed as a graph, where every reconstruction
957 node (unit) in the neuron was connected with its parent node with an edge specified by the
958 topological connection of the parent-child pair with the edge weight, or 'similarity' s, set to be
959 the exponential of the negative 3D Euclidean distance, d, of these two nodes, i.e. $s = \exp(-d)$.
960 Then, we considered the normalized graph-cut method [82] to extract "clusters" of reconstruction
961 nodes so that the within-cluster "total similarity" of nodes would be maximized and cross-cluster
962 total similarity would be minimized. As a result, each such coherent cluster corresponds to one
963 neuron arbor, which was also visually checked to ensure its correctness. Third, to automatically
964 determine the number of such clusters, for a presumed number of clusters, we calculated the
965 normalized score of total cross-cluster similarity divided by the total within-cluster similarity,
966 followed by trial-testing a range (between 2 to 8) of such presumed cluster-numbers to determine
967 the optimal number that would minimizes this normalized score. In the final result, the detected
968 arbor that contains the soma is called soma-arbor; the remaining arbors are called non-soma
969 arbors.

970

971 **Feature quantification of cortical arbors**

972 We divided the cortex into consecutive coronal slices of 100μm thick. Anchor points were
973 evenly sampled along the outer border of each slice, with normal vectors that are perpendicular
974 to the local cortical surface and pointing to the inside of the brain. Nodes of arbors were assigned
975 to their neighbor anchors and projected onto the surface by corresponding normal vectors. Depth
976 of nodes were determined by the length of projection along normal vectors. We also estimated
977 the area of an anchor by their distance to neighbor anchors and slice thickness. The 2D cortical
978 area of an arbor was determined by the total areas of unique anchors occupied by its nodes. To
979 determine the radius of an arbor, we assigned arbor 'center' as the node that has the shortest
980 average distance to other nodes. Radius was determined by a growing sphere until 70% segments
981 are inside it. For neurons with tufted apical dendrite, we vertically shifted the arbors, so the top
982 of apical dendrites reached L1. We manually confirmed that all tufted apical dendrites reached
983 L1 in the original image.

984

985 **Clustering of cortical arbors**

986 For local (soma-neighboring) arbors, the following features were used for clustering: '2d_area',
987 'axon_length', 'dend_length', 'radius', 'axon_depth_mean', 'axon_depth_std', 'dend_depth_mean',
988 'dend_depth_std'. Axon/Dendrite depth features were normalized by the average thickness of
989 cortical areas where the arbor locates. We performed PCA to reduce the effect of noise. Top
990 principle components were selected to recover 95% of variance. We applied UMAP dimension
991 reduction using python package 'UMAP' (McInnes et al., 2018). The 'n_neighbors' parameter
992 was set at 4. K-means clustering was performed using the UMAP embeddings as input.

993 For distal arbors, we profiled the axon density distribution along cortical depth as input features
994 of clustering. We did not use the same features as local arbors as distal arbors exhibited much
995 higher diversity. Clustering approach for distal arbors is the same as local arbors.

Neuron-beta

we developed the Neuron-beta metric by borrowing the concept of the 'Beta' value from the finance field [83]. For each group, defined by brain areas and/or cortical layers, we calculated the average of mesoscale experiments as $M = [m_1, \ldots, m_p]$, p = number of brain areas. For one single cell $S = [s_1, \ldots, s_p]$, we define the neuron-beta value as:

$$Neuronbeta = \frac{Cov(M,S)}{Var(M)}$$

Clustering of cortical L6 Car3 and claustral neurons

Data normalization: Morphological features were normalized by the mean and standard variation in a feature-wise manner. Projection pattern features were normalized by the total length per 100 μm in a sample-wise manner and scaled by logarithm. Soma locations were flipped to the same hemisphere.

Similarity metrics: For each feature set, we first calculated the Euclidean distance matrix. Then a ranked K-nearest neighbor (KNN) matrix was created. We then applied the Shared Nearest Neighbor (SNN) approach to measure the similarity between each pair of samples x_i and x_j. The SNN metric was defined as the maximum average rank among their common neighbors:

$$\mathrm{S}(x_i, x_j) = \max_{v \in NN(x_i) \cap NN(x_j)} \{k - \frac{1}{2}[rank_{NN(x_i)}(v) + \frac{1}{2}(rank_{NN(x_j)}(v)]\}$$

Similarity scores were set as 0 for pairs with non-overlapping KNN sets and a weighted SNN graph is created.

Co-clustering analysis: Co-clustering matrix for each feature set was calculated by iterative random sampling. During each iteration, 95% of samples were randomly selected to create an SNN graph. We then applied the "Fast-greedy" community detection algorithm using python package "python-igraph" for clustering assignment. For each pair of samples, the co-clustering score was defined as the times of co-clustering normalized by the iterations of co-occurring. Resampling was performed 5,000 times to reach saturation. The overall co-clustering matrix is a weighted average of the four feature sets. Agglomerative clustering was performed to the co-clustering matrix to get clusters.

Outlier removal: Outliers were detected by comparing the Euclidean distance between a sample and the other samples with the same cluster identity. We used overall within-cluster distance as the background distribution. Samples with significantly higher (one-sided Mann-Whitney test) within-cluster distance were filtered out as outliers. Agglomerative clustering was performed for the remaining co-clustering matrix. This process iterated until no new outlier could be detected.

Characterization of cell types: For each feature set, we performed two-sided Mann-Whitney tests: claustrum vs. cortical neurons; each cluster vs. other clusters. P-values were adjusted by Bonferroni correction.

Anterograde tracing and retrograde labeling

1034 For anterograde projection mapping, we injected AAV2/1-pCAG-FLEX-EGFP-WPRE-pA (Oh
1035 et al., 2014) into CLA, SSs or SSp of Gnb4-IRES2-Cre or Gnb4-IRES2-CreERT2 mice at P37-
1036 P65 respectively. Stereotaxic injection procedures were performed as previous described [16] and
1037 stereotaxic coordinates used for each experiment can be found in the data portal. For the Gnb4-
1038 IRES2-CreERT2 mice, tamoxifen induction was conducted one week post injection at full dose
1039 (0.2 mg/g body weight) for 5 consecutive days. Mice survived 3 weeks (or 4 weeks for the
1040 tamoxifen-induced mice) post injection, and brains were perfused and collected for TissueCyte
1041 imaging.

1042 For retrograde labeling, we injected several different types of retrograde viral tracers, including
1043 AAV2-retro-EF1a-dTomato or AAV2-retro-EF1a-Cre [84], RVdGdL-Cre or RVdL-FlpO [85], or
1044 CAV2-Cre [86], into specific target regions of defined transgenic mice (**Supplementary Table 4**).
1045 We FACS-sorted and collected RFP+ or RFP+/GFP+ cells from defined source regions for
1046 single-cell RNA-sequencing. Stereotaxic injection procedures were performed as previous
1047 described [16]. Mice were injected at P40 or older, with 16-31 days survival post-injection.

1048

1049 **Single-cell RNA-sequencing**

1050 Cells from transgenic mice or transgenic mice injected with retrograde tracers were collected by
1051 microdissection of different cortical regions. Single-cell suspensions were created and cells were
1052 collected using fluorescence activated cell sorting (FACS). FACS gates were selective for cells
1053 with fluorescent protein expression from transgenic and/or viral reporters.

1054 Cells were then frozen at -80° C, and were later processed for scRNA-seq using the SMART-Seq
1055 v4 method [4]. After sequencing, raw data was quantified using STAR v2.5.3 and were aligned to
1056 both a Ref-Seq transcriptome index for the mm10 genome, and a custom index consisting of
1057 transgene sequences. PCR duplicates were masked and removed using STAR option
1058 'bamRemoveDuplicates'. Only uniquely aligned reads were used for gene quantification. Gene
1059 read counts were quantified using the summarizeOverlaps function from R GenomicAlignments
1060 package using both intronic and exonic reads, and QC was performed as described [4].

1061 Clustering was performed using house developed R package scrattch.hicat (available via GitHub
1062 https://github.com/AllenInstitute/scrattch.hicat). In additional to classical single-cell clustering
1063 processing steps provided by other tools such as Seurat, this package features automatically
1064 iterative clustering by making finer and finer splits while ensuring all pairs of clusters, even at
1065 the finest level, are separable by fairly stringent differential gene expression criteria. The
1066 package also performs consensus clustering by repeating iterative clustering step on 80%
1067 subsampled set of cells 100 times, and derive the final clustering result based on cell-cell co-
1068 clustering probability matrix. This feature enables us to both fine tune clustering boundaries and
1069 to assess clustering uncertainty. One critical criterion that determines the clustering resolution is
1070 the minimal differential gene expression (DGE) requirement between all pairs of clusters. Using
1071 stringent DGE requirement results in fewer clusters with more prominent differences between
1072 clusters, while using more relaxed DGE expression result in more clusters captured by more
1073 subtle differences. For the whole cortical and hippocampal dataset with ~75,000 cells, we used
1074 the standard DGE requirement as in [4]. More specifically, q1.th = 0.5 (minimal fraction of cells in
1075 a given cluster that express the positive markers), q.diff.th=0.7 (normalized differences in
1076 fraction of cells expressing the positive markers between the foreground and background cluster,

maximal value is 1), and de.score.th=150 (overall assessment of the statistical significance of all DGE genes).

Data and Code availability

The raw and TeraFly converted fMOST image datasets of all mouse brains used in this study, as well as the CCFv3 registered single neuron reconstructions, are available at BICCN's Brain Image Library (BIL) at Pittsburgh Supercomputing Center (www.brainimagelibrary.org). The single neuron reconstructions, the CCFv3 registered version of these reconstructions, as well as 3D navigation movie-gallery of these data are available at SEU-ALLEN Joint Center, Institute for Brain and Intelligence (https://braintell.org/projects/fullmorpho/). The neuron reconstructions are also released through NeuroMorpho.Org. The Vaa3D platform along with the TeraFly and TeraVR reconstruction software is available through the GitHub release page of vaa3d.org. The mBrainAligner and mBrainAnalyzer software packages are available upon request.

Mesoscale tracing data (including high resolution images, segmentation, registration to CCFv3, and automated quantification of injection size, location, and distribution across brain structures) are available through the Allen Mouse Brain Connectivity Atlas portal (http://connectivity.brain-map.org/). Retro-seq SMART-Seq v4 data have been deposited to BICCN's NeMO Archive.

Reference

1 Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature reviews* **18**, 530-546, doi:10.1038/nrn.2017.85 (2017).

2 Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030 e1016, doi:10.1016/j.cell.2018.07.028 (2018).

3 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature neuroscience* **19**, 335-346, doi:10.1038/nn.4216 (2016).

4 Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78, doi:10.1038/s41586-018-0654-5 (2018).

5 Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022, doi:10.1016/j.cell.2018.06.021 (2018).

6 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (New York, N.Y* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).

7 Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience* **22**, 1182-1195, doi:10.1038/s41593-019-0417-0 (2019).

8 Jiang, X. *et al.* Principles of connectivity among morphologically defined cell types in adult neocortex. *Science (New York, N.Y* **350**, aac9462, doi:10.1126/science.aac9462 (2015).

9 Markram, H. *et al.* Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* **163**, 456-492, doi:10.1016/j.cell.2015.09.029 (2015).

10 Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nature biotechnology* **34**, 199-203, doi:10.1038/nbt.3445 (2016).

11 Fuzik, J. *et al.* Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nature biotechnology* **34**, 175-183, doi:10.1038/nbt.3443 (2016).

12 Ramón y Cajal, S. *Histologie Du Système Nerveux de L'homme & Des Vertébrés.*, (Paris: Maloine [Translated by N. Swanson and L.W. Swanson, Oxford University Press, 1995], 1909).

13 Bota, M., Sporns, O. & Swanson, L. W. Architecture of the cerebral cortical association connectome underlying cognition. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E2093-2101, doi:10.1073/pnas.1504394112 (2015).

14 Gamanut, R. *et al.* The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. *Neuron* **97**, 698-715 e610, doi:10.1016/j.neuron.2017.12.037 (2018).

15 Markov, N. T. *et al.* A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb Cortex* **24**, 17-36, doi:10.1093/cercor/bhs270 (2014).

16 Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207-214, doi:10.1038/nature13186 (2014).

17 Zeng, H. Mesoscale connectomics. *Current opinion in neurobiology* **50**, 154-162, doi:10.1016/j.conb.2018.03.003 (2018).

18 Zingg, B. *et al.* Neural networks of the mouse neocortex. *Cell* **156**, 1096-1111, doi:10.1016/j.cell.2014.02.023 (2014).

19 Minciacchi, D., Molinari, M., Bentivoglio, M. & Macchi, G. The organization of the ipsi- and contralateral claustrocortical system in rat with notes on the bilateral claustrocortical projections in cat. *Neuroscience* **16**, 557-576, doi:10.1016/0306-4522(85)90192-7 (1985).

20 Yamashita, T. *et al.* Membrane potential dynamics of neocortical projection neurons driving target-specific signals. *Neuron* **80**, 1477-1490, doi:10.1016/j.neuron.2013.10.059 (2013).

21 Zingg, B., Dong, H. W., Tao, H. W. & Zhang, L. I. Input-output organization of the mouse claustrum. *The Journal of comparative neurology* **526**, 2428-2443, doi:10.1002/cne.24502 (2018).

22 Han, Y. *et al.* The logic of single-cell projections from visual cortex. *Nature* **556**, 51-56, doi:10.1038/nature26159 (2018).

23 Li, L. *et al.* A robot for high yield electrophysiology and morphology of single neurons in vivo. *Nature communications* **8**, 15604, doi:10.1038/ncomms15604 (2017).

24 Oberlaender, M. *et al.* Three-dimensional axon morphologies of individual layer 5 neurons indicate cell type-specific intracortical pathways for whisker motion and touch. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4188-4193, doi:10.1073/pnas.1100647108 (2011).

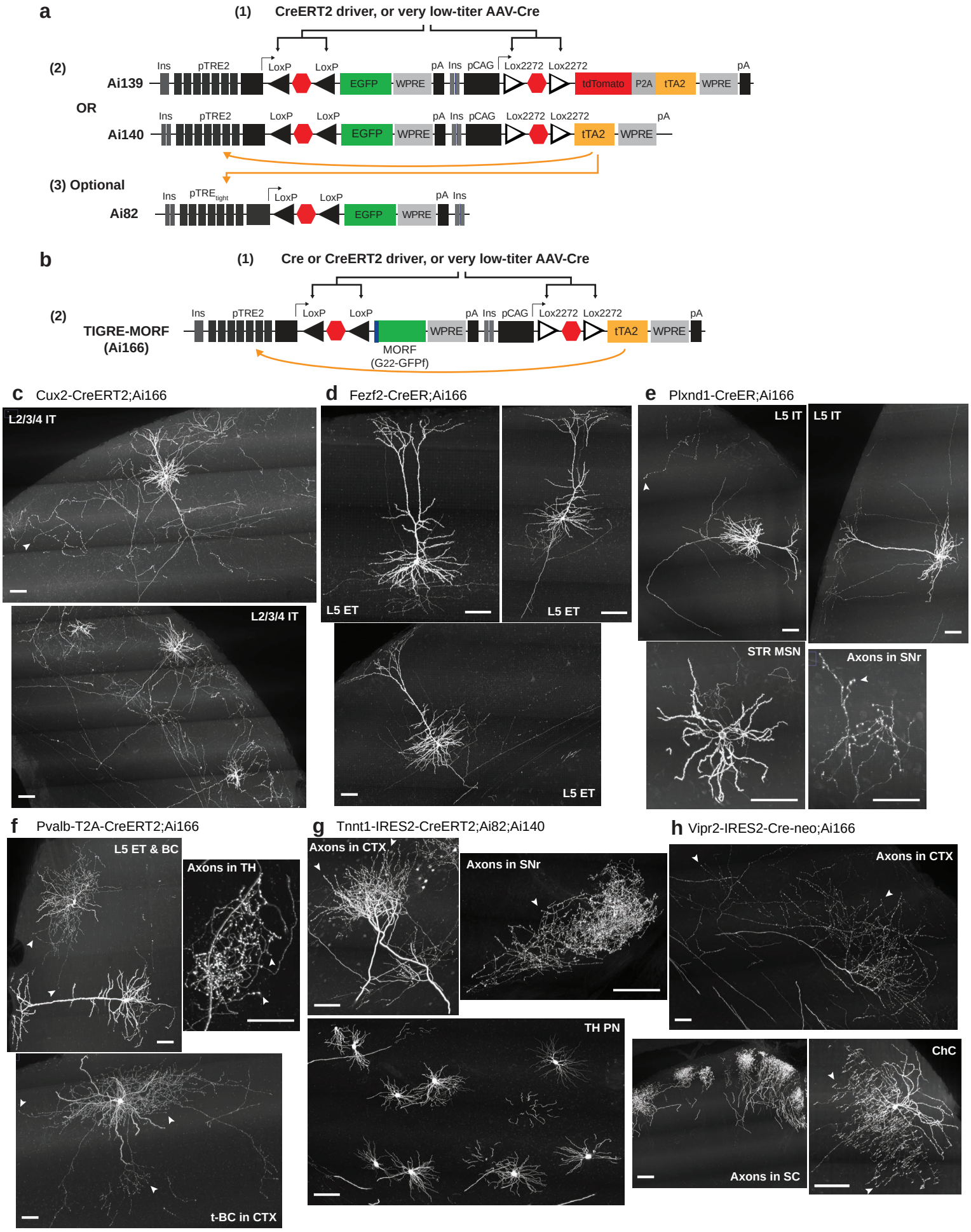
1161 25 Rotolo, T., Smallwood, P. M., Williams, J. & Nathans, J. Genetically-directed, cell type-
1162 specific sparse labeling for the analysis of neuronal morphology. *PLoS ONE* **3**, e4099,
1163 doi:10.1371/journal.pone.0004099 (2008).
1164 26 Aransay, A., Rodriguez-Lopez, C., Garcia-Amado, M., Clasca, F. & Prensa, L. Long-range
1165 projection neurons of the mouse ventral tegmental area: a single-cell axon tracing
1166 analysis. *Frontiers in neuroanatomy* **9**, 59, doi:10.3389/fnana.2015.00059 (2015).
1167 27 Ghosh, S. *et al.* Sensory maps in the olfactory cortex defined by long-range viral tracing
1168 of single neurons. *Nature* **472**, 217-220, doi:10.1038/nature09945 (2011).
1169 28 Kuramoto, E. *et al.* Two types of thalamocortical projections from the motor thalamic
1170 nuclei of the rat: a single neuron-tracing study using viral vectors. *Cereb Cortex* **19**, 2065-
1171 2077, doi:10.1093/cercor/bhn231 (2009).
1172 29 Economo, M. N. *et al.* A platform for brain-wide imaging and reconstruction of individual
1173 neurons. *eLife* **5**, e10566, doi:10.7554/eLife.10566 (2016).
1174 30 Lin, H. M. *et al.* Reconstruction of Intratelencephalic Neurons in the Mouse Secondary
1175 Motor Cortex Reveals the Diverse Projection Patterns of Single Neurons. *Frontiers in*
1176 *neuroanatomy* **12**, 86, doi:10.3389/fnana.2018.00086 (2018).
1177 31 Lin, R. *et al.* Cell-type-specific and projection-specific brain-wide reconstruction of single
1178 neurons. *Nature methods* **15**, 1033-1036, doi:10.1038/s41592-018-0184-y (2018).
1179 32 Economo, M. N. *et al.* Distinct descending motor cortex pathways and their roles in
1180 movement. *Nature* **563**, 79-84, doi:10.1038/s41586-018-0642-9 (2018).
1181 33 Kita, T. & Kita, H. The subthalamic nucleus is one of multiple innervation sites for long-
1182 range corticofugal axons: a single-axon tracing study in the rat. *J Neurosci* **32**, 5990-
1183 5999, doi:10.1523/JNEUROSCI.5717-11.2012 (2012).
1184 34 Kuramoto, E. *et al.* Ventral medial nucleus neurons send thalamocortical afferents more
1185 widely and more preferentially to layer 1 than neurons of the ventral anterior-ventral
1186 lateral nuclear complex in the rat. *Cereb Cortex* **25**, 221-235, doi:10.1093/cercor/bht216
1187 (2015).
1188 35 Ohno, S. *et al.* A morphological analysis of thalamocortical axon fibers of rat posterior
1189 thalamic nuclei: a single neuron tracing study with viral vectors. *Cereb Cortex* **22**, 2840-
1190 2857, doi:10.1093/cercor/bhr356 (2012).
1191 36 Winnubst, J. *et al.* Reconstruction of 1,000 Projection Neurons Reveals New Cell Types
1192 and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* **179**, 268-281
1193 e213, doi:10.1016/j.cell.2019.07.042 (2019).
1194 37 Wittner, L., Henze, D. A., Zaborszky, L. & Buzsaki, G. Three-dimensional reconstruction of
1195 the axon arbor of a CA3 pyramidal cell recorded and filled in vivo. *Brain structure &*
1196 *function* **212**, 75-83, doi:10.1007/s00429-007-0148-y (2007).
1197 38 Wu, H., Williams, J. & Nathans, J. Complete morphologies of basal forebrain cholinergic
1198 neurons in the mouse. *eLife* **3**, e02444, doi:10.7554/eLife.02444 (2014).
1199 39 Nakamura, H., Hioki, H., Furuta, T. & Kaneko, T. Different cortical projections from three
1200 subdivisions of the rat lateral posterior thalamic nucleus: a single-neuron tracing study
1201 with viral vectors. *The European journal of neuroscience* **41**, 1294-1310,
1202 doi:10.1111/ejn.12882 (2015).

1203 40 Gong, H. *et al.* High-throughput dual-colour precision imaging for brain-wide
1204 connectome with cytoarchitectonic landmarks at the cellular level. *Nature*
1205 *communications* **7**, 12142, doi:10.1038/ncomms12142 (2016).
1206 41 Badea, T. C. *et al.* New mouse lines for the analysis of neuronal morphology using
1207 CreER(T)/loxP-directed sparse labeling. *PLoS ONE* **4**, e7859,
1208 doi:10.1371/journal.pone.0007859 (2009).
1209 42 Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced
1210 Brain-Cell-Type Targeting and Functionality. *Cell* **174**, 465-480 e422,
1211 doi:10.1016/j.cell.2018.06.035 (2018).
1212 43 Veldman, M. B. *et al.* Brainwide Genetic Sparse Cell Labeling to Illuminate the
1213 Morphology of Neurons and Glia with Cre-Dependent MORF Mice. *Neuron*,
1214 doi:10.1016/j.neuron.2020.07.019 (2020).
1215 44 Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and
1216 effectors with high specificity and performance. *Neuron* **85**, 942-958,
1217 doi:10.1016/j.neuron.2015.02.022 (2015).
1218 45 Matho, K. S. *et al.* Genetic dissection of glutamatergic neuron subpopulations and
1219 developmental trajectories in the cerebral cortex. *bioRxiv*,
1220 doi:https://doi.org/10.1101/2020.04.22.054064 (2020).
1221 46 Martersteck, E. M. *et al.* Diverse Central Projection Patterns of Retinal Ganglion Cells.
1222 *Cell reports* **18**, 2058-2072, doi:10.1016/j.celrep.2017.01.075 (2017).
1223 47 Harris, J. A. *et al.* Anatomical characterization of Cre driver mice for neural circuit
1224 mapping and manipulation. *Frontiers in neural circuits* **8**, 76,
1225 doi:10.3389/fncir.2014.00076 (2014).
1226 48 Guo, Z. V. *et al.* Maintenance of persistent activity in a frontal thalamocortical loop.
1227 *Nature* **545**, 181-186, doi:10.1038/nature22324 (2017).
1228 49 Sherman, S. M. Thalamus plays a central role in ongoing cortical functioning. *Nature*
1229 *neuroscience* **19**, 533-541, doi:10.1038/nn.4269 (2016).
1230 50 Peng, H., Bria, A., Zhou, Z., Iannello, G. & Long, F. Extensible visualization and analysis
1231 for multidimensional images using Vaa3D. *Nature protocols* **9**, 193-208,
1232 doi:10.1038/nprot.2014.011 (2014).
1233 51 Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. V3D enables real-time 3D
1234 visualization and quantitative analysis of large-scale biological image data sets. *Nature*
1235 *biotechnology* **28**, 348-353, doi:10.1038/nbt.1612 (2010).
1236 52 Bria, A., Iannello, G., Onofri, L. & Peng, H. TeraFly: real-time three-dimensional
1237 visualization and annotation of terabytes of multidimensional volumetric images. *Nature*
1238 *methods* **13**, 192-194, doi:10.1038/nmeth.3767 (2016).
1239 53 Peng, H. *et al.* Virtual finger boosts three-dimensional imaging and microsurgery as well
1240 as terabyte volume image visualization and analysis. *Nature communications* **5**, 4342,
1241 doi:10.1038/ncomms5342 (2014).
1242 54 Wang, Y. *et al.* TeraVR empowers precise reconstruction of complete 3-D neuronal
1243 morphology in the whole brain. *Nature communications* **10**, 3474, doi:10.1038/s41467-
1244 019-11443-y (2019).
1245 55 Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference
1246 Atlas. *Cell* **181**, 936-953 e920, doi:10.1016/j.cell.2020.04.007 (2020).

56 Gerfen, C. R. & Surmeier, D. J. Modulation of striatal projection systems by dopamine. *Annual review of neuroscience* **34**, 441-466, doi:10.1146/annurev-neuro-061010-113641 (2011).
57 Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195-202, doi:10.1038/s41586-019-1716-z (2019).
58 Jones, E. G. Viewpoint: the core and matrix of thalamic organization. *Neuroscience* **85**, 331-345, doi:10.1016/s0306-4522(97)00581-2 (1998).
59 Hunnicutt, B. J. *et al.* A comprehensive thalamocortical projection map at the mesoscopic level. *Nature neuroscience* **17**, 1276-1285, doi:10.1038/nn.3780 (2014).
60 Bennett, C. *et al.* Higher-Order Thalamic Circuits Channel Parallel Streams of Visual Information in Mice. *Neuron* **102**, 477-492 e475, doi:10.1016/j.neuron.2019.02.010 (2019).
61 Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nature neuroscience* **18**, 170-181, doi:10.1038/nn.3917 (2015).
62 Narayanan, R. T., Udvary, D. & Oberlaender, M. Cell Type-Specific Structural Organization of the Six Layers in Rat Barrel Cortex. *Frontiers in neuroanatomy* **11**, 91, doi:10.3389/fnana.2017.00091 (2017).
63 Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D. & Shepherd, G. M. A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. *eLife* **3**, e05422, doi:10.7554/eLife.05422 (2014).
64 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *bioRxiv*, doi:https://doi.org/10.1101/2020.03.30.015214 (2020).
65 Zhang, M. *et al.* Molecular, spatial and projection diversity of neurons in primary motor cortex revealed by in situ single-cell transcriptomics. *bioRxiv*, doi:https://doi.org/10.1101/2020.06.04.105700 (2020).
66 Scala, F. *et al.* Phenotypic variation within and across transcriptomic cell types in mouse motor cortex. *bioRxiv*, doi:https://doi.org/10.1101/2020.02.03.929158 (2020).
67 Zhang, Z. *et al.* Epigenomic Diversity of Cortical Projection Neurons in the Mouse Brain. *bioRxiv*, doi:https://doi.org/10.1101/2020.04.01.019612 (2020).
68 Wang, Q. *et al.* Organization of the connections between claustrum and cortex in the mouse. *The Journal of comparative neurology* **525**, 1317-1346, doi:10.1002/cne.24047 (2017).
69 Bayer, S. A. & Altman, J. Development of the endopiriform nucleus and the claustrum in the rat brain. *Neuroscience* **45**, 391-412, doi:10.1016/0306-4522(91)90236-h (1991).
70 Smith, J. B. *et al.* The relationship between the claustrum and endopiriform nucleus: A perspective towards consensus on cross-species homology. *The Journal of comparative neurology* **527**, 476-499, doi:10.1002/cne.24537 (2019).
71 Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772-786 e719, doi:10.1016/j.cell.2019.09.023 (2019).
72 Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975-987, doi:10.1016/j.neuron.2016.07.036 (2016).

1290 73 Vavladeli, A., Daigle, T., Zeng, H., Crochet, S. & Petersen, C. C. H. Projection-specific
1291 Activity of Layer 2/3 Neurons Imaged in Mouse Primary Somatosensory Barrel Cortex
1292 During a Whisker Detection Task. *Function* **1**, zqaa008, doi:10.1093/function/zqaa008
1293 (2020).
1294 74 Kim, E. J. *et al.* Extraction of Distinct Neuronal Cell Types from within a Genetically
1295 Continuous Population. *Neuron* **107**, 274-282 e276, doi:10.1016/j.neuron.2020.04.018
1296 (2020).
1297 75 Li, H. *et al.* Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell
1298 RNA Sequencing. *Cell* **171**, 1206-1220 e1222, doi:10.1016/j.cell.2017.10.019 (2017).
1299 76 Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain
1300 imaging. *Nature methods* **9**, 255-258, doi:10.1038/nmeth.1854 (2012).
1301 77 Xiong, H. *et al.* Chemical reactivation of quenched fluorescent protein molecules
1302 enables resin-embedded fluorescence microimaging. *Nature communications* **5**, 3992,
1303 doi:10.1038/ncomms4992 (2014).
1304 78 Gang, Y. *et al.* Embedding and Chemical Reactivation of Green Fluorescent Protein in the
1305 Whole Mouse Brain for Optical Micro-Imaging. *Frontiers in neuroscience* **11**, 121,
1306 doi:10.3389/fnins.2017.00121 (2017).
1307 79 Li, A. *et al.* Micro-optical sectioning tomography to obtain a high-resolution atlas of the
1308 mouse brain. *Science (New York, N.Y* **330**, 1404-1408, doi:10.1126/science.1191776
1309 (2010).
1310 80 Qu, L. & Peng, H. A principal skeleton algorithm for standardizing confocal images of
1311 fruit fly nervous systems. *Bioinformatics* **26**, 1091-1097,
1312 doi:10.1093/bioinformatics/btq072 (2010).
1313 81 Peng, H. *et al.* BrainAligner: 3D registration atlases of Drosophila brains. *Nature methods*
1314 **8**, 493-500, doi:10.1038/nmeth.1602 (2011).
1315 82 Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on pattern*
1316 *analysis and machine intelligence* **22**, 888-905 (2000).
1317 83 Blume, M. E. Betas and Their Regression Tendencies. *The Journal of Finance* **30**, 785-795
1318 (1975).
1319 84 Tervo, D. G. *et al.* A Designer AAV Variant Permits Efficient Retrograde Access to
1320 Projection Neurons. *Neuron* **92**, 372-382, doi:10.1016/j.neuron.2016.09.021 (2016).
1321 85 Chatterjee, S. *et al.* Nontoxic, double-deletion-mutant rabies viral vectors for retrograde
1322 targeting of projection neurons. *Nature neuroscience* **21**, 638-646, doi:10.1038/s41593-
1323 018-0091-7 (2018).
1324 86 Hnasko, T. S. *et al.* Cre recombinase-mediated restoration of nigrostriatal dopamine in
1325 dopamine-deficient mice reverses hypophagia and bradykinesia. *Proceedings of the*
1326 *National Academy of Sciences of the United States of America* **103**, 8858-8863,
1327 doi:10.1073/pnas.0603081103 (2006).
1328
1329
1330

Figure 1


a
(1) CreERT2 driver, or very low-titer AAV-Cre
(2)
Ai139
Ins pTRE2 LoxP LoxP EGFP WPRE pA Ins pCAG Lox2272 Lox2272 tdTomato P2A tTA2 WPRE pA
OR
Ai140
Ins pTRE2 LoxP LoxP EGFP WPRE pA Ins pCAG Lox2272 Lox2272 tTA2 WPRE pA
(3) Optional
Ai82
Ins pTRE tight LoxP LoxP EGFP WPRE pA Ins
b
(1) Cre or CreERT2 driver, or very low-titer AAV-Cre
(2)
TIGRE-MORF (Ai166)
Ins pTRE2 LoxP LoxP MORF (G22-GFPf) WPRE pA Ins pCAG Lox2272 Lox2272 tTA2 WPRE pA
c Cux2-CreERT2;Ai166
L2/3/4 IT
L2/3/4 IT
d Fezf2-CreER;Ai166
L5 ET
L5 ET
L5 ET
e Plxnd1-CreER;Ai166
L5 IT
L5 IT
STR MSN
Axons in SNr
f Pvalb-T2A-CreERT2;Ai166
L5 ET & BC
Axons in TH
t-BC in CTX
g Tnnt1-IRES2-CreERT2;Ai82;Ai140
Axons in CTX
Axons in SNr
TH PN
h Vipr2-IRES2-Cre-neo;Ai166
Axons in CTX
Axons in SC
ChC

1331 **Figure 1. Sparse, robust and consistent labeling and visualization of the dendritic and**
1332 **axonal arborizations of a wide range of neuronal types. a,** Schematic diagram showing the
1333 combination of CreERT2 transgenic driver line or Cre-expressing AAV (1) with the GFP-
1334 expressing TIGRE2.0 reporter line Ai139 or Ai140 (2). Very low dose tamoxifen induction of
1335 CreERT2 or very low-titer AAV-Cre delivery results in activation of the reporter in a spatially
1336 sparse manner. Transgenic reporter expression of GFP is robust and consistent across different
1337 cells. An optional addition is to cross in the GFP-expressing TIGRE1.0 reporter line Ai82 (3), so
1338 that the tTA2 from Ai139 or Ai140 will activate the expression of GFP from two alleles –
1339 Ai139/Ai140 and Ai82, further increasing the level of GFP within Cre+ cells. **b**, Schematic
1340 diagram showing the combination of Cre or CreERT2 transgenic driver line or Cre-expressing
1341 AAV (1) with the GFP-expressing sparse reporter line TIGRE-MORF/Ai166 (2). Due to the
1342 intrinsic sparse expression of MORF (G_{22}-GFPf), some conventional Cre lines, moderate doses
1343 of tamoxifen induction of CreERT2, or moderate titers of AAV-Cre delivery can result in very
1344 sparse labeling. **c,** Cortical L2/3/4 IT neurons and their extensive local axon collaterals clearly
1345 labeled in a Cux2-CreERT2;Ai166 brain. **d,** Cortical L5 ET neurons and their sparse local axon
1346 collaterals seen in a Fezf2-CreER;Ai166 brain. **e,** Cortical L5 IT neurons and their local axon
1347 collaterals seen in a Plxnd1-CreER;Ai166 brain. Striatal medium spiny neurons (STR MSN) are
1348 also sparsely labeled, and their individual axons are clearly seen in substantia nigra (SN). **f,**
1349 Cortical inhibitory basket cells (BC) and translaminar basket cells (t-BC), as well as L5 ET
1350 excitatory neurons, seen in a Pvalb-T2A-CreERT2;Ai166 brain. The L5 ET neurons form
1351 driving-type axon clusters with large boutons in the thalamus (TH). **g,** Thalamic projection
1352 neurons (TH PN) with their dense axon terminal clusters in cortex seen in a Tnnt1-IRES2-
1353 CreERT2;Ai82;Ai140 brain. Some STR MSNs are also labeled and they form intense axon
1354 clusters in SN. **h,** In a Vipr2-IRES2-Cre-neo;Ai166 brain, axon clusters from projection neurons
1355 in visual thalamic nuclei are seen in CTX, axon clusters likely from retinal ganglion cells are
1356 seen in superior colliculus (SC), and a cortical chandelier cell (ChC) is fully labeled with its
1357 characteristic axonal branches. Images shown in c-h are 100-µm maximum intensity projection
1358 (MIP) images (*i.e.*, projected from 100 consecutive 1-µm image planes). Arrowheads indicate
1359 observed terminal boutons at the end of the axon segments. Tamoxifen doses are shown in
1360 **Supplementary Table 1**. Scale bars, 100 µm.

1361

1362

Figure 2

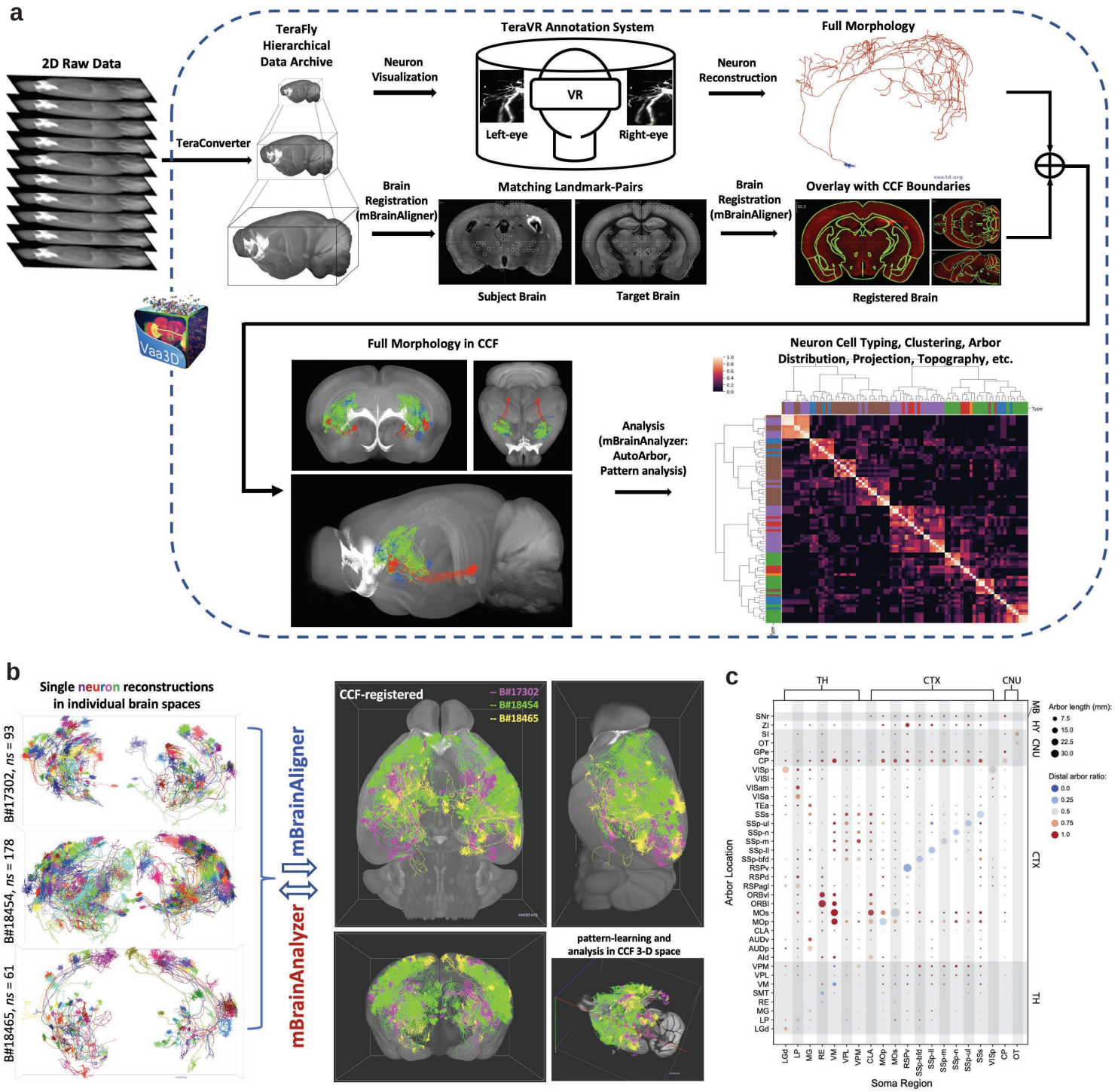

a
2D Raw Data
TeraFly Hierarchical Data Archive
TeraConverter
Neuron Visualization
TeraVR Annotation System
VR
Left-eye
Right-eye
Neuron Reconstruction
Full Morphology
Brain Registration (mBrainAligner)
Matching Landmark-Pairs
Subject Brain
Target Brain
Brain Registration (mBrainAligner)
Overlay with CCF Boundaries
Registered Brain
Vaa3D
Full Morphology in CCF
Analysis (mBrainAnalyzer: AutoArbor, Pattern analysis)
Neuron Cell Typing, Clustering, Arbor Distribution, Projection, Topography, etc.
b
Single neuron reconstructions in individual brain spaces
B#17302, ns = 93
B#18454, ns = 178
B#18465, ns = 61
mBrainAligner
mBrainAnalyzer
CCF-registered
B#17302
B#18454
B#18465
pattern-learning and analysis in CCF 3-D space
c
TH
CTX
CNU
MB
HY
Arbor Location
Soma Region
Arbor length (mm):
7.5
15.0
22.5
30.0
Distal arbor ratio:
0.0
0.25
0.5
0.75
1.0

Figure 2. Platform and workflow of the brain-wide full morphology reconstruction, registration and analysis pipeline. a, The workflow of neuron visualization, reconstruction, mapping to Common Coordinate Framework (CCF) and analysis. A complete fMOST image dataset is first converted to TeraFly file format by TeraConverter, the data formatting tool in TeraFly. Then annotators work in the TeraVR annotation system to reconstruct the full morphology of each neuron. In parallel, the whole brain image dataset is registered to CCF using mBrainAligner, which uses both RLM (Reliable-Landmark-Matching) and LQM (Little-Quick-Warp) modules in brain alignment. Following registration of the image dataset to CCFv3, all the reconstructed morphologies from the same brain are also registered for subsequent visualization and quantitative analysis. **b**, Demonstration of single neuron reconstructions shown in varying colors from individual mouse brains registered to the CCF space by mBrainAligner, which allows integrated analysis by the morphology analysis toolbox, mBrainAnalyzer. **c**, Overview of projection patterns. Cells are grouped by curated soma locations, first by major brain areas (e.g. CTX: cortex, TH: thalamus, CNU: Cerebral nuclei) and then by refined areas (e.g. LGd nucleus of the thalamus). Sizes of dots represent group-average of arbor length. Colors represent the ratio of dendritic arbors (red: low; blue: high). Note that some minor "projections" indicated by tiny dots may be false positives due to passing fibers or not completely precise registration.

Figure 3

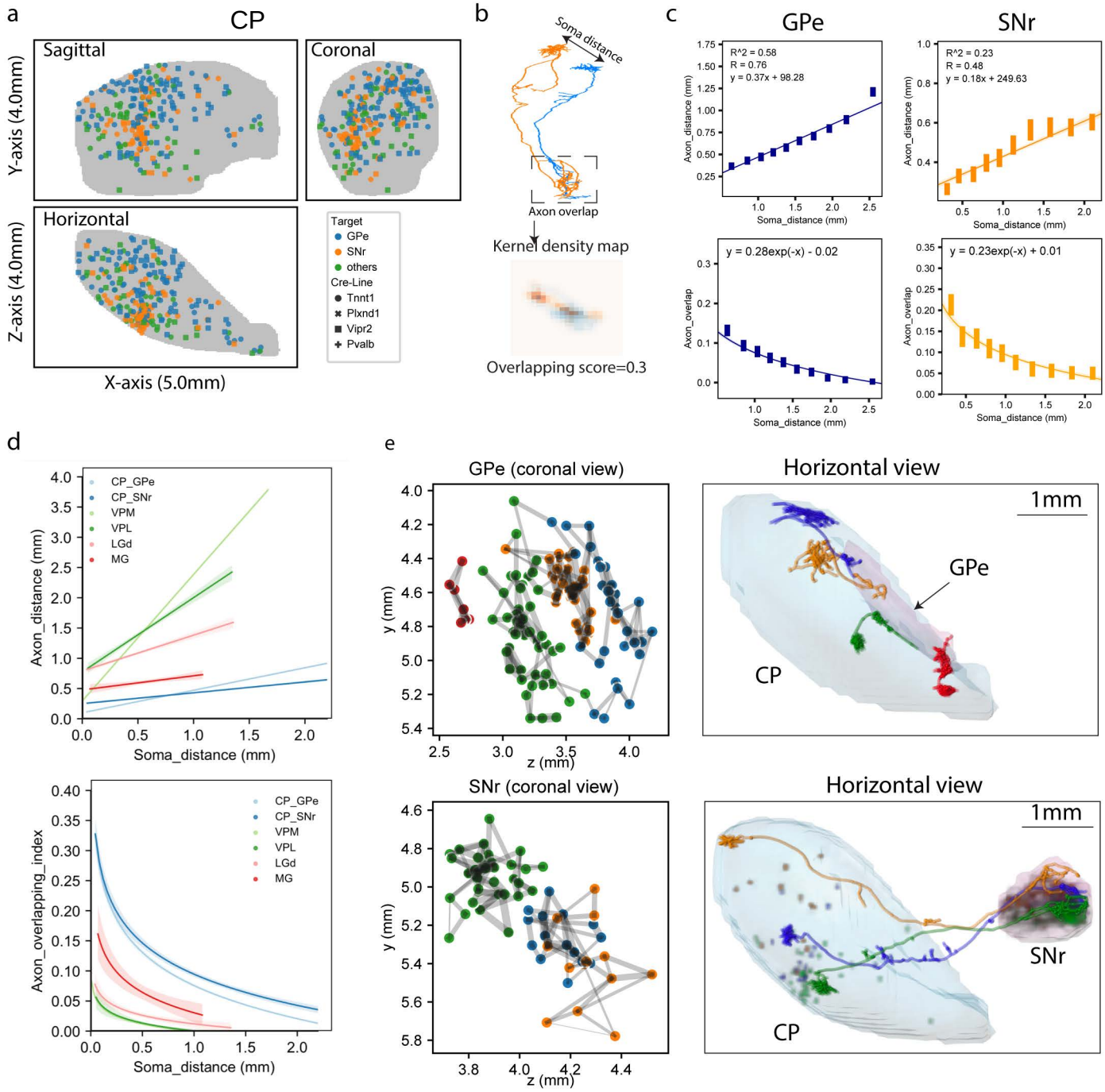

a
CP
Sagittal
Coronal
Horizontal
Y-axis (4.0mm)
Z-axis (4.0mm)
X-axis (5.0mm)
Target
GPe
SNr
others
Cre-Line
Tnnt1
Plxnd1
Vipr2
Pvalb
b
Soma distance
Axon overlap
Kernel density map
Overlapping score=0.3
c
GPe
SNr
R^2 = 0.58
R = 0.76
y = 0.37x + 98.28
R^2 = 0.23
R = 0.48
y = 0.18x + 249.63
Axon_distance (mm)
Soma_distance (mm)
y = 0.28exp(-x) - 0.02
y = 0.23exp(-x) + 0.01
Axon_overlap
d
CP_GPe
CP_SNr
VPM
VPL
LGd
MG
Axon_distance (mm)
Soma_distance (mm)
Axon_overlapping_index
e
GPe (coronal view)
Horizontal view
1mm
GPe
CP
y (mm)
z (mm)
SNr (coronal view)
SNr

Figure 3. Striatal neuron morphology analysis. a, Coronal, sagittal and horizontal views of soma distribution of CP neurons (X-axis: anterior to posterior, Y-axis: dorsal to ventral, Z-axis: medial to lateral. The same axis labels were used throughout the paper). **b,** Overlapping score of axons is calculated by estimating the kernel density map of individual axon arbors and the density-weighted average of overlapping areas for each arbor pair. **c,** Regression of distance between arbor centers (top panels) or overlapping score (bottom panels) by soma distance. Linear and negative exponential models are used for distance and overlapping score, respectively. **d,** Comparison of arbor convergence across cell types. Regression curves generated by the same approach as in c. Colors represent cell types. Light-shaded bands represent 95% confidence intervals. **e,** Clustering of axon overlapping by Louvain algorithm. Left panels, coronal views of axon arbor locations colored by clusters. Width of grey lines represents overlapping scores between arbor pairs. Right panels, representative single neurons to illustrate topography of CP neuron projections. Cells are colored by cluster identities.

Figure 4

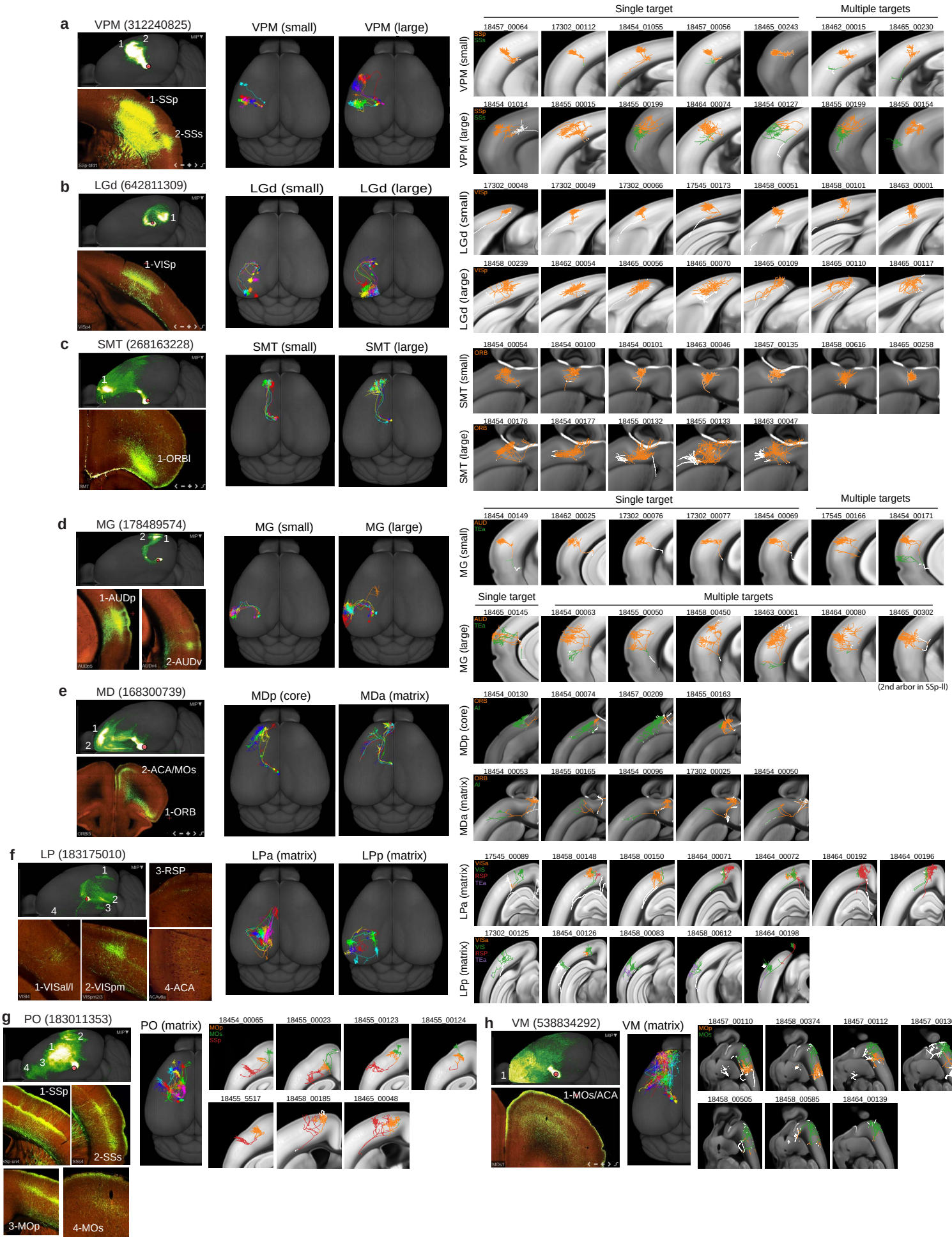

a
VPM (312240825)
1-SSp
2-SSs
VPM (small)
VPM (large)
Single target
Multiple targets
VPM (small)
VPM (large)
b
LGd (642811309)
1-VISp
LGd (small)
LGd (large)
LGd (small)
LGd (large)
c
SMT (268163228)
1-ORBl
SMT (small)
SMT (large)
SMT (small)
SMT (large)
d
MG (178489574)
1-AUDp
2-AUDv
MG (small)
MG (large)
Single target
Multiple targets
MG (small)
MG (large)
(2nd arbor in SSp-ll)
e
MD (168300739)
2-ACA/MOs
1-ORB
MDp (core)
MDa (matrix)
MDp (core)
MDa (matrix)
f
LP (183175010)
3-RSP
1-VISal/l
2-VISpm
4-ACA
LPa (matrix)
LPp (matrix)
LPa (matrix)
LPp (matrix)
g
PO (183011353)
1-SSp
2-SSs
3-MOp
4-MOs
PO (matrix)
h
VM (538834292)
1-MOs/ACA
VM (matrix)

Figure 4. Long-range projection patterns of individual thalamic neurons in comparison with mesoscale population-level projections. a-h, Axonal morphologies and projections of reconstructed single neurons compared with population projection patterns for nucleus VPM (a), LGd (b), SMT (c), MG (d), MD (e), LP (f), PO (g) and VM (h). For each nucleus, left panels, representative mesoscale experiments shown in a maximum projection whole-brain top-down view and individual higher-power images showing axon termination patterns in major target regions; middle panels, all example single neurons shown together in a maximum projection whole-brain top-down view; right panels, each neuron is shown in a chosen plane to best capture the perpendicular (to pial surface) orientation of the main axon arbor with superimposed maximum projection view of the neuron's axon arbors. The chosen plane can be coronal (for a, b, d), horizontal (for c, e), sagittal (for h) or tilted (for f, g), based on the main cortical target region. Different cortical target regions are indicated by different colors. Small, small axon arbors. Large, large axon arbors. MDa and MDp, or LPa and LPp are the anterior and posterior parts of MD or LP respectively.

Figure 5

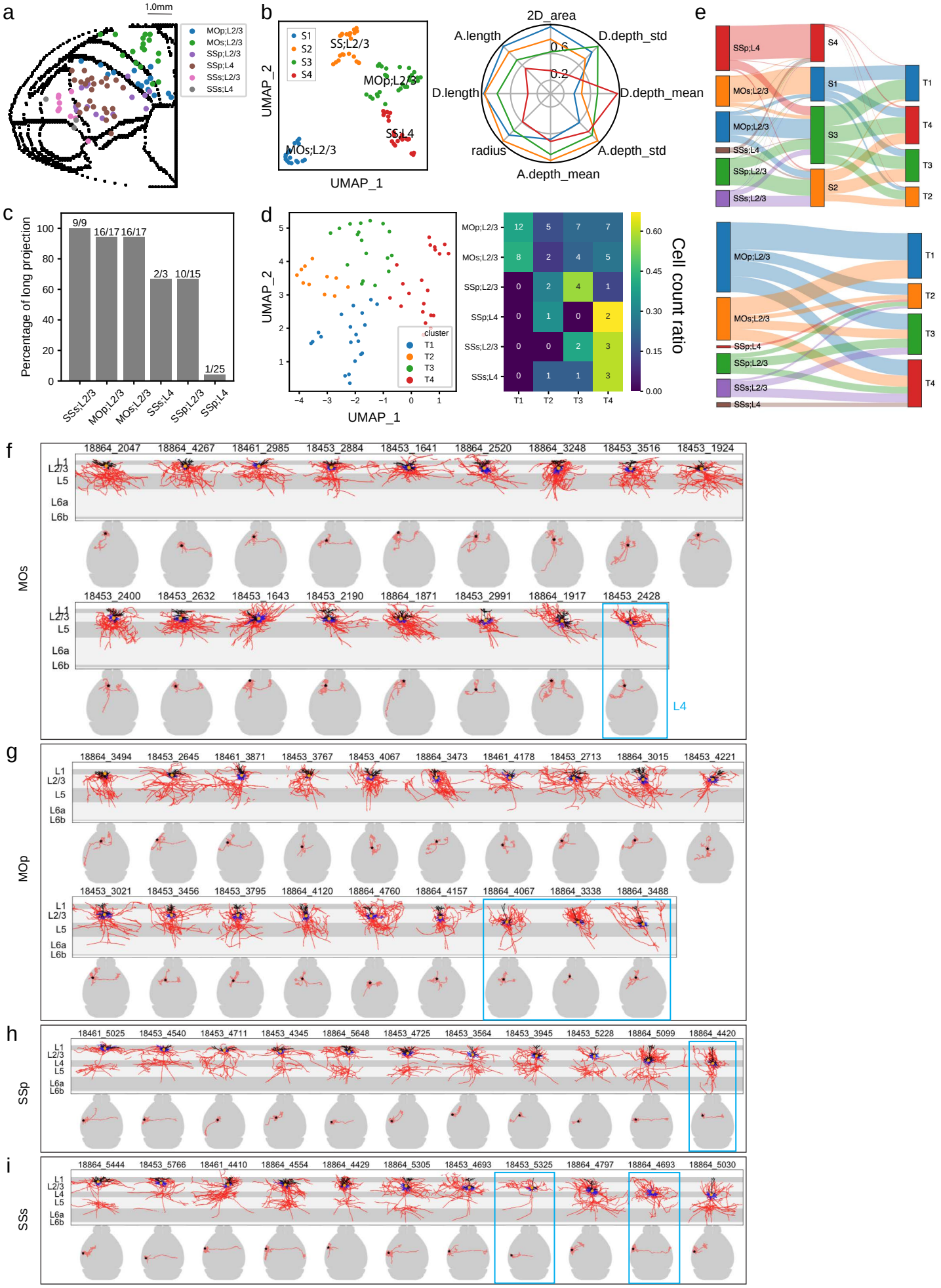

a
1.0mm
MOp;L2/3
MOs;L2/3
SSp;L2/3
SSp;L4
SSs;L2/3
SSs;L4
b
S1
S2
S3
S4
SS;L2/3
MOp;L2/3
SS;L4
MOs;L2/3
UMAP_2
UMAP_1
2D_area
A.length
D.depth_std
D.length
D.depth_mean
radius
A.depth_std
A.depth_mean
0.6
0.2
e
SSp;L4
MOs;L2/3
MOp;L2/3
SSs;L4
SSp;L2/3
SSs;L2/3
S4
S1
S3
S2
T1
T4
T3
T2
MOp;L2/3
MOs;L2/3
SSp;L4
SSp;L2/3
SSs;L2/3
SSs;L4
T1
T2
T3
T4
c
Percentage of long projection
9/9
16/17
16/17
2/3
10/15
1/25
SSs;L2/3
MOp;L2/3
MOs;L2/3
SSs;L4
SSp;L2/3
SSp;L4
d
UMAP_2
UMAP_1
cluster
T1
T2
T3
T4
MOp;L2/3 12 5 7 7
MOs;L2/3 8 2 4 5
SSp;L2/3 0 2 4 1
SSp;L4 0 1 0 2
SSs;L2/3 0 0 2 3
SSs;L4 0 1 1 3
T1 T2 T3 T4
Cell count ratio
0.60
0.45
0.30
0.15
0.00
f
MOs
18864_2047 18864_4267 18461_2985 18453_2884 18453_1641 18864_2520 18864_3248 18453_3516 18453_1924
18453_2400 18453_2632 18453_1643 18453_2190 18864_1871 18453_2991 18864_1917 18453_2428
L1
L2/3
L5
L6a
L6b
L4
g
MOp
18864_3494 18453_2645 18461_3871 18453_3767 18453_4067 18864_3473 18461_4178 18453_2713 18864_3015 18453_4221
18453_3021 18453_3456 18453_3795 18864_4120 18864_4760 18864_4157 18864_4067 18864_3338 18864_3488
L1
L2/3
L5
L6a
L6b
h
SSp
18461_5025 18453_4540 18453_4711 18453_4345 18864_5648 18453_4725 18453_3564 18453_3945 18453_5228 18864_5099 18864_4420
L1
L2/3
L4
L5
L6a
L6b
i
SSs
18864_5444 18453_5766 18461_4410 18864_4554 18864_4429 18864_5305 18453_4693 18453_5325 18864_4797 18864_4693 18864_5030
L1
L2/3
L4
L5
L6a
L6b

Figure 5. Local morphological and long-range projection analysis of cortical L2/3/4 IT neurons. a, Cortical surface flatmap showing the soma locations of reconstructed L2/3/4 IT neurons from MOp, MOs, SSp and SSs. **b**, Clustering based on local dendritic, axonal and soma location features divides L2/3/4 IT cells into 4 clusters. UMAP dimension reduction was performed, followed by k-means clustering using UMAP embeddings as input features. Clustering results shown by the UMAP representation and polar plot of main features. **c**, Percentage of cells from each region and each layer that have long-range projections. **d**, Clustering based on long-range projection targets, shown by the UMAP representation and confusion matrix of soma location and projection clusters. Clustering is performed by UMAP embedding and k-means. **e**, Sankey plots showing the correspondence among soma locations, local clusters and long-range projection clusters for individual neurons. **f-i**, Comparison of local morphologies (upper panels; apical dendrite in black, basal dendrite in blue, axon in red, soma as an orange dot) and whole-brain projections (lower panels; axon in red, soma as a star) for MOs (f), MOp (g), SSp (h) and SSs (i) neurons. Neurons are ordered based on the depths from pial surface of their somas. L4 IT neurons are marked by the blue boxes. The others are L2/3 IT neurons.

Figure 6

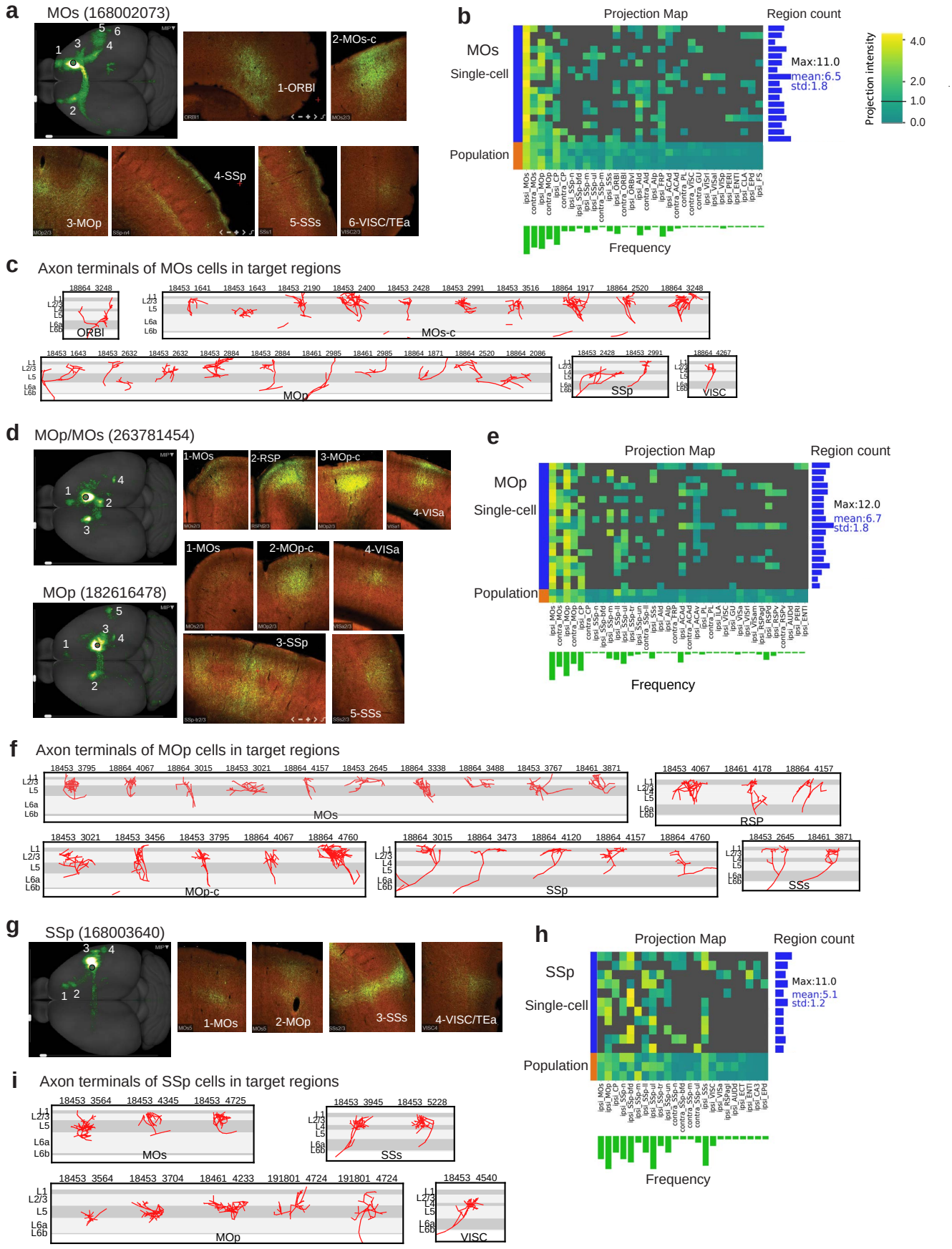

a
MOs (168002073)
1-ORBl
2-MOs-c
3-MOp
4-SSp
5-SSs
6-VISC/TEa
b
Projection Map
Region count
MOs
Single-cell
Population
Max:11.0
mean:6.5
std:1.8
Projection intensity
4.0
2.0
1.0
0.0
Frequency
c
Axon terminals of MOs cells in target regions
ORBl
MOs-c
MOp
SSp
VISC
d
MOp/MOs (263781454)
1-MOs
2-RSP
3-MOp-c
4-VISa
MOp (182616478)
1-MOs
2-MOp-c
4-VISa
3-SSp
5-SSs
e
Projection Map
Region count
MOp
Single-cell
Population
Max:12.0
mean:6.7
std:1.8
Frequency
f
Axon terminals of MOp cells in target regions
MOs
RSP
MOp-c
SSp
SSs
g
SSp (168003640)
1-MOs
2-MOp
3-SSs
4-VISC/TEa
h
Projection Map
Region count
SSp
Single-cell
Population
Max:11.0
mean:5.1
std:1.2
Frequency
i
Axon terminals of SSp cells in target regions
MOs
SSs
MOp
VISC

Figure 6. Comparison of long-range projection patterns between individual cortical L2/3/4 IT neurons and mesoscale population-level projections. a, A representative mesoscale experiment from MOs, shown in a maximum projection whole-brain top-down view and individual higher power images showing axon termination patterns in major target regions. **b**, Projection matrix of MOs single cells and population-level mesoscale experiments (168002073, 266645328, 272822110 and 587659400) along with some basic statistics for the single cells (i.e. target region counts per cell, frequencies per target region). Projection strengths are scaled to 0-4 and visualized with a cutoff at 1 to minimize spurious false positives due to issues such as imaging artifact (in mesoscale experiments), imprecise registration and minor passing fibers. Cells are ordered by soma depth from pial surface. Only brain regions targeted (relative projection strength >1) by any single/mesoscale samples are included. **c**, Axon terminals of MOs neurons in specified target regions. Because not all neurons project to all target regions, all detected axon terminals from any neurons for each target region are shown here. MOs-c, contralateral MOs. **d-f**, Same panels as a-c but for MOp. Two representative mesoscale experiments (182616478 and 263781454) from MOp are shown in d, which are also used as population projections in e. **g-i**, Same panels as a-c but for SSp. Mesoscale experiments used as population projections in h are 168003640, 298830161 and 278317945.

Figure 7

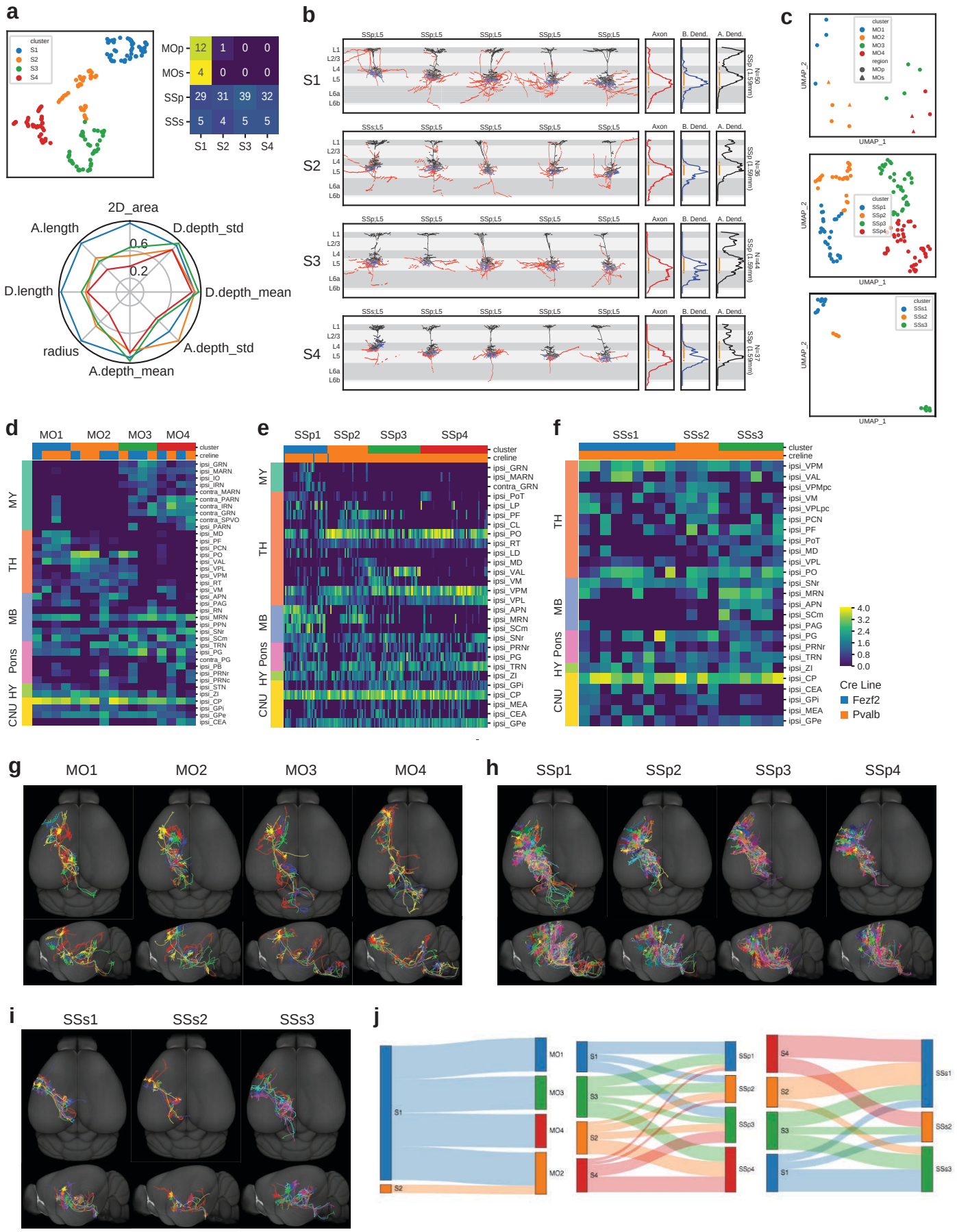

a
cluster
S1
S2
S3
S4
MOp 12 1 0 0
MOs 4 0 0 0
SSp 29 31 39 32
SSs 5 4 5 5
S1 S2 S3 S4
2D_area
A.length
D.depth_std
0.6
0.2
D.length
D.depth_mean
radius
A.depth_std
A.depth_mean
b
S1
S2
S3
S4
SSp;L5
SSs;L5
Axon
B. Dend.
A. Dend.
L1
L2/3
L4
L5
L6a
L6b
N=50
N=36
N=44
N=37
c
cluster
MO1
MO2
MO3
MO4
region
MOp
MOs
SSp1
SSp2
SSp3
SSp4
SSs1
SSs2
SSs3
UMAP_1
UMAP_2
d
MO1 MO2 MO3 MO4
cluster
creline
MY
TH
MB
Pons
HY
CNU
ipsi_GRN
ipsi_MARN
ipsi_IO
ipsi_IRN
contra_MARN
contra_PARN
contra_IRN
contra_GRN
contra_SPVO
ipsi_PARN
ipsi_MD
ipsi_PF
ipsi_PCN
ipsi_PO
ipsi_VAL
ipsi_VPL
ipsi_VPM
ipsi_RT
ipsi_VM
ipsi_APN
ipsi_PAG
ipsi_RN
ipsi_MRN
ipsi_PPN
ipsi_SNr
ipsi_SCm
ipsi_TRN
ipsi_PG
contra_PG
ipsi_PB
ipsi_PRNr
ipsi_PRNc
ipsi_STN
ipsi_ZI
ipsi_CP
ipsi_GPi
ipsi_GPe
ipsi_CEA
e
SSp1 SSp2 SSp3 SSp4
cluster
creline
MY
TH
MB
Pons
HY
CNU
ipsi_GRN
ipsi_MARN
contra_GRN
ipsi_PoT
ipsi_LP
ipsi_PF
ipsi_CL
ipsi_PO
ipsi_RT
ipsi_LD
ipsi_MD
ipsi_VAL
ipsi_VM
ipsi_VPM
ipsi_VPL
ipsi_APN
ipsi_MRN
ipsi_SCm
ipsi_SNr
ipsi_PRNr
ipsi_PG
ipsi_TRN
ipsi_ZI
ipsi_GPi
ipsi_CP
ipsi_MEA
ipsi_CEA
ipsi_GPe
f
SSs1 SSs2 SSs3
cluster
creline
TH
MB
Pons
HY
CNU
ipsi_VPM
ipsi_VAL
ipsi_VPMpc
ipsi_VM
ipsi_VPLpc
ipsi_PCN
ipsi_PF
ipsi_PoT
ipsi_MD
ipsi_VPL
ipsi_PO
ipsi_SNr
ipsi_MRN
ipsi_APN
ipsi_SCm
ipsi_PAG
ipsi_PG
ipsi_PRNr
ipsi_TRN
ipsi_ZI
ipsi_CP
ipsi_CEA
ipsi_GPi
ipsi_MEA
ipsi_GPe
4.0
3.2
2.4
1.6
0.8
0.0
Cre Line
Fezf2
Pvalb
g
MO1 MO2 MO3 MO4
h
SSp1 SSp2 SSp3 SSp4
i
SSs1 SSs2 SSs3
j
S1
S2
MO1
MO3
MO4
MO2
S1
S3
S2
S4
SSp1
SSp2
SSp3
SSp4
S4
S2
S3
S1
SSs1
SSs2
SSs3

Figure 7. Local morphological and long-range projection analysis of cortical L5 ET neurons. a, Clustering based on local dendritic, axonal and soma location features divides L5 ET cells into 4 clusters, as shown by the UMAP representation, soma-cluster confusion matrix and polar plot of main features. Clustering approach is the same as used for Fig. 5b. **b**, Local morphologies of example neurons and average vertical profiles for each local cluster. Broken lines are due to the substantially tilted nature of some of these neurons. **c**, Clustering based on long-range projection targets, carried out separately for each region (MOp/MOs combined, SSp and SSs), as shown by the UMAP representations. **d-f**, Projection matrix heatmaps for MOp/MOs (d), SSp (e) and SSs (f) neurons, showing representative target brain regions of each neuron. Columns represent single cells sorted by cluster assignments. Rows represent target regions. **g-i**, Whole-brain projection overview of individual neurons in each cluster for MOp/MOs (g), SSp (h) and SSs (i). **j**, Sankey plots showing the correspondence between local clusters and long-range projection clusters for individual neurons.

Figure 8

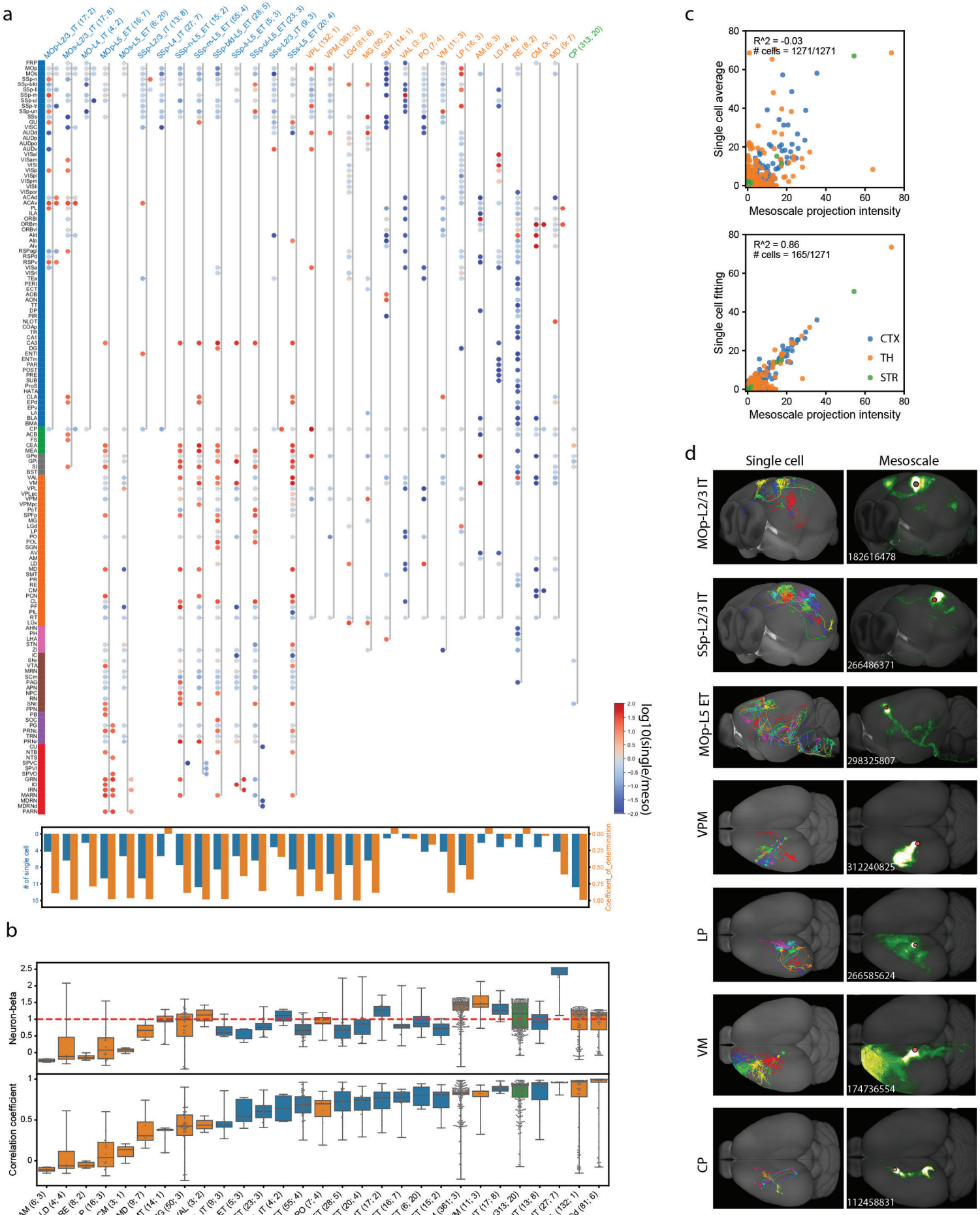

a
log10(single/mesos)
# of single cell
Coefficient_of_determination
b
Neuron-beta
Correlation coefficient
c
R^2 = -0.03
cells = 1271/1271
Single cell average
Mesoscale projection intensity
R^2 = 0.86
cells = 165/1271
Single cell fitting
CTX
TH
STR
d
Single cell
Mesoscale
MOp-L2/3 IT
182616478
SSp-L2/3 IT
266486371
MOp-L5 ET
298325807
VPM
312240825
LP
266585624
VM
174736554
CP
112458831

Figure 8. Combination of single neuron morphologies recapitulates population-level mesoscale projection patterns. a, Comparative projection map of single cell and mesoscale data. Individual samples are grouped by brain areas and/or cortical layers based on soma locations (single cell) and injection sites (mesoscale). Each group is represented by a stretch of connected dots with ipsilateral and contralateral targets on the left and right hemisphere, respectively. Projection intensities are quantified as log(percentage+1). Selected regions are defined at a cutoff 0.5 and targeted by at least 50% mesoscale or 10% single cells. Dot colors are scaled by the log10 of single cell and mesoscale strength ratio. (lower panel) Coefficients of determination (orange bars) and number of cells (blue bars) of mesoscale regression by single cells (described in c). **b**, Boxplots of neuron-beta and correlation coefficients between single cells and group-average of mesoscale data. Individual comparisons shown as swarm plots overlapped with boxes. The first and second numbers in the group labels in a and b indicate the numbers of single cells and mesoscale experiments, respectively. **c**, Approximation of mesoscale projections by single cell projection strengths (1,271 cells used) by group-average (upper) or by linear regression (lower). LASSO regularity was applied during regression, to reduce the number of single cells with non-zero weights (representative cells). **d**, Visualization of projection patterns constituted by representative cells and mesoscale projection intensities.

Figure 9

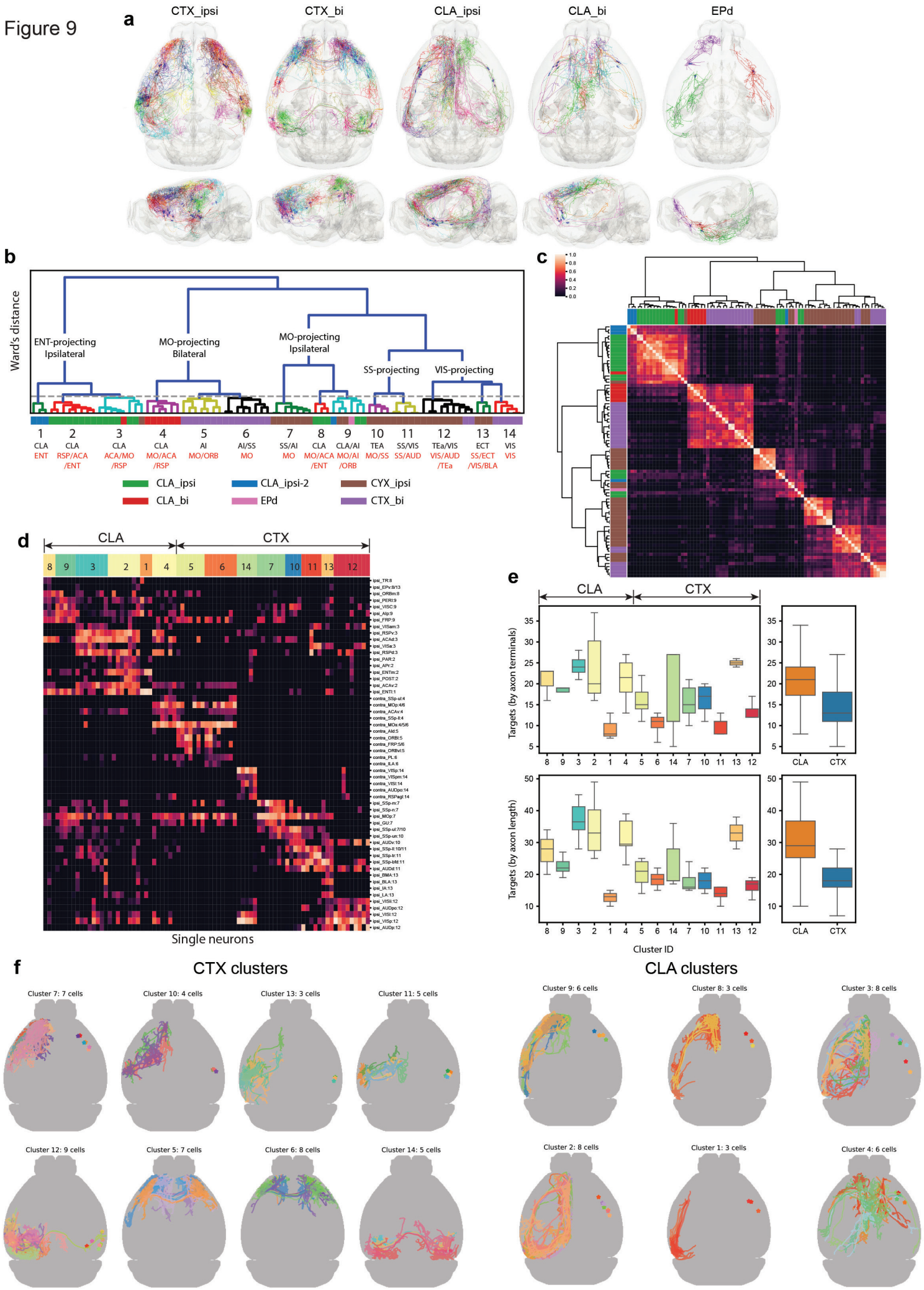

a
CTX_ipsi
CTX_bi
CLA_ipsi
CLA_bi
EPd
b
Ward's distance
ENT-projecting Ipsilateral
MO-projecting Bilateral
MO-projecting Ipsilateral
SS-projecting
VIS-projecting
1 CLA ENT
2 CLA RSP/ACA /ENT
3 CLA ACA/MO /RSP
4 CLA MO/ACA /RSP
5 AI MO/ORB
6 AI/SS MO
7 SS/AI MO
8 CLA MO/ACA /ENT
9 CLA/AI MO/AI /ORB
10 TEA MO/SS
11 SS/VIS SS/AUD
12 TEa/VIS VIS/AUD /TEa
13 ECT SS/ECT /VIS/BLA
14 VIS VIS
CLA_ipsi
CLA_ipsi-2
CYX_ipsi
CLA_bi
EPd
CTX_bi
c
d
CLA
CTX
Single neurons
e
CLA
CTX
Targets (by axon terminals)
Targets (by axon length)
Cluster ID
f
CTX clusters
CLA clusters
Cluster 7: 7 cells
Cluster 10: 4 cells
Cluster 13: 3 cells
Cluster 11: 5 cells
Cluster 9: 6 cells
Cluster 8: 3 cells
Cluster 3: 8 cells
Cluster 12: 9 cells
Cluster 5: 7 cells
Cluster 6: 8 cells
Cluster 14: 5 cells
Cluster 2: 8 cells
Cluster 1: 3 cells
Cluster 4: 6 cells

Figure 9. Extensive projection diversity of the L6 Car3 subclass of cortical and claustral neurons. a, Individual L6 Car3 neurons are shown in manually divided groups: CTX_ipsi (n=33), cortical neurons projecting to ipsilateral cortex only; CTX_bi (n=26), cortical neurons projecting bilaterally; CLA_ipsi (n=24), claustral neurons projecting to ipsilateral cortex only; CLA_bi (n=10), claustral neurons projecting bilaterally; EPd neurons (n=3) form a distinct morphological group, with specific projections to olfactory areas and limbic cortical areas. **b-c**, Integrated co-clustering dendrogram (b) and matrix (c) obtained by averaging the co-clustering matrices of four feature sets: projection pattern, soma location, axon morphology and dendrite morphology. Threshold for cluster calls is shown as the dashed line. Each cluster is annotated by the brain regions where somata (black) and axon clusters (red) reside. Regions were selected to represent >50% of cluster members. Side bars indicate manually assigned types with color codes shown below the matrix. **d,** Projection matrix heatmap for representative target brain regions of each neuron. Columns represent single cells sorted by cluster assignments. Rows represent targets, and the number following each target name indicates the dominant cluster ID for the row. **e,** Total number of cortical targets innervated by each neuron grouped by clusters. Ipsilateral and contralateral targets are counted separately. A minimum of one axon terminal (top panel) or 1,000 µm of axon length (bottom panel) is used as the threshold to label a region as "targeted". **f,** Top-down views of neurons in each cluster. Neurons are all flipped to the left hemisphere for comparison of axon projection patterns. Stars indicate soma locations and are flipped to the right hemisphere for visualization purpose.

1506 **Supplemental Information**

1507

1508 **Supplementary Table 1.** Transgenic mice used for the generation of fMOST imaging datasets,
1509 including main metadata information and tamoxifen dosing (see Methods) for sparse labeling.

1510 **Supplementary Table 2.** List of reconstructed neurons, with each neuron's 3D coordinates,
1511 annotated soma location in CCFv3 after registration and manual correction, transgenic line and
1512 brain ID, neuron subclass or type assignment, and projection matrix.

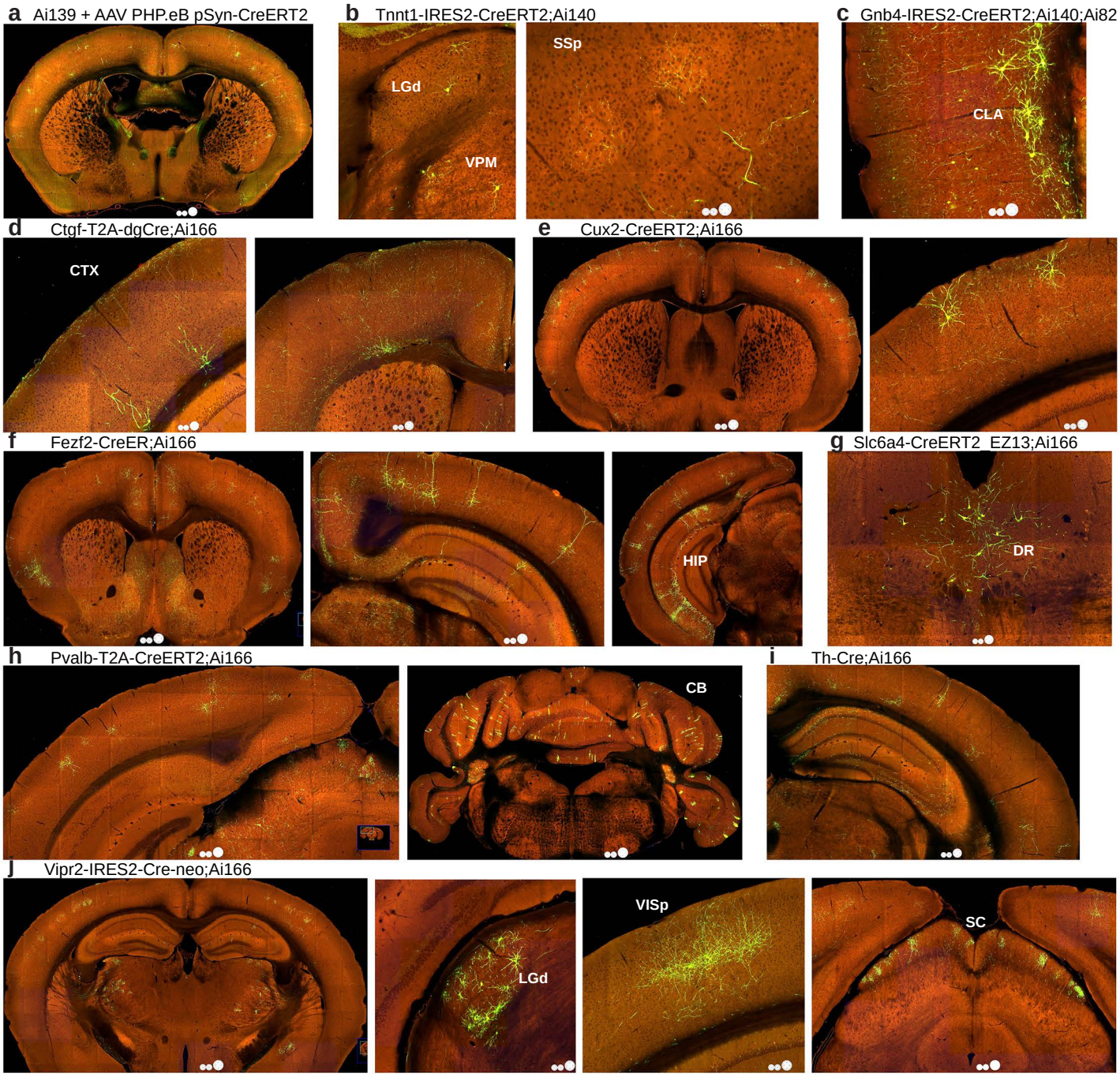
1513 **Supplementary Table 3.** Mesoscale anterograde tracing experiments used in this study for
1514 comparison with single neuron projection patterns, including main metadata information and
1515 projection matrix.

1516 **Supplementary Table 4.** Retro-seq cells for scRNA-seq analysis, with relevant metadata
1517 including retrograde labeling information.

1518

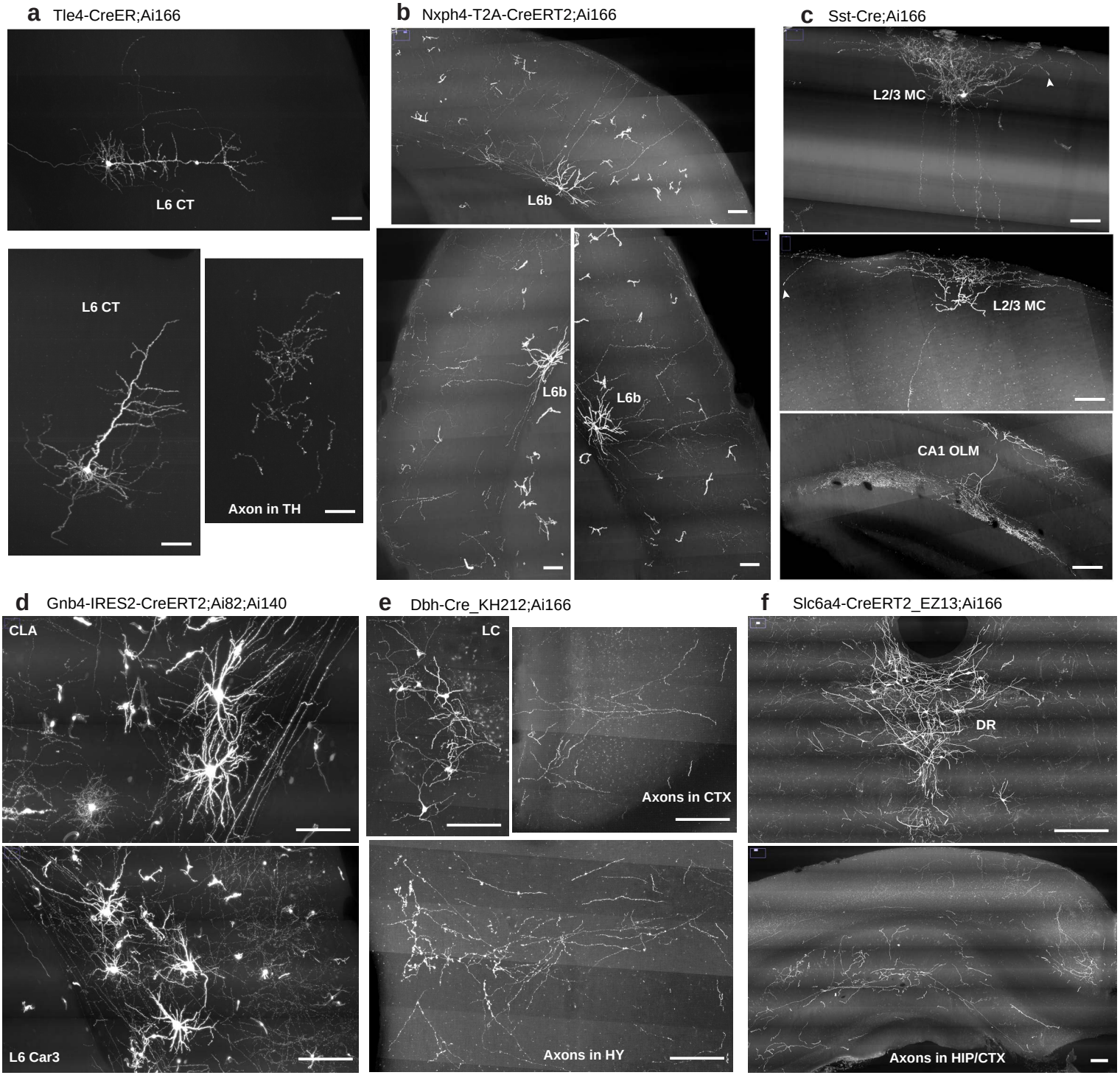
1519

Extended Data Figure 1



Extended Data Figure 1. Representative TissueCyte images with sparse and strong labeling of various neuron types using these approaches. **a**, A Synapsin I promoter-driven CreERT2-expressing AAV serotyped with PHP.eB was delivered at a dilution of 1:1000 by retroorbital injection into an Ai139 mouse, followed by a 1-day tamoxifen induction one week post injection, resulting in random sparse labeling of neurons throughout the brain. **b**, In a Tnnt1-IRES2-CreERT2;Ai140 brain, low-dose tamoxifen induction results in sparse labeling of thalamic projection neurons (left panel) with their axon terminal clusters in cortex clearly visible (right panel). **c**, In a Gnb4-IRES2-CreERT2;Ai140;Ai82 brain, low-dose tamoxifen induction results in sparse labeling of *Gnb4*+ claustral and cortical neurons with their widely dispersed axon fibers clearly visible. **d**, Cortical L6b neurons in a Ctgf-T2A-dgCre;Ai166 brain. **e**, Cortical L2/3/4 neurons in a Cux2-CreERT2;Ai166 brain. **f**, Cortical L5 ET neurons in a Fezf2-CreER;Ai166 brain. **g**, Serotonergic neurons in dorsal raphe (DR) in a Slc6a4-CreERT2_EZ13;Ai166 brain. **h**, Interneurons in cortex and cerebellum in a Pvalb-T2A-CreERT2;Ai166 brain. **i**, *Th*+ cortical interneurons in a Th-Cre;Ai166 brain. **j**, Projection neurons in LGd and other thalamic nuclei in a Vipr2-IRES2-Cre-neo;Ai166 brain. Third panel, axon projections from LGd neurons are seen in primary visual cortex (VISp). Fourth panel, axon projections likely from retinal ganglion cells are seen in superior colliculus (SC). Tamoxifen doses for CreERT2-containing mice are shown in **Supplementary Table 1**.

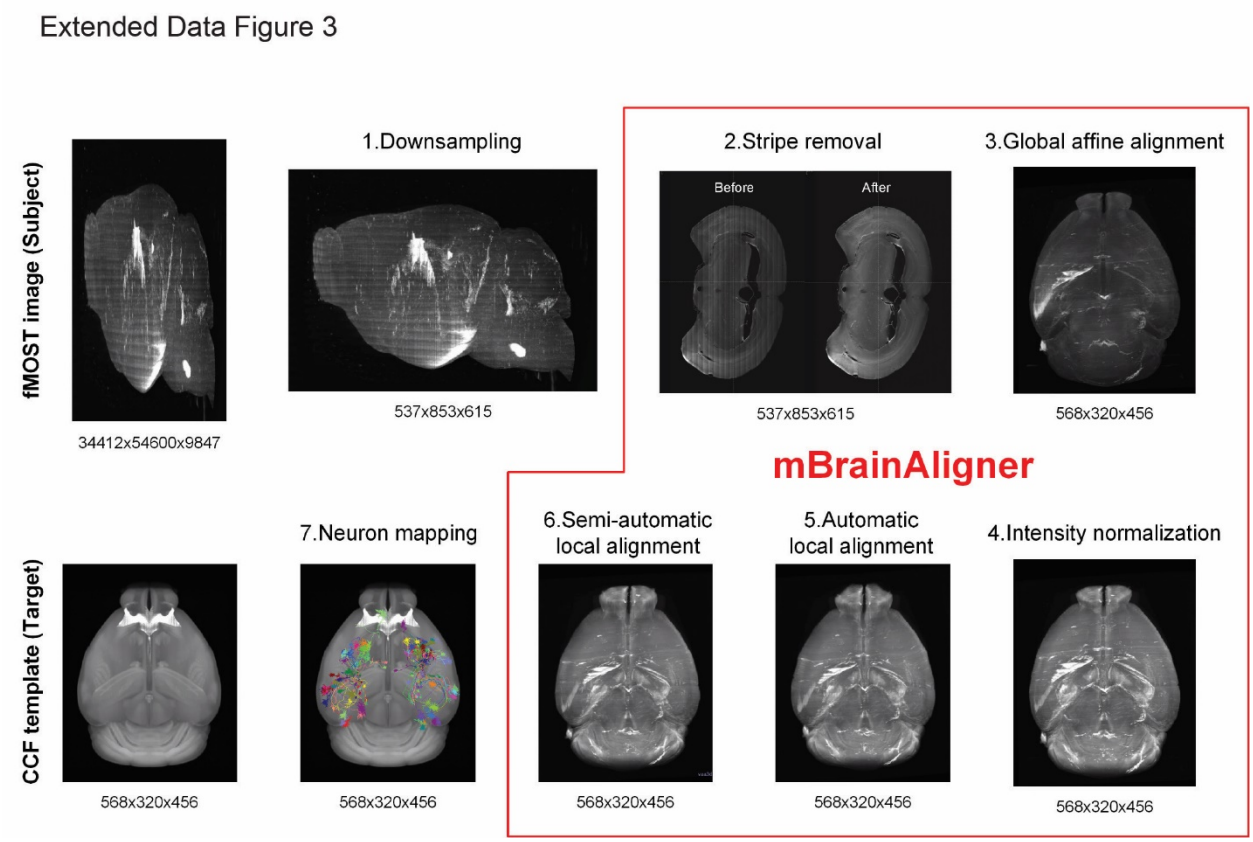
Extended Data Figure 2



1540 **Extended Data Figure 2.** Sparse, robust and consistent labeling and visualization of the
1541 dendritic and axonal arborizations of neuronal types in additional Cre lines. **a,** Cortical L6 CT
1542 neurons and their characteristic apical dendrites not reaching L1, as well as local axon collaterals
1543 and long-range axon projections into thalamus (TH), labeled in a Tle4-CreER;Ai166 brain. **b,**
1544 Cortical L6b neurons and their local axon projections up into L1 seen in a Nxph4-T2A-
1545 CreERT2;Ai166 brain. **c,** Cortical inhibitory Martinotti cells (MC) and hippocampal CA1 OLM
1546 cells labeled in a Sst-Cre;Ai166 brain. **d,** *Gnb4*+ claustral (CLA) and cortical (L6PC) neurons
1547 with their widely dispersed axon fibers seen in a Gnb4-IRES2-CreERT2;Ai140;Ai82 brain. **e,**
1548 Noradrenergic neurons labeled in the locus ceruleus (LC), and their long-range axon fibers seen
1549 in cortex (CTX) and hypothalamus (HY) in a Dbh-Cre_KH212;Ai166 brain. **f,** Serotonergic
1550 neurons labeled in the dorsal raphe (DR), and their long-range axon fibers seen in hippocampus
1551 (HIP) and cortex (CTX) in a Slc6a4-CreERT2_EZ13;Ai166 brain. Images shown are 100-µm
1552 maximum intensity projection (MIP) images (*i.e.*, projected from 100 consecutive 1-µm image
1553 planes). Arrowheads indicate observed terminal boutons at the end of the axon segments.
1554 Tamoxifen doses are shown in **Supplementary Table 1**. Scale bars, 100 µm.

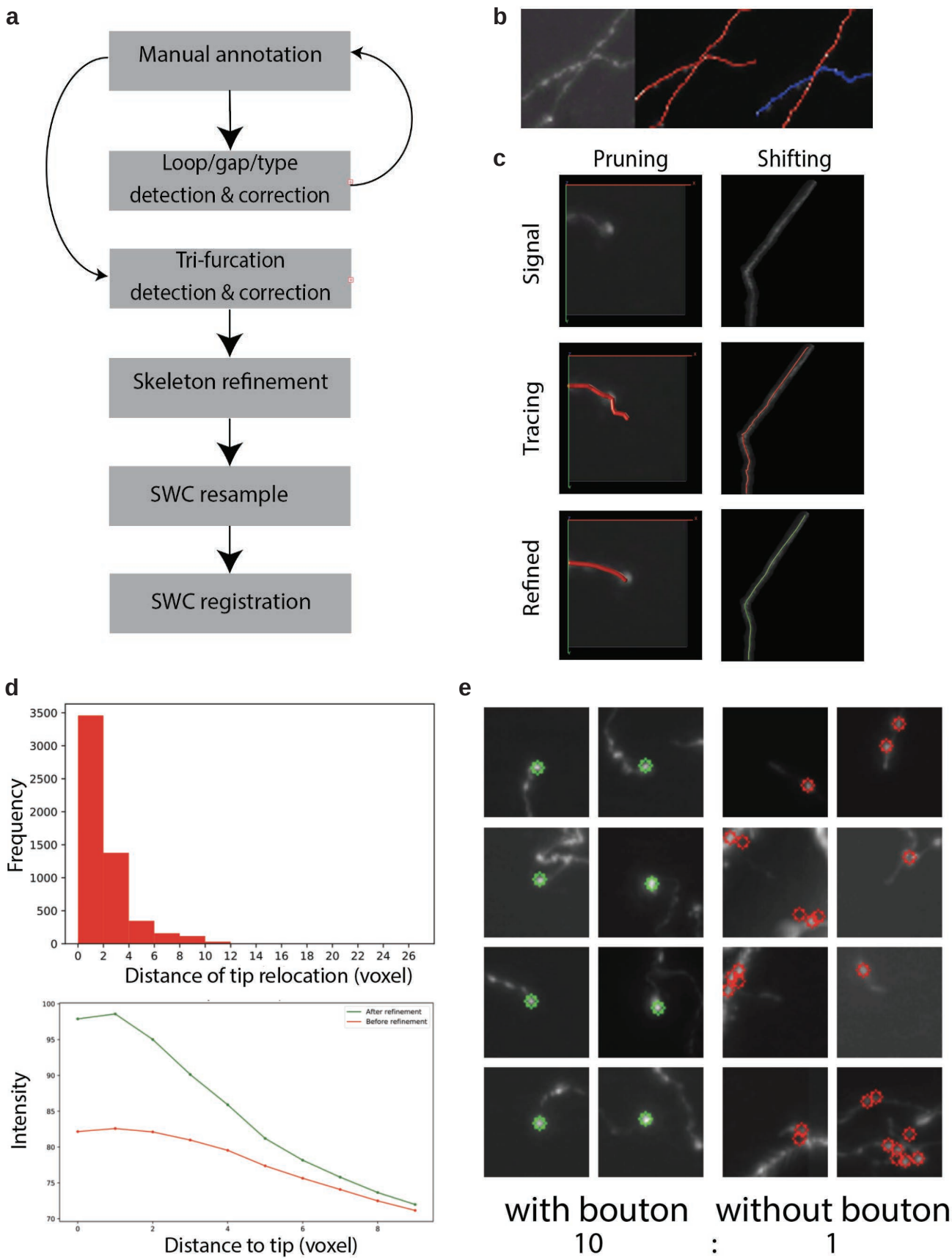
1555

1556



Extended Data Figure 3. CCF registration workflow. Pipeline of 3D registration from fMOST image (subject) to average mouse brain template of CCFv3 (target). Numbers below each panel indicate the pixel sizes in the order of X*Y*Z. See Methods for explanation of each step.

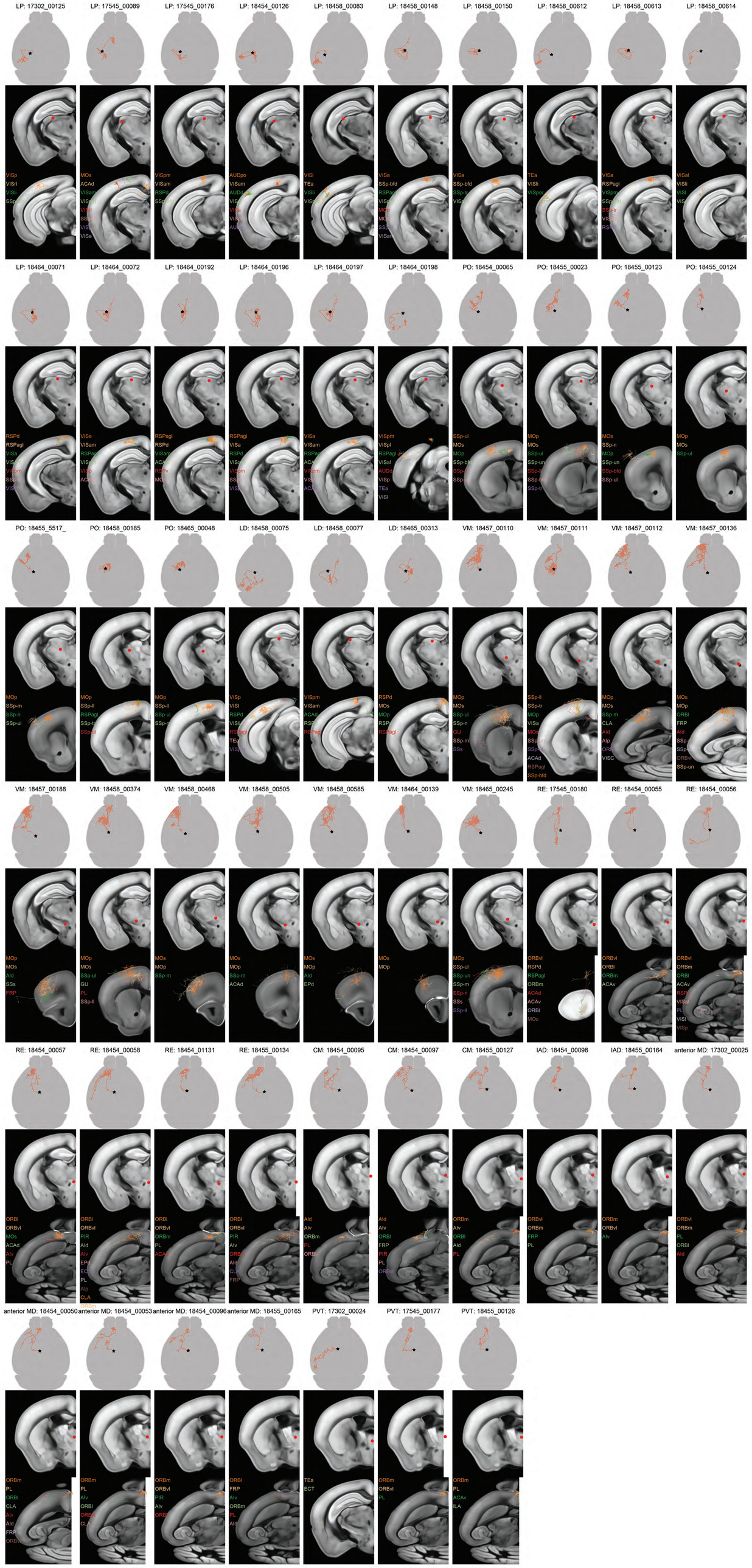
Extended Data Figure 4



Extended Data Figure 4. QC of reconstructed morphologies. **a,** Workflow of SWC post-processing process for QC: 1. automatic detection and correction of basic reconstruction errors including loops, gaps and incorrect node types. 2. Corrections are sent back to manual verification. 3. Automatic detection and correction of trifurcation, which are usually overlapping neurites, instead of branching points. 4. Refinement of SWC files, including pruning of over-traced terminals and shifting skeleton to fit the center of image signals. 5. Resampling of SWC to achieve evenly distributed nodes. 6. SWC registration to the standard CCFv3 mouse brain template. **b,** Examples of trifurcation before (middle) and after (right) correction. Blue and red branches do not cross. **c,** Examples of refinement before and after pruning (left) and shifting (right). **d,** Refinement leads to more precisely defined axon termination. Upper, distribution of terminal relocation distance by pruning. Lower, radius-decay curve of terminal signals shows that after refinement the axon ends at a brighter spot (indicating a bouton) rather than tapering off. **e,** Examples of axonal terminals that end with or without a bouton.

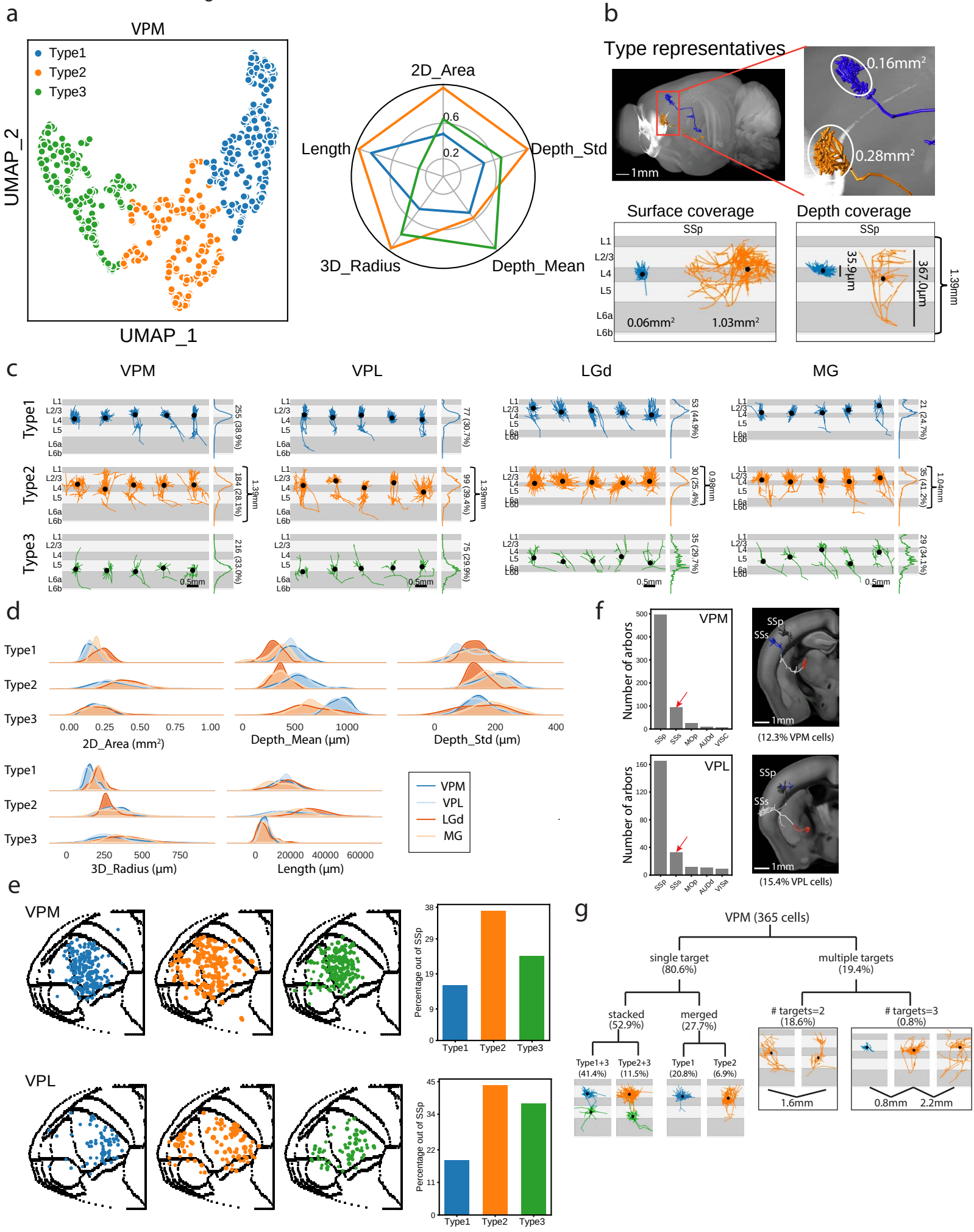
VPM: 18457_00064
VPM: 17302_00112
VPM: 18454_01055
VPM: 18457_00056
VPM: 18465_00243
VPM: 18462_00015
VPM: 18465_00230
VPM: 18454_01014
VPM: 18455_00015
VPM: 18455_00199
VPM: 18464_00074
VPM: 18454_00127
VPM: 18455_00159
VPM: 18455_00154
VPL: 18455_00120
VPL: 18454_01033
VPL: 18457_00202
VPL: 18463_00034
VPL: 18454_00109
VPL: 18463_00073
VPL: 18455_00152
VPL: 18463_00056
VPL: 18465_00177
VPL: 18457_00099
VPL: 18465_00226
VPL: 18458_00456
VPL: 18464_00133
VPL: 18457_00205
VPMpc: 18454_00132
VPMpc: 18455_00158
VPLpc: 18454_00073
VPLpc: 18455_5730_
VPLpc: 18455_5898_
VPLpc: 18455_6001_
VPLpc: 18464_00132
LGd: 17302_00048
LGd: 17302_00049
LGd: 17302_00066
LGd: 17545_00173
LGd: 18458_00051
LGd: 18458_00101
LGd: 18463_00001
LGd: 18458_00239
LGd: 18462_00054
LGd: 18465_00056
LGd: 18465_00070
LGd: 18465_00109
LGd: 18465_00110
LGd: 18465_00117
MG: 18454_00149
MG: 18462_00025
MG: 17302_00076
MG: 17302_00077
MG: 18454_00069
MG: 17545_00166
MG: 18454_00171
MG: 18465_00145
MG: 18454_00063
MG: 18455_00050
MG: 18458_00450
MG: 18463_00061
MG: 18464_00080
MG: 18465_00302
VAL: 17545_00071
VAL: 18454_00051
VAL: 18454_00075
AM: 17302_00023
AM: 18454_00099
AM: 18455_00166
AM: 18455_4982_
AM: 18458_00375
AM: 18458_00376
SMT: 17545_00175
SMT: 18454_00054
SMT: 18454_00100
SMT: 18454_00101
SMT: 18463_00046
SMT: 18457_00135
SMT: 18458_00616
SMT: 18465_00258
SMT: 18454_00176
SMT: 18454_00177
SMT: 18455_00132
SMT: 18455_00133
SMT: 18463_00047
posterior MD: 18454_00070
posterior MD: 18454_00130
posterior MD: 18455_00163
posterior MD: 18457_00209

Extended Data Figure 5. Tri-views of reconstructed neurons from "core" thalamic nuclei, visualized within the CCFv3 3D reference space. Each tri-view contains three views of the same neuron ordered from top to bottom: a whole-brain top-down view (soma indicated by a star, axon in red), a coronal plane showing the location of the soma (red dot), and a chosen coronal or horizontal plane close to the center of the main axon arbor with superimposed maximum projection view of the axon arbors. Cortical target regions with axon length >1 mm are indicated by different colors while other axon branches are shown in white.

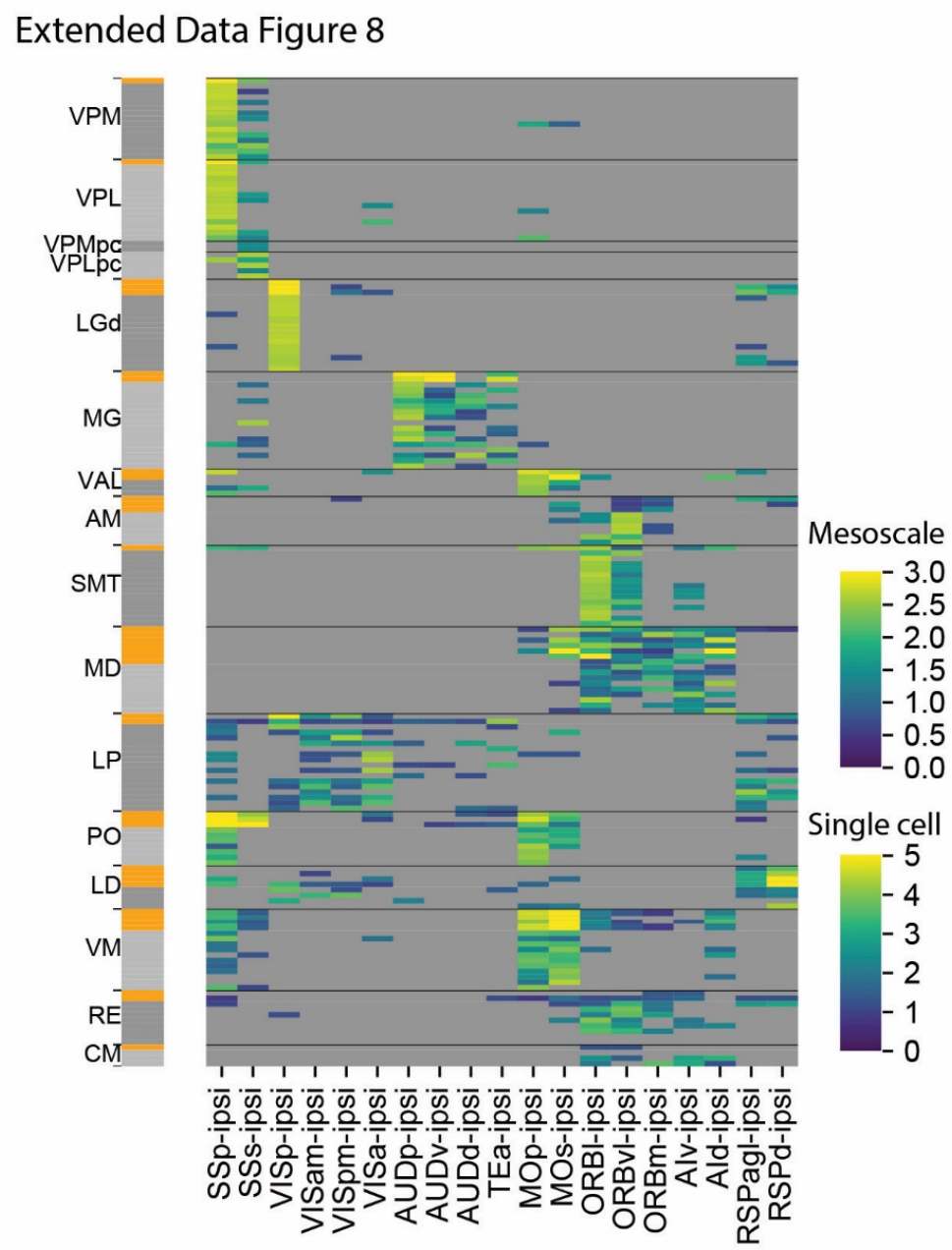

LP: 17302_00125
LP: 17545_00089
LP: 17545_00176
LP: 18454_00126
LP: 18458_00083
LP: 18458_00148
LP: 18458_00150
LP: 18458_00612
LP: 18458_00613
LP: 18458_00614
LP: 18464_00071
LP: 18464_00072
LP: 18464_00192
LP: 18464_00196
LP: 18464_00197
LP: 18464_00198
PO: 18454_00065
PO: 18455_00023
PO: 18455_00123
PO: 18455_00124
PO: 18455_5517_
PO: 18458_00185
PO: 18465_00048
LD: 18458_00075
LD: 18458_00077
LD: 18465_00313
VM: 18457_00110
VM: 18457_00111
VM: 18457_00112
VM: 18457_00136
VM: 18457_00188
VM: 18458_00374
VM: 18458_00468
VM: 18458_00505
VM: 18458_00585
VM: 18464_00139
VM: 18465_00245
RE: 17545_00180
RE: 18454_00055
RE: 18454_00056
RE: 18454_00057
RE: 18454_00058
RE: 18454_01131
RE: 18455_00134
CM: 18454_00095
CM: 18454_00097
CM: 18455_00127
IAD: 18454_00098
IAD: 18455_00164
anterior MD: 17302_00025
anterior MD: 18454_00050
anterior MD: 18454_00053
anterior MD: 18454_00096
anterior MD: 18455_00165
PVT: 17302_00024
PVT: 17545_00177
PVT: 18455_00126

Extended Data Figure 6. Tri-views of reconstructed neurons from "matrix" thalamic nuclei, visualized within the CCFv3 3D reference space. Each tri-view contains three views of the same neuron ordered from top to bottom: a whole-brain top-down view (soma indicated by a star, axon in red), a coronal plane showing the location of the soma (red dot), and a chosen coronal plane close to the center of the main axon arbor with superimposed maximum projection view of the axon arbors. Cortical target regions with axon length >1 mm are indicated by different colors while other axon branches are shown in white.

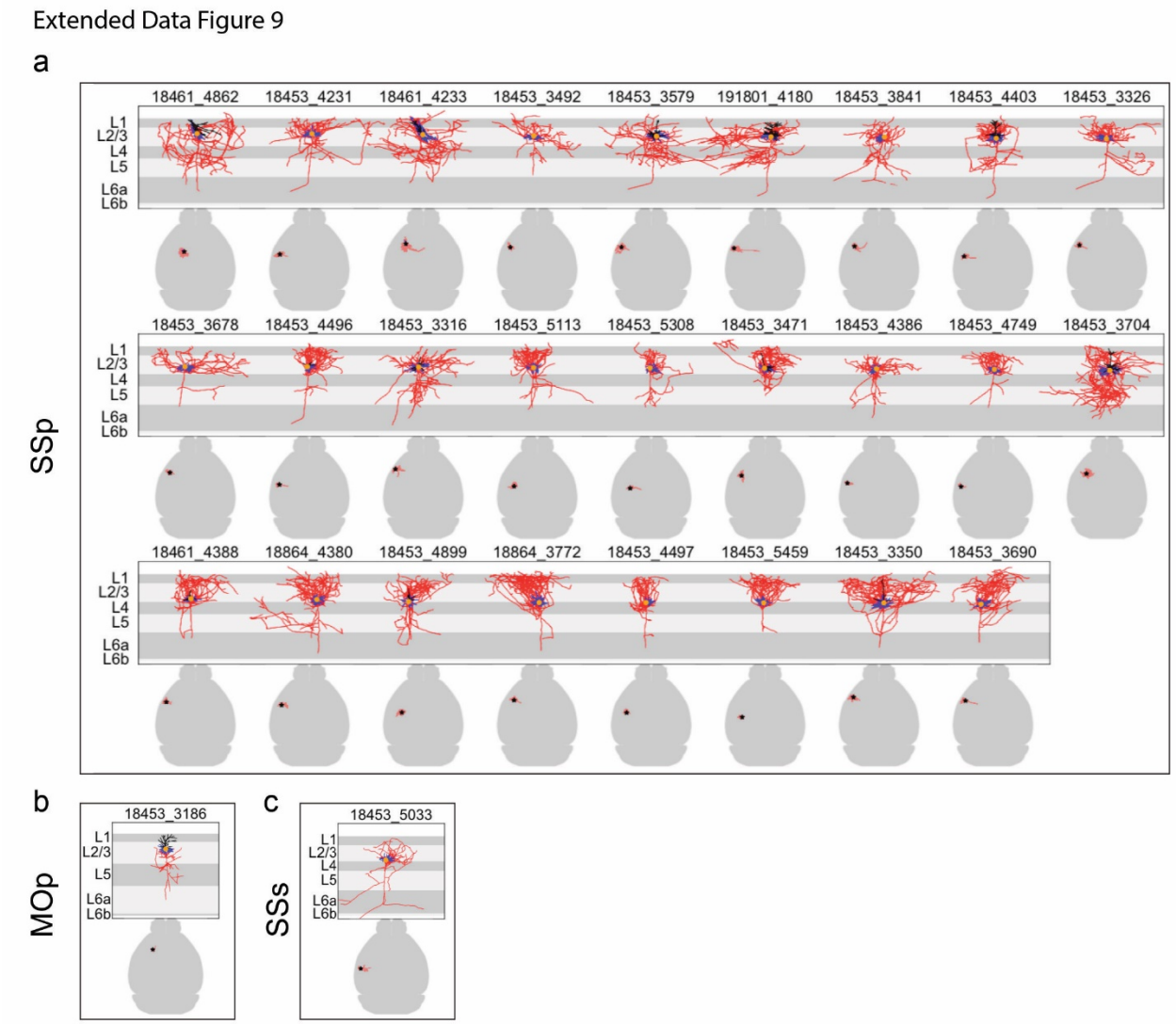
Extended Data Figure 7



Extended Data Figure 7. Thalamocortical axon arbor analysis. **a,** Clustering result indicates three types of cortical axon arbors in VPM neurons. Left, UMAP representation of VPM axon arbors, colored by cluster ID's. Right, Polar plot of main features, values as normalized cluster averages. **b,** Representative (upper) and extreme (lower) examples of VPM cortical arbors. **c,** Examples grouped by thalamic nuclei and arbor types. In each sub-panel, vertical views are shown for 5 representative arbors, with branch length distribution for all neurons of the same cluster on the right side. Arbor numbers and percentage of the group are shown on the right side. **d**, Distribution of features grouped by thalamic nuclei and arbor types. **e**, Arbor locations of VPM and VPL neurons in 2D cortical map grouped by arbor types. Each dot represents the center of an arbor. Right panels show percentage of arbors outside of the primary target of VPM/VPL neurons. **f**, (Left) Counts of VPM/VPL arbors in cortical regions. (Right) Examples of neurons with double arbors, one in SSp and the other in SSs. **g**, Variation of VPM neurons by arbor compositions. 'Single target' neurons are described as 'stacked' or 'merged' by bi-layer or single-layer distribution. The stacked and merged groups can be further separated by arbor types. The 'multiple targets' group is divided by number of targets.

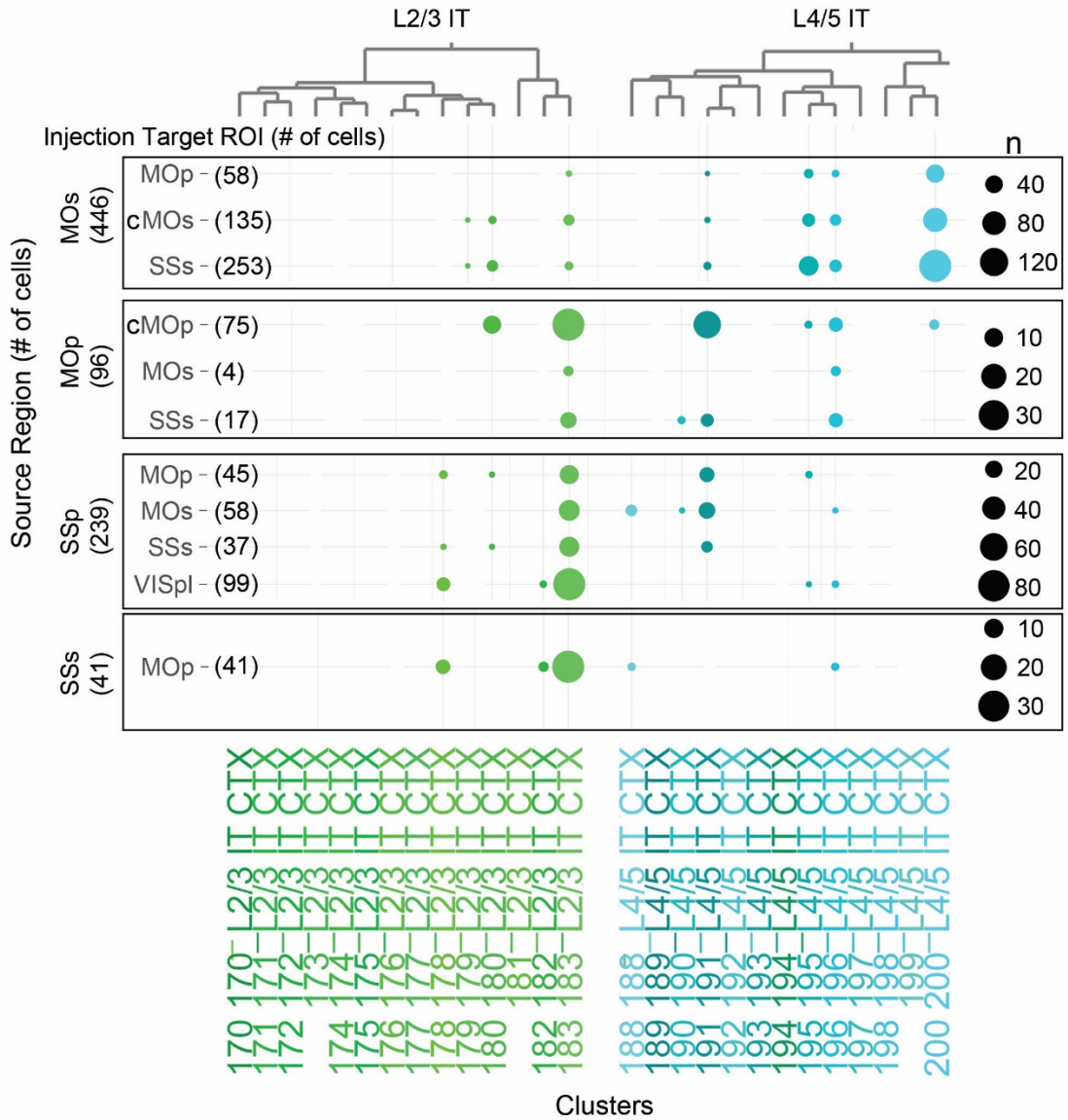


Extended Data Figure 8. Comparison of thalamocortical projection patterns between mesoscale experiments and single neurons as well as among individual neurons, for each listed thalamic nucleus. Heatmap colors represent log (percentage of projection strength + 1), scaled to 0-3 for mesoscale experiments and 0-5 for single cell reconstructions. We set regions below a cutoff 0.5 as grey. The side bar to the left of heatmap indicates experiment types (single cell: grey, mesoscale: orange).

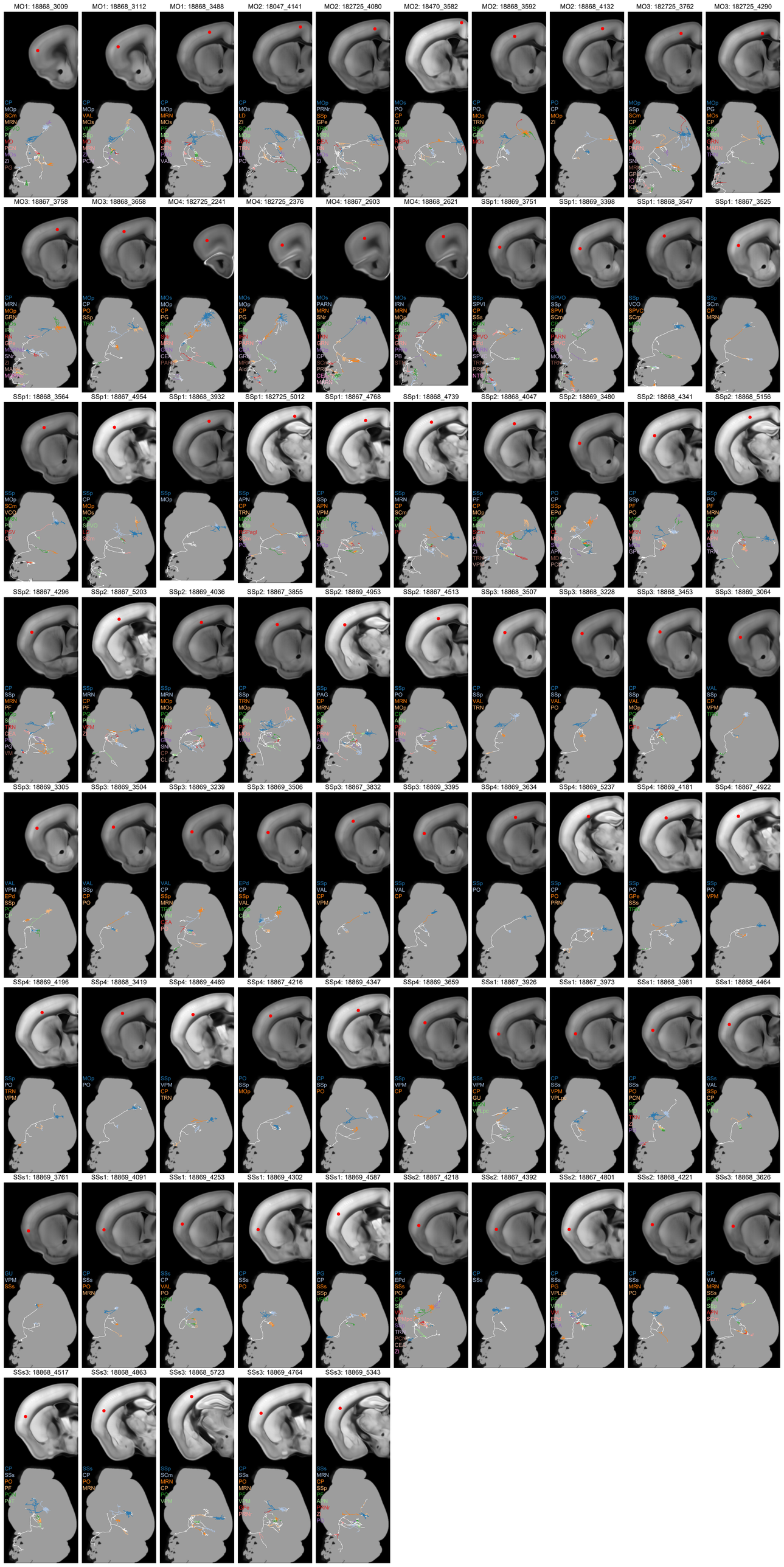


Extended Data Figure 9. Reconstructed cortical L2/3 and L4 IT neurons of SSp, SSs and MOp without substantial long-range axon projections. For each neuron, both local morphologies (upper panels; apical dendrite in black, basal dendrite in blue, axon in red, soma as an orange dot) and whole-brain projections (lower panels; axon in red, soma as a star) are shown.

Extended Data Figure 10

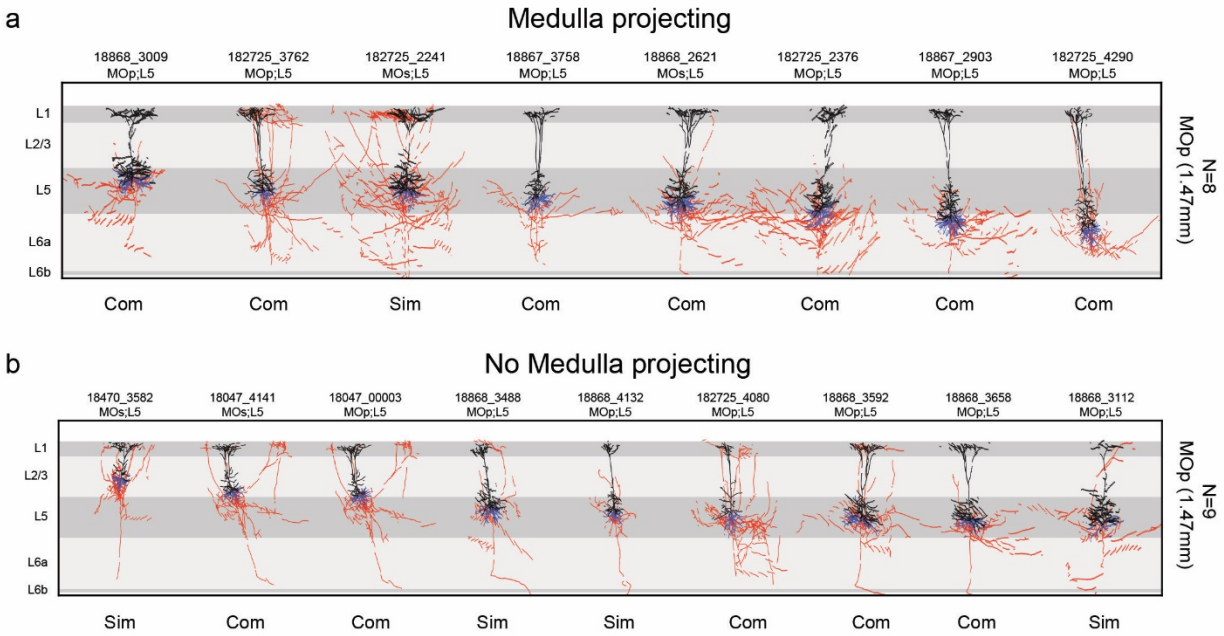


Extended Data Figure 10. Retro-seq characterization of cortical L2/3 and L4/5 IT neurons from MOs, MOp, SSp and SSs. Transcriptomes of retrogradely labeled neurons were obtained by single cell or nucleus RNA-sequencing and then mapped to our transcriptomic taxonomy [64] to identify the transcriptomic type of each neuron (shown as clusters at the bottom of the dot plot). Cells are grouped by their source region, and they mainly belong to a few transcriptomic types – L2/3 IT clusters 178, 180 and 183, and L4/5 IT clusters 191, 195, 196 and 200, with some interareal difference. Within each source region, cells labeled from different projection targets (injection target region of interest, ROI) are compared, and found to be assigned to a similar subset of transcriptomic types without major distinction. cMOs or cMOp denotes contralateral MOs or MOp, respectively.


MO1: 18868_3009
CP MOp SCm MRN SPVO PF MO PCN MV ZI PG
MO1: 18868_3112
CP MOp VAL MOs SCm SPVI MO MRN Z PCN
MO1: 18868_3488
CP MOp MRN MOs PF VPM GPe SSN PAG VAL
MO2: 18047_4141
CP MOs LD ZI SCm MRN AVN TRN LL PO
MO2: 182725_4080
MOp PRNr SSp GPe TRN MRN CEA RN MV ZI
MO2: 18470_3582
MOs PO CP ZI VPM MRN RSPd VPL
MO2: 18868_3592
CP PO MOp TRN SSp GU MOs
MO2: 18868_4132
PO CP MOp ZI
MO3: 182725_3762
MOp SSp SCm CP SPVI PF MOs PRNr Z SNr MRN GPe IO LC IOK
MO3: 182725_4290
MOp PG MOs CP SSp MRN GRN MRN TRN
MO3: 18867_3758
CP MRN MOp GRN IRN PF RN GPe MV SNr ZI MARN MRN
MO3: 18868_3658
MOp CP PO SSp TRN
MO4: 182725_2241
MOs MOp CP PG SCm SNr MRN CEA PARN
MO4: 182725_2376
MOs MOp CP PG PF SNr IPN PARN CP GRN MRN Ald
MO4: 18867_2903
MOs PARN MRN SNr SPVO TRN GRN MV CP SCm PRNr CEA MRN
MO4: 18868_2621
MOs IRN MRN MOp PARN SNr CP GRN PB STN
SSp1: 18869_3751
SSp SPVI CP SSs SCm SPVO EPd Z SPVC TRN PRNr NTS
SSp1: 18869_3398
SPVO SSp SPVI SCm CP GRN PARN SPVC Z MOp TRN
SSp1: 18868_3547
SSp VCO SPVC SCm MRN PF
SSp1: 18867_3525
SSp SCm CP MRN
SSp1: 18868_3564
SSp MOp SCm VCO MRN PF PSV CP
SSp1: 18867_4954
SSp CP MOp MOs SPVO LD SCm
SSp1: 18868_3932
SSp MOp
SSp1: 182725_5012
SSp APN CP TRN MRN MOp RSPagl SCm PCN
SSp1: 18867_4768
CP SSp APN VPM MRN PF PO ZI MV
SSp1: 18868_4739
SSp MRN CP SCm PF VPM PT
SSp2: 18868_4047
SSp PF CP MOp MRN SCm PO Z ZI TRN VPM
SSp2: 18869_3480
PO CP SSp EPd VPM CP MOp MRN Z APN MD PCN
SSp2: 18868_4341
CP SSp PF PO CP MOp PF MRN VPM GPe
SSp2: 18868_5156
SSp PO MRN CP PRNr VPM APN Z TRN
SSp2: 18867_4296
CP SSp MRN PF MOp SSp TRN CEA PO PG VM
SSp2: 18867_5203
SSp MRN CP PF PRNr VPM ZI
SSp2: 18869_4036
SSp MRN MOp MOs PF TRN AVN PF PAG SNr CP CL
SSp2: 18867_3855
CP SSp TRN MOp PF MOs VPM
SSp2: 18869_4953
SSp PAG CP MRN PF SSs PF PRNr AVN ZI
SSp2: 18867_4513
SSp PO MRN MOp PF APN PF TRN CL
SSp3: 18868_3507
CP SSp VAL TRN
SSp3: 18868_3228
CP SSp VAL PO
SSp3: 18868_3453
CP SSp VAL MOp PF SPe
SSp3: 18869_3064
VAL SSp CP VPM TRN
SSp3: 18869_3305
VAL VPM EPd SSp PF CP
SSp3: 18869_3504
VAL SSp CP PO
SSp3: 18869_3239
VAL CP SSp MRN TRN VPM CEA PF
SSp3: 18869_3506
EPd CP SSp VAL MOs CEA
SSp3: 18867_3832
SSp VAL CP VPM
SSp3: 18869_3395
SSp VAL CP
SSp4: 18869_3634
SSp PO
SSp4: 18869_5237
SSp CP PO PRNr
SSp4: 18869_4181
SSp PO GPe SSs TRN
SSp4: 18867_4922
SSp PO VPM
SSp4: 18869_4196
SSp PO TRN VPM
SSp4: 18868_3419
MOp PO
SSp4: 18869_4469
SSp VPM CP TRN
SSp4: 18867_4216
PO SSp MOp
SSp4: 18869_4347
CP SSp PO
SSp4: 18869_3659
SSp VPM CP
SSs1: 18867_3926
SSs VPM CP GU MRN VPLpc
SSs1: 18867_3973
CP SSs VPM VPLpc
SSs1: 18868_3981
CP SSs PO PCN PF MV MRN Z EPd
SSs1: 18868_4464
SSs VAL SSp CP PF VPM
SSs1: 18869_3761
GU VPM SSs
SSs1: 18869_4091
CP SSs PO MRN
SSs1: 18869_4253
SSs CP VAL PO VPM ZI
SSs1: 18869_4302
CP SSs PO
SSs1: 18869_4587
PG CP SSs SSp VPM
SSs2: 18867_4218
PF EPd SSs PO CP SPI VM VPM SSp SNr TRN PCN CEA ZI
SSs2: 18867_4392
CP SSs
SSs2: 18867_4801
CP SSs PG VPLpc PF VPM EPd CEA
SSs2: 18868_4221
CP SSs MRN PO
SSs3: 18868_3626
CP VAL MRN SSs PCN SPI APN SCm
SSs3: 18868_4517
CP SSs PO PF PCN PG
SSs3: 18868_4863
SSs CP PO MRN
SSs3: 18868_5723
SSp SCm MRN CP PF VPM
SSs3: 18869_4764
CP SSs PO MRN PF VPM SCm PRNr
SSs3: 18869_5343
SSs MRN CP SSp PF APN PRNr ZI PG

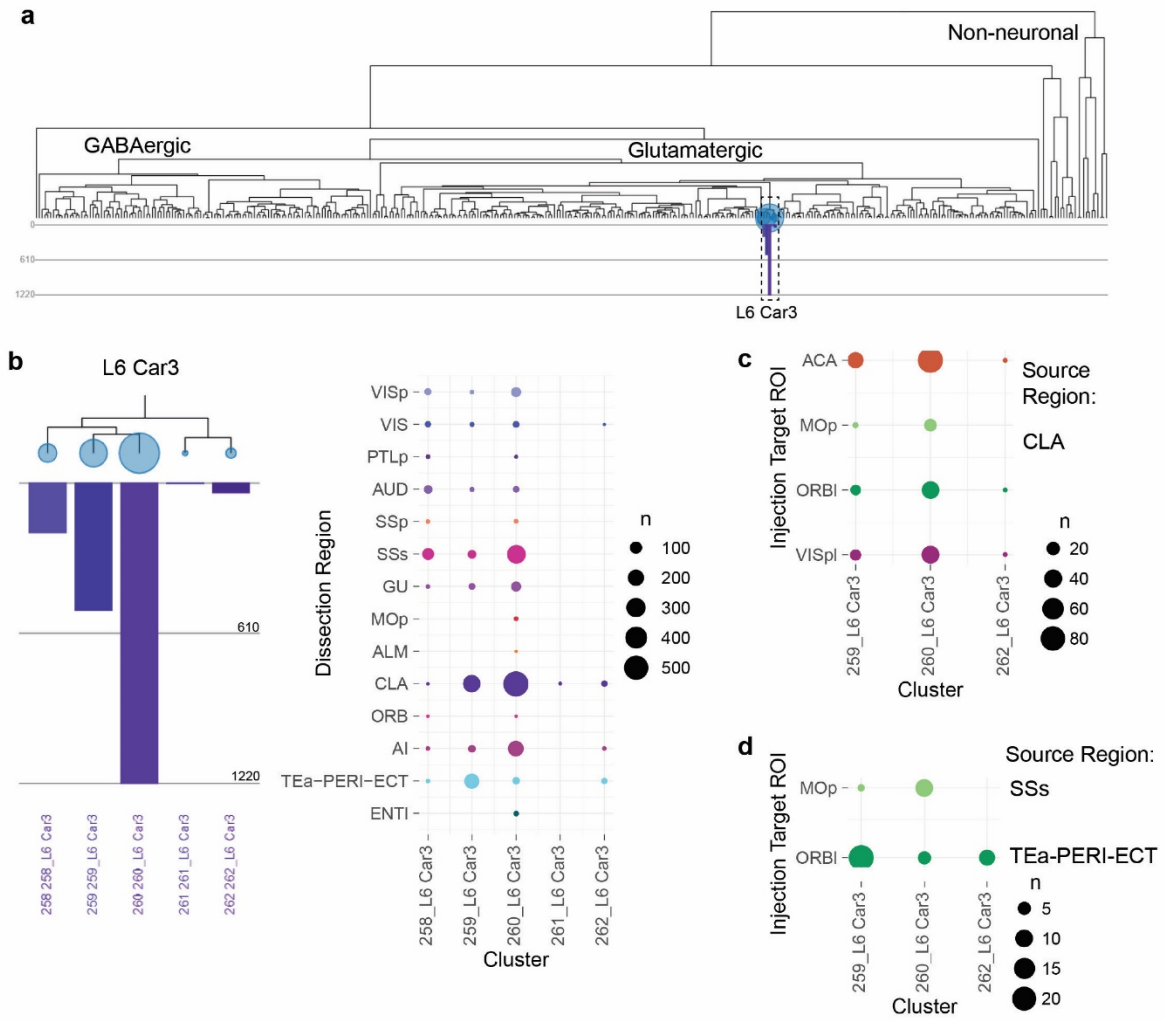
Extended Data Figure 11. Overview of reconstructed cortical L5 ET neurons, visualized within the CCFv3 3D reference space. Each neuron is shown in two views: top, a coronal plane showing the location of the soma (red dot); bottom, a sagittal maximum projection view showing the brain-wide axon projection pattern. Cortical target regions with axon length >1 mm are indicated by different colors while other axon branches are shown in white.

Extended Data Figure 12



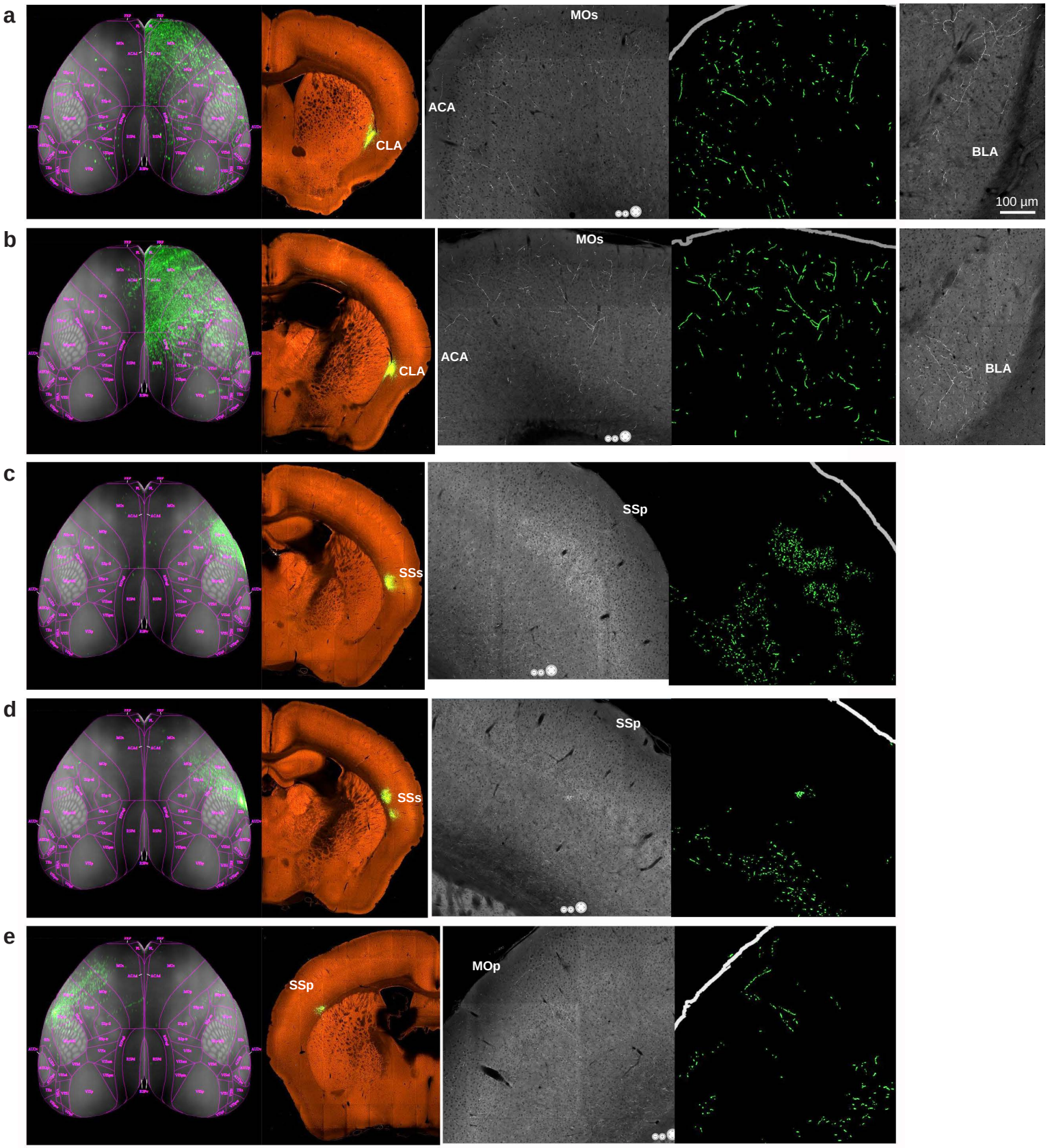
Extended Data Figure 12. Local morphologies of motor cortex (MOp and MOs) L5 ET neurons separated into medulla-projecting and non-medulla-projecting groups. Neurons with a single apical dendrite (not branching until reaching L1) are assigned as "simple" (Sim). Neurons with two or more apical dendrites (with early branching) are assigned as "complex" (Com). Apical dendrite in black, basal dendrite in blue, axon in red. Broken lines are due to the substantially tilted nature of these MOp and MOs neurons.

Extended Data Figure 13



Extended Data Figure 13. Single-cell RNA-seq characterization of L6 Car3 subclass of cortical and claustral neurons. **a,** Dendrogram of the co-clustering analysis of SMART-Seq v4 data from ~75,000 cortical, hippocampal and claustral cells reveals a distinct branch of L6 Car3 subclass (dashed box). **b**, The L6 Car3 subclass consists of 5 clusters with variable numbers of cells in each. Dot plot (right panel) shows the number of cells from each cortical region or claustrum contributing to each cluster. Clusters with large numbers of cells are contributed by cells coming from nearly all sampled regions. **c,** Retro-seq of claustral (CLA) cells shows that CLA cells projecting to different targets (ACA, MOp, ORBl or VISpl) are mapped to the same set of transcriptomic clusters without major distinction. **d,** Retro-seq of cortical cells shows that cells from SSs projecting to MOp and cells from TEa-PERI-ECT region projecting to ORBl are mapped to a similar set of transcriptomic clusters.

Extended Data Figure 14



Extended Data Figure 14. Anterograde bulk AAV tracing of projections from *Gnb4*+ neurons in claustrum or lateral cortex. **a-e,** AAV2/1-pCAG-FLEX-GFP tracer was injected into the claustrum (a-b), SSs (c-d) or SSp (e) in Gnb4-IRES2-Cre or Gnb4-IRES2-CreERT2 mice. Brains were imaged by the TissueCyte STPT system. First panel in each row: top-down view of segmented GFP-labeled axon projections in the cortex. Second panel: injection site. Third panel: the fine axon fibers in a target cortical area. Fourth panel: the segmented image of the third panel to visualize and quantify the axon fibers. Fifth panel in a and b: axon fibers observed in BLA. Full STPT image datasets are available at the Allen Mouse Brain Connectivity Atlas web portal (http://connectivity.brain-map.org/) with the following experiment IDs: a, 514505957; b, 485902743; c, 553446684; d, 581327676; e, 656688345. These 5 selected datasets all had small, spatially specific, injection sites that were located very close to each other. These small bulk injections demonstrate very distinct projection patterns between claustral and cortical *Gnb4*+ neurons.