1   **Title: A multimodal cell census and atlas of the mammalian primary motor cortex**
2
3   **Authors: BRAIN Initiative Cell Census Network (BICCN)**
4
5
6   **ABSTRACT**
7
8   We report the generation of a multimodal cell census and atlas of the mammalian primary motor
9   cortex (MOp or M1) as the initial product of the BRAIN Initiative Cell Census Network
10   (BICCN). This was achieved by coordinated large-scale analyses of single-cell transcriptomes,
11   chromatin accessibility, DNA methylomes, spatially resolved single-cell transcriptomes,
12   morphological and electrophysiological properties, and cellular resolution input-output mapping,
13   integrated through cross-modal computational analysis. Together, our results advance the
14   collective knowledge and understanding of brain cell type organization: First, our study reveals a
15   unified molecular genetic landscape of cortical cell types that congruently integrates their
16   transcriptome, open chromatin and DNA methylation maps. Second, cross-species analysis
17   achieves a unified taxonomy of transcriptomic types and their hierarchical organization that are
18   conserved from mouse to marmoset and human. Third, cross-modal analysis provides compelling
19   evidence for the epigenomic, transcriptomic, and gene regulatory basis of neuronal phenotypes
20   such as their physiological and anatomical properties, demonstrating the biological validity and
21   genomic underpinning of neuron types and subtypes. Fourth, *in situ* single-cell transcriptomics
22   provides a spatially-resolved cell type atlas of the motor cortex. Fifth, integrated transcriptomic,
23   epigenomic and anatomical analyses reveal the correspondence between neural circuits and
24   transcriptomic cell types. We further present an extensive genetic toolset for targeting and fate
25   mapping glutamatergic projection neuron types toward linking their developmental trajectory to
26   their circuit function. Together, our results establish a unified and mechanistic framework of
27   neuronal cell type organization that integrates multi-layered molecular genetic and spatial
28   information with multi-faceted phenotypic properties.
29
30
31   **INTRODUCTION**
32
33   Unique among body organs, the human brain is a vast network of information processing units,
34   comprising billions of neurons interconnected through trillions of synapses. Across the brain,
35   diverse neuronal and non-neuronal cells display a wide range of molecular, anatomical, and
36   physiological properties that together shape the network dynamics and computations underlying
37   mental activities and behavior. A remarkable feature of brain networks is their self-assembly
38   through the developmental process, which leverages genomic information shaped by evolution to
39   build a set of stereotyped network scaffolds largely identical among individuals of the same
40   species; life experiences then sculpt neural circuits customized to each individual. An essential

41   step toward understanding the architecture, development, function and neuropsychiatric diseases
42   of the brain is to discover and map its constituent neuronal elements together with the many
43   other cell types that comprise the full organ system.
44
45   The notion of "neuron types", cells with similar properties as the basic units of brain circuits, has
46   been an important concept since the discovery of stereotyped neuronal morphology over a
47   century ago [1,2]. However, a rigorous and quantitative definition of neuron types has remained
48   surprisingly elusive [3–7]. Neurons are remarkably complex and heterogeneous, both locally and in
49   their long-range axonal projections that can span the entire brain and connect to many target
50   regions. Many conventional technologies analyze one neuron at a time, and often study only one
51   or two cellular phenotypes in an incomplete way (*e.g.* missing axonal arbors in distant targets).
52   As a result, despite major advances in past decades, until recently phenotypic analyses of neuron
53   types remained severely limited in resolution, robustness, comprehensiveness, and throughput.
54   Besides technical challenges, complexities in the relationship among different cellular
55   phenotypes (multi-modal correspondence) have fueled long-standing debates on how neuron
56   types should be defined [8]. These debates reflect the lack of a biological framework of cell type
57   organization for understanding brain architecture and function.
58
59   In the past decade, single-cell genomics technologies have rapidly swept across many areas of
60   biology including neuroscience, promising to catalyze a transformation from phenotypic
61   description and classification to a mechanistic and explanatory molecular genetic framework for
62   the cellular basis of brain organization [9–12]. These technologies provide unprecedented resolution
63   and throughput to measure the molecular profiles of individual cells, including the complete sets
64   of actively transcribed genes (the transcriptome) and genome-wide epigenetic landscape (the
65   epigenome). Application of single cell RNA-sequencing (scRNA-seq) to the neocortex,
66   hippocampus, hypothalamus and other brain regions has revealed a complex but tractable
67   hierarchical organization of transcriptomic cell types that are consistent overall with knowledge
68   from decades of anatomical, physiological and developmental studies but with an unmatched
69   level of granularity [13–19]. Similarly, single-cell DNA methylation and chromatin accessibility
70   studies have begun to reveal cell type-specific genome-wide epigenetic landscapes and gene
71   regulatory networks in the brain [20–25]. Importantly, the scalability and high information content
72   of these methods allow comprehensive quantitative analysis and classification of cell types, both
73   neuronal and non-neuronal, revealing the molecular basis of cellular phenotypes and properties.
74   Further, these methods are readily applicable to brain tissues across species including humans,
75   providing a quantitative means for comparative analysis that has revealed compelling
76   conservation of cellular architecture as well as specialization of cell types across mammalian
77   species.
78
79   Other recent technological advances have crossed key thresholds to provide the resolution and
80   throughput to tackle brain complexity as well, for example for whole-brain neuronal morphology

A multimodal cell census and atlas of the mammalian primary motor cortex

81   and comprehensive projection mapping [26,27]. Furthermore, powerful new methods, including
82   imaging-based single-cell transcriptomics, the combination of single-cell transcriptome imaging
83   and functional imaging, and the integration of electrophysiological recording and single-cell
84   sequencing, allow mapping of the spatial organization, function, and electrophysiological,
85   morphological and connectional properties of molecularly defined cell types [28–32]. Finally, the
86   molecular classification of cell types allows the generation of models for genetic access to
87   specific cell types using transgenic mice and, more recently, short enhancer sequences [33–39]. All
88   of these methods have been applied to brain tissues in independent studies, but not yet in a
89   coordinated fashion to establish how different modalities correspond with one another, and how
90   explanatory a molecular genetic framework is for other functionally important cellular
91   phenotypes.
92
93   Recognizing the unprecedented opportunity to tackle brain complexity brought by these
94   technological advances, the overarching goal of the BRAIN Initiative Cell Census Network
95   (BICCN) is to generate an open-access reference brain cell atlas that integrates molecular,
96   spatial, morphological, connectional, and functional data for describing cell types in mouse,
97   human, and non-human primate brains [40]. A key concept is the Brain Cell Census, similar
98   conceptually to a population census, which accounts for the population of constituent neuronal
99   and non-neuronal cell types, along with their spatial locations and defining phenotypic
100   characteristics that can be aggregated as cellular populations that make up each brain region.
101   This cell type classification scheme, organized as a taxonomy, should aim for a consensus across
102   modalities and across mammalian species (for conserved types). Beyond the cell census, a Brain
103   Cell Atlas would be embedded in a 3D Common Coordinate Framework (CCF) of the brain [41], in
104   which the precise location and distribution of all cell types and their multi-modal features are
105   registered and displayed. Such a cell-type resolution spatial framework will greatly facilitate
106   integration, interpretation and navigation of various types of information for understanding brain
107   network organization and function.
108
109   Here we present the cell census and atlas of cell types in one region of the mammalian brain, the
110   primary motor cortex (MOp or M1) of mouse, marmoset and human, through an analysis with
111   unprecedented scope, depth and range of approaches (**Fig. 1, Table 1**). MOp is important in the
112   control of complex movement and is well conserved across species. Decades of accumulated
113   anatomical, physiological, and functional studies have provided a rich knowledge base for the
114   integration and interpretation of cell type information in MOp [42,43]. This manuscript describes a
115   synthesis of results and findings derived from eleven core companion papers through a multi-
116   laboratory coordinated data generation within BICCN. We derive a cross-species consensus
117   transcriptomic taxonomy of cell types and identify conserved and divergent gene expression and
118   epigenomic regulatory signatures from a large and comprehensive set of single-cell/nucleus
119   RNA-sequencing, DNA methylation and chromatin accessibility data. Focusing on mouse MOp,
120   we map the spatial organization of transcriptomic cell types by multiplexed error-robust

A multimodal cell census and atlas of the mammalian primary motor cortex

121 fluorescence in situ hybridization (MERFISH) and their laminar, morphological and
122 electrophysiological properties by Patch-seq; we report the cell-type resolution input-output
123 wiring diagram of this region by anterograde and retrograde tracing and investigate how axon
124 projection patterns of glutamatergic excitatory neurons correlate with molecularly-defined cell
125 types by Epi-Retro-Seq, Retro-MERFISH (the combination of MERFISH and retrograde
126 labeling), and single-neuron full morphology reconstruction; we describe transgenic driver lines
127 systematically targeting glutamatergic cell types based on marker genes and lineages. Finally, we
128 integrate this vastly diverse array of information into a cohesive depiction of cell types in the
129 MOp region with correlated molecular genetic, spatial, morphological, connectional, and
130 physiological properties and relating them to traditionally described cell types. Such integration
131 is illustrated in detail in example cell types with unique features in MOp: the layer 4
132 intratelencephalic-projecting (L4 IT) cells and layer 5 extratelencephalic-projecting (L5 ET)
133 cells. This multitude of datasets are organized by the BRAIN Cell Data Center (BCDC) and
134 made public through the BICCN web portal www.biccn.org. Key concepts and terms are
135 described in **Table 2**, including anatomical terms for input and output brain regions for MOp,
136 and hierarchical cell class, subclass and type definitions.

137

138 **Table 1. Experimental and computational techniques used in this study and associated**
139 **datasets**

Feature	Experimental or analytic technique(s)	Abbreviations	References	Samples (e.g. # of cells or nuclei) in MOp/M1	Total samples in flagship and companion papers
Transcription	Single-cell mRNA sequencing	scRNA-Seq: SMART-Seq v4, 10x Chromium v2, v3	Background: [15,44] Companion: [45]	**SMART-seq v4**: 6,288 cells (mouse) **10x Chromium v2, v3**: 193,824 cells (mouse)	1,163,727 cells
	Single nucleus mRNA sequencing	snRNA-Seq: SMART-Seq v4, 10x Chromium v2, v3	Background: [18,46,47] Companion: [45,48]	**SMART-seq v4**: 6,171 nuclei (mouse) 10,534 nuclei (human) **10x Chromium v2, v3**: 294,717 nuclei (mouse) 69,279 nuclei (marmoset) 15,842 nuclei (macaque) 76,533 nuclei (human)	1,100,168 nuclei

A multimodal cell census and atlas of the mammalian primary motor cortex

DNA methylation	Single-nucleus methylcytosine sequencing 2	snmC-Seq2	Background: [49] Companion: [45,48,50]	9,941 nuclei (mouse) 5,324 nuclei (marmoset) 5,222 nuclei (human)	110,294 nuclei
Open chromatin	Single nucleus Assay for Transposase-Accessible Chromatin	snATAC-Seq	Background: [21,51] Companion: [45,52]	79,625 nuclei (mouse)	813,799 nuclei
Combined transcription/ Open chromatin	Single-nucleus chromatin accessibility and mRNA expression sequencing	SNARE-seq2	Background: [53] Companion: [48]	9,946 nuclei (marmoset) 84,178 nuclei (human)	94,124 nuclei
Spatially resolved single-cell transcriptomics	Multiplexed error-robust fluorescence in situ hybridization	MERFISH	Background: [28,29] Companion: [54]	~300,000 cells (mouse)	~300,000 cells
Clustering and data integration methods	Clustering - Hierarchical iterative clustering	scrattch.hicat	Background: [15,44] Companion: [45,48]		
	Clustering - Metacell hierarchical clustering with dynamic tree pruning	tree-based method	Companion: [48]		
	Clustering of snATAC-seq data	SnapATAC	Background: [55] Companion: [52]		
	Clustering - Leiden clustering		Background: [56] Companion: [48]		
	Multimodality and cross-species integration	LIGER, Seurat, SingleCellFusion (SCF),scrattch. hicat	Background: [44,47,57–60] Companion:[45,48]		
Statistical validation	Cross-dataset replicability analysis	MetaNeighbor	Background: [61] Companion: [45,48]		
Electrophysiology, cellular morphology and	Combined in vitro slice physiology, biocytin cell filling, cytoplasm extraction	Patch-Seq, Smart-seq v2	Background: [30,62,63] Companion: [48,64,65]	1,237 cells (mouse) 6 cells (macaque) 6 cells (human)	133 cells (mouse) 6 cells (macaque) 391 cells (human)

A multimodal cell census and atlas of the mammalian primary motor cortex

transcriptomics	and RNA-sequencing				
Cellular morphology and projection	Whole brain single cell full morphology reconstructions Barcoded anatomy resolved by sequencing	fMOST, MouseLight BARseq	 Background: [26,66,67] Companion: [68,69]	151 cells (full morphology)	1,708 cells (full morphology) 10,299 neurons (BARseq)
Inter-areal circuit mapping	Anterograde tracing: PHAL, Viral tracers: AAV, Cre-dependent AAV, monosynaptic anterograde AAV-Cre	AAV, PHAL	Background: [70–74] Companion: [69,75]	22 experiments (mouse)	
	Retrograde tracing: CTB, viral tracers	RV, rabies, TRIO	Background: [70,76–78] Companion: [69,75]	40 experiments (mouse)	
Projection-specific profiling	Retrograde viral labeling of neurons with defined projections followed by epigenome profiling Combined retrograde labeling and MERFISH	Epi-Retro-Seq Retro-MERFISH	 Companion: [54,79]	2,111 cells (mouse)	11,827 cells
Genetic tools	Transgenic mouse lines	FlpO, Cre, CreER knockin lines; TIGRE-MORF/Ai166, MORF3 reporter line	Background: [80] Companion: [68,75] Stafford, Daigle, Chance et al., in preparation	6 knock-in driver lines 1 reporter line	26 knock-in lines

140

141 **Table 2. Glossary**

Glossary			Definition	InterLex Identifiers
Neuroanatomical regions described				
	MOp (mouse), M1 (human and	Primary motor cortex, the main target of cellular diversity analyses	Primary motor cortex as defined in Fig 1 of the paper.	ILX:0770115

A multimodal cell census and atlas of the mammalian primary motor cortex

	non-human primate)			
	L1, L2/3, L4, L5, L6, L6b	Layers within MOp	Cortical layers in primary motor cortex	ILX:0770170 ILX:0770171 ILX:0770172 ILX:0770173 ILX:0770179 ILX:0770180
Brain regions receiving axonal projections from MOp targeted for retrograde labeling studies				ILX:0770177
Cortical	SSp (Primary somatosensory cortex), MOs (Secondary motor cortex), TEa (Temporal association area), ACA (Anterior cingulate area)		Subset of cortical regions that receive axonal projections from primary motor cortex that were targeted in BICCN retrograde labeling studies.	ILX:0770178 ILX:0770117 ILX:0770116 ILX:0770118 ILX:0770120
Subcortical	STR (Striatum), TH (Thalamus), SC (Superior colliculus), VTA (Ventral tegmental area), HY (Hypothalamus), MB (Midbrain), P (Pons), MY (Medulla), claustrum		Subset of subcortical regions that receive axonal projections from primary motor cortex that were targeted in BICCN retrograde labeling studies.	ILX:0770167 ILX:0770122 ILX:0770123 ILX:0770124 ILX:0770137 ILX:0770165 ILX:0770126 ILX:0770127 ILX:0770125 ILX:0770128
Germinal sources of cortical GABAergic neurons				
	MGE, CGE	Medial and caudal ganglionic eminences	The MGE is a progenitor domain within the ventral telencephalon that, together with the lateral ganglionic eminence (LGE), are the source of the majority of interneurons in the neocortex, hippocampus and olfactory bulb. In addition, oligodendrocytes arise from these regions and migrate into the developing cortex. The CGE is a progenitor domain within the ventral telencephalon that is a source of cortical interneurons in the striatum, neocortex and limbic system. The CGE is defined as a posterior region in which the medial and lateral eminences are fused to one structure. This structure is also a source of oligodendrocytes. Adapted from https://discovery.lifemapsc.com.	ILX:0770129 ILX:0770130

A multimodal cell census and atlas of the mammalian primary motor cortex

Definition of terms used to describe cellular hierarchy				
Cell class: Top branches of hierarchical tree			The top branches of theCN transcriptomic cell type hierarchy comprising neuronal and non-neuronal cells. Neuronal cells comprise inhibitory GABAergic cortical neurons and excitatory glutamatergic cortical neurons. Non-neuronal cells comprise glia and non-neural cells.	ILX:0770094
Neuronal	Inh	GABAergic or inhibitory neurons, derived from MGE and CGE	Neuronal cell with a soma located in the cortex that uses GABA as a neurotransmitter and which exerts an inhibitory post-synaptic effect and derived from MGE and CGE.	ILX:0770098
	Exc	Glutamatergic or excitatory neurons	Neuronal cell with a soma located in the cortex that uses glutamate as a neurotransmitter and exerts an excitatory post-synaptic effect.	ILX:0770097
Non-neuronal	Glia	Non-neuronal cells of neuroectoderm origin	Non-neuronal brain cells of neuroectoderm origin	ILX:0770169
	Non-neural	Cells of mesoderm, neural crest or yolk sac origin	Non-neuronal brain cells of mesoderm, neural crest or yolk sac origin	ILX:0770187
Cell subclass: Subset of class, major groupings with highly convergent evidence across data modalities			Subset of class, major groupings with highly convergent evidence across data modalities	ILX:0770095
	GABAergic subclasses (Lamp5, Sncg, Vip, Sst, Sst Chodl, Pvalb, and Meis2)		Subclasses of GABAergic neurons distinguished by one or more marker genes	ILX:0770149 ILX:0770150 ILX:0770151 ILX:0770152 ILX:0770153 ILX:0770154 ILX:0770155
	Glutamatergic subclasses (L2/3 IT, L4/5 IT, L5 IT, L6 IT, L6 IT Car3, L5 ET, L5/6 NP, L6 CT, L6b)		Subclasses of glutamatergic neurons distinguished by anatomical location and projection pattern.	ILX:0770156 ILX:0770174 ILX:0770157 ILX:0770158 ILX:0770159 ILX:0770160 ILX:0770161 ILX:0770162 ILX:0770163
		IT: Intratelencephalic projecting	Excitatory glutamatergic neuron that projects to other telencephalic structures.	ILX:0770100
		ET: Extratelencephalic projecting	Excitatory glutamatergic neuron that projects to structures not derived from telencephalon	ILX:0770101

A multimodal cell census and atlas of the mammalian primary motor cortex

		NP: Near-projecting	Excitatory glutamatergic neuron that projects axons locally rather than long distance	ILX:0770103
		CT: Corticothalamic projecting	Excitatory glutamatergic neurons that project to the thalamus	ILX:0770102
	Glial subclasses (Astro, Oligo, OPC)		Subclasses of glial cells including astrocytes (Astro), oligodendrocytes (Oligo) and OPC cells (OPC)	ILX:0770141 ILX:0770140 ILX:0770139
	Non-neural subclasses (Endo, VLMC, SMC, Peri, Micro, PVM)		Subclasses of non-neural cells including endothelial cells (Endo), vascular leptomeningeal cells (VLMC), smooth muscle cells (SMC), pericytes (Peri), microglia (Micro) and perivascular myeloid cells (PVM)	ILX:0770142 ILX:0770143 ILX:0770144 ILX:0770145 ILX:0770146 ILX:0770147
Cell type: Subset of subclass, finest resolution clustering achieved for a modality or a consensus clustering across modalities and/or species			Subset of subclass, finest resolution clustering achieved for a modality or a consensus clustering across modalities and/or species	ILX:0770096
Cluster: Data-driven cell set, synonymous with type			Data-driven cell set, synonymous with type	ILX:0770164

142

143 Major findings from this coordinated consortium project include:

144 ● Combined single-cell transcriptomic and epigenomic analysis reveals a unified molecular
145 genetic landscape of adult cortical cell types that integrates gene expression, chromatin
146 state and DNA methylation maps.
147 ● Combination of single-cell -omics, MERFISH-based spatially resolved single-cell
148 transcriptomics and Patch-seq generates a census and atlas of cell types, including their
149 population demographics of type, proportion, and spatial distribution across cortical
150 layers and sublayers.
151 ● Comparative analysis of mouse, marmoset and human transcriptomic types achieves a
152 unified cross-species taxonomy of cortical cell types with their hierarchical organization
153 that reflects developmental origins; transcriptional similarity of cell type granularity
154 across species varies as a function of evolutionary distance.
155 ● We observed both highly conserved gene expression and epigenomic signatures of cell
156 identity across species, as well as a large set of species-specific cell type gene expression
157 profiles suggesting a high degree of evolutionary specialization.
158 ● The overall correspondence among transcriptomic, epigenetic, spatial transcriptomic,
159 morphological, and intrinsic physiological datasets reinforces the transcriptomic
160 classification of neuronal subclasses and distinctive types, demonstrating their biological
161 validity and genomic underpinnings, and also reveals continuously varying properties
162 along these axes among some neuronal subclasses and types.
163 ● Multi-faceted anatomic studies yield a cellular resolution wiring diagram of mouse MOp
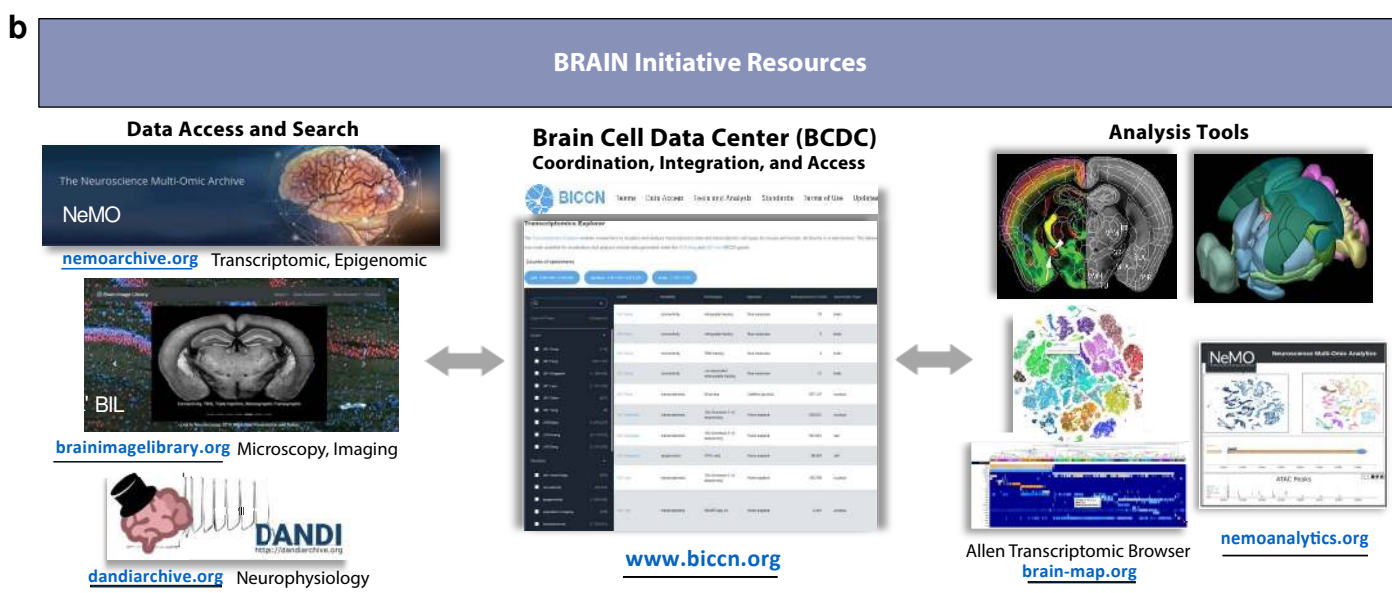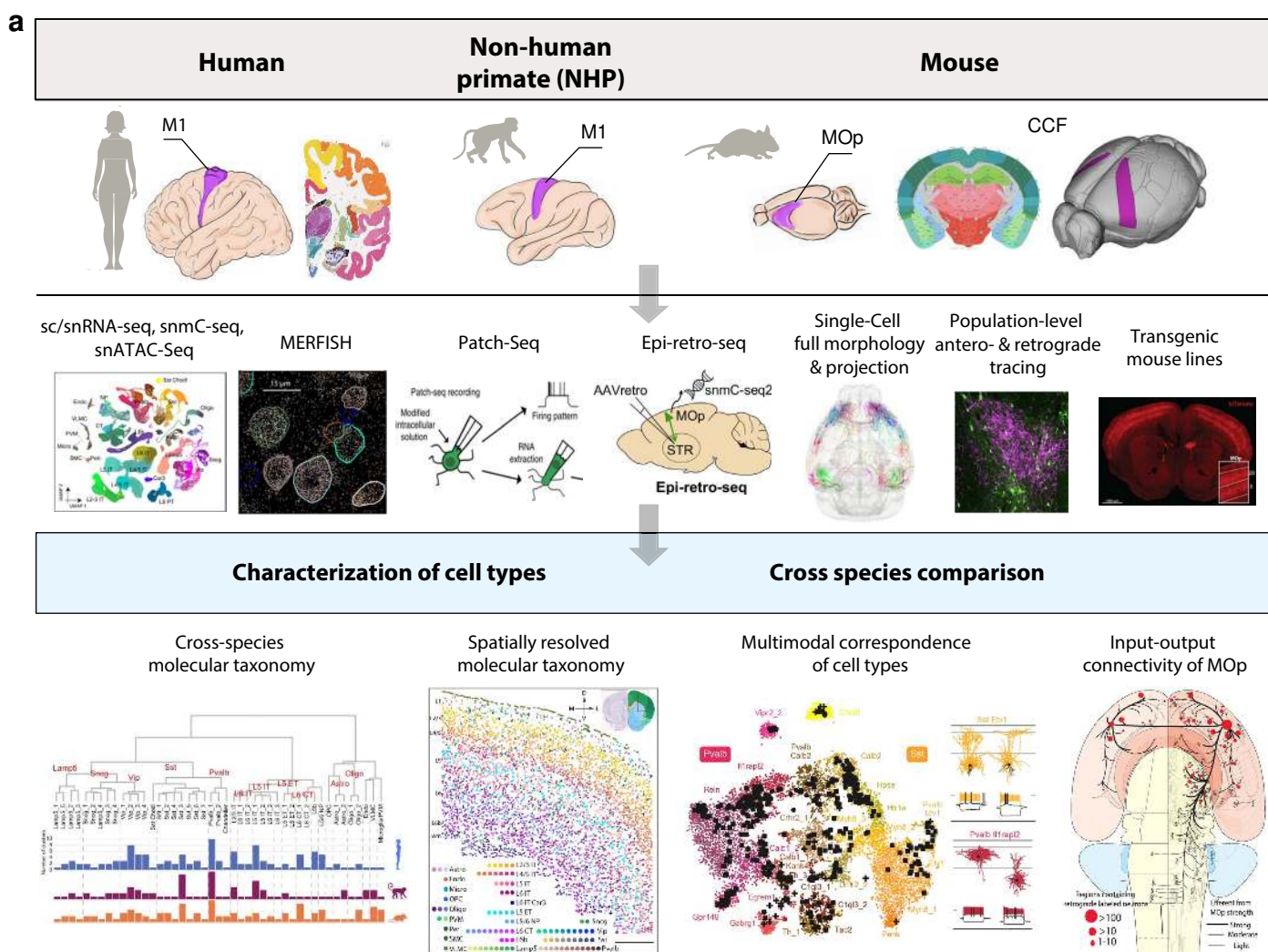164 anchored on major transcriptome-defined projection types, including input-output

A multimodal cell census and atlas of the mammalian primary motor cortex

165   connectivity at subpopulation level and output pathways at genetically-defined single-cell
166   level.
167 ● The long-range axon projection patterns of individual glutamatergic excitatory neurons
168   exhibit a complex and diverse range of relationships (between one-to-one and many-to-
169   many) with transcriptomic and epigenetic types, suggesting another level of regulation in
170   defining single-cell connectional specificity.
171 ● Cell type transcriptional and epigenetic signatures can guide the generation of an
172   extensive genetic toolkit for targeting glutamatergic pyramidal neuron types and fate
173   mapping their progenitor types.
174 ● Multi-site coordination within BICCN and data archives allows a high degree of
175   standardization, computational integration, and creation of open data resources for
176   community dissemination of data, tools and knowledge.
177
178
179 **Figure 1. Summary of experimental and computational approaches taken as well as**
180 **community resources generated by the BICCN. a**, Comprehensive characterization of cell
181 types in the primary motor cortex (MOp) of three mammalian species using multiple approaches
182 spanning molecular, genetic, physiological and anatomical domains. Integration of these datasets
183 leads to a cohesive multimodal description of cell types in the mouse MOp and a cross-species
184 molecular taxonomy of MOp cell types. **b**, The multimodal datasets are organized by the Brain
185 Cell Data Center (BCDC), archived in the Neuroscience Multi-omic (NeMO) Archive (for
186 molecular datasets), Brain Image Library (BIL, for imaging datasets) and Distributed Archive for
187 Neurophysiology Data Integration (DANDI, for electrophysiology data), and made publicly
188 available through the BICCN web portal www.biccn.org.
189
190

a

Human	Non-human primate (NHP)	Mouse

M1 ... M1 ... MOp ... CCF

sc/snRNA-seq, snmC-seq, snATAC-Seq — MERFISH — Patch-Seq — Epi-retro-seq — Single-Cell full morphology & projection — Population-level antero- & retrograde tracing — Transgenic mouse lines

AAVretro ... snmC-seq2 ... MOp ... STR ... Epi-retro-seq

Characterization of cell types   **Cross species comparison**

Cross-species molecular taxonomy — Spatially resolved molecular taxonomy — Multimodal correspondence of cell types — Input-output connectivity of MOp

b

BRAIN Initiative Resources

Data Access and Search

The Neuroscience Multi-Omic Archive
NeMO
nemoarchive.org Transcriptomic, Epigenomic

BIL
brainimagelibrary.org Microscopy, Imaging

DANDI
http://dandiarchive.org
dandiarchive.org Neurophysiology

Brain Cell Data Center (BCDC)
Coordination, Integration, and Access

BICCN

www.biccn.org

Analysis Tools

Allen Transcriptomic Browser
brain-map.org

NeMO
nemoanalytics.org

A multimodal cell census and atlas of the mammalian primary motor cortex

191 **RESULTS**
192

193 **Molecular definition of cell types in MOp**
194 The mouse MOp molecular taxonomy is derived from 9 datasets, including seven sc/snRNA-seq
195 datasets and one each of snmC-Seq2 and snATAC-Seq datasets (companion paper [45]). The
196 combined seven sc/snRNA-seq datasets (>700,000 cells total) had the advantages of large
197 number of cells profiled using the droplet-based 10x Chromium v2 or v3 method and deep full-
198 length sequencing using the plate-based SMART-Seq v4 method, resulting in a consensus
199 transcriptomic taxonomy for the mouse MOp with the greatest resolution compared to other data
200 types, containing 116 clusters or transcriptomic types (t-types), 90 of which were neuronal types
201 [45]. We used this mouse MOp transcriptomic taxonomy as the anchor for comparison and cross-
202 correlation of cell-type classification results across all data types. We further utilized two
203 computational approaches, SingleCellFusion (SCF) and LIGER, to combine the seven
204 transcriptomic with two epigenomic datasets and derive an integrated molecular taxonomy
205 consisting of 56 neuronal cell types (corresponding to the 90 transcriptomic neuronal types) for
206 the mouse MOp, with highly consistent molecular profiles based on transcriptomics, DNA-
207 methylation, and open chromatin [45] (**Fig. 2a**). Critically, this integrated taxonomy enabled us to
208 link RNA transcripts with epigenomic marks identifying potential cell-type-specific cis-
209 regulatory elements (CREs) and transcriptional regulatory networks. Similarly, we established
210 M1 cell type taxonomies for human (127 t-types) and marmoset (94 t-types) by unsupervised
211 clustering of snRNA-seq data, followed by integration with epigenomic datasets (companion
212 paper [48]).
213

214 To establish a consensus classification of MOp/M1 cell types among mouse, human and
215 marmoset, we integrated snRNA-seq datasets across species and identified 45 conserved
216 transcriptomic types that spanned three major cell classes, including 24 GABAergic, 13
217 glutamatergic, and 8 non-neuronal types (**Fig. 2b, Extended Data Fig. 1**). These types were
218 grouped into broader subclasses based on shared developmental origin for GABAergic inhibitory
219 neurons [i.e., three caudal ganglionic eminence (CGE)-derived subclasses (Lamp5, Sncg and
220 Vip) and two medial ganglionic eminence (MGE)-derived subclasses (Sst and Pvalb)], layer and
221 projection pattern in mouse for glutamatergic excitatory neurons [i.e., intratelencephalic (IT),
222 extratelencephalic (ET), corticothalamic (CT), near-projecting (NP) and layer 6b (L6b)], and
223 non-neuronal functional subclass (e.g., oligodendrocytes and astrocytes) (**Table 2**). Note that the
224 layer 5 extratelencephalic (L5 ET) neurons had been named as pyramidal tract (PT) neurons or
225 subcerebral projection neurons (SCPN) in the literature [81,82]; in this study we chose to use the
226 name L5 ET for this subclass of neurons to be more representative across cortical areas and
227 species (**Supplementary Notes**). The resolution of this cross-species conserved taxonomy was
228 lower than that derived from each species alone, due to gene expression variations among
229 species. The degree of species alignments varied across consensus types (**Fig. 2c**); some types
230 could be aligned one-to-one (e.g., Lamp5_1, L6 IT_3), while others aligned several-to-several

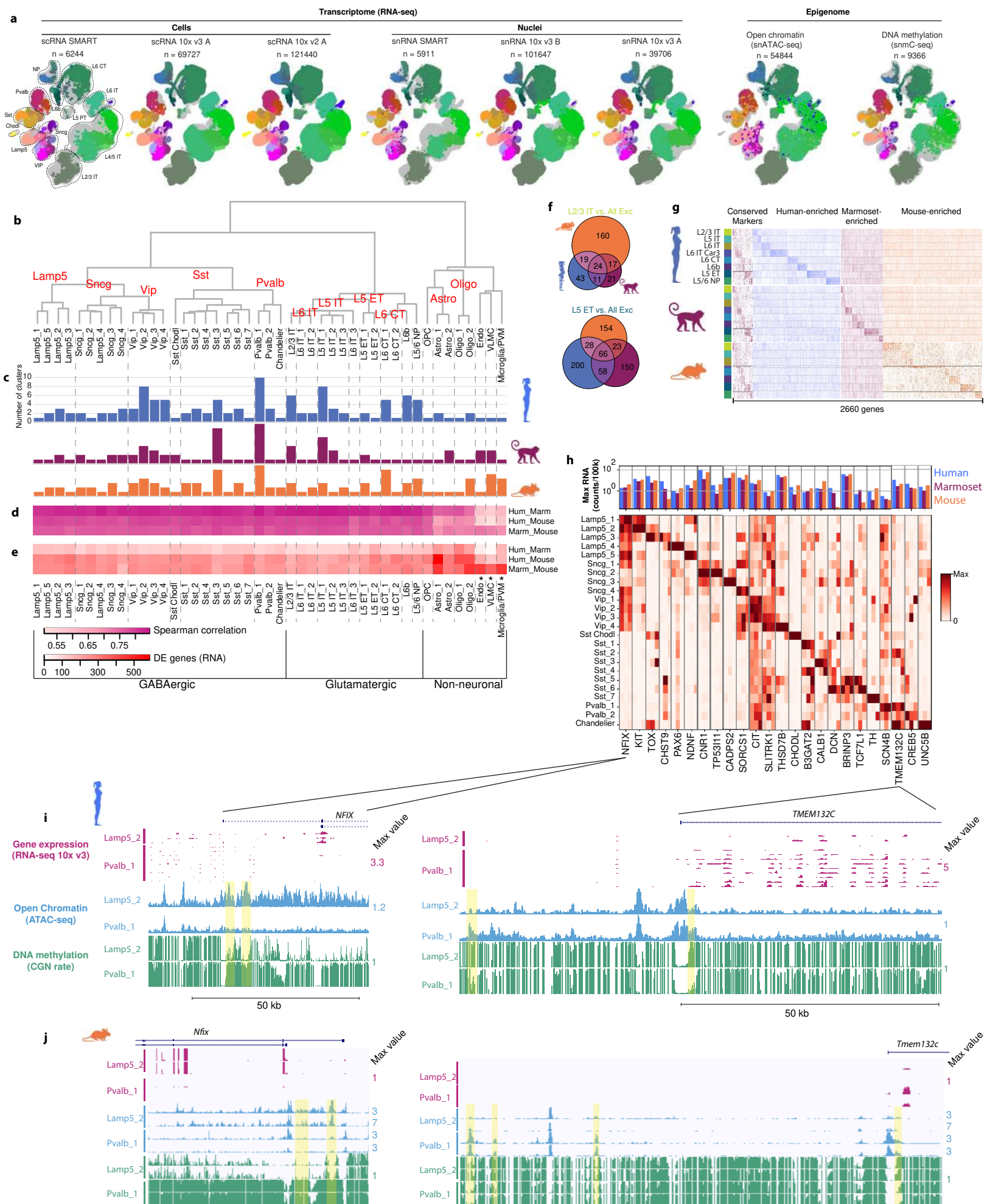A multimodal cell census and atlas of the mammalian primary motor cortex

231 (e.g., Pvalb_1, L2/3 IT, L5 IT_1). This may reflect over- or under-clustering, limitations in
232 aligning highly similar cell types or species-specific expansion of cell-type diversity (companion
233 paper [65]).
234
235 We hypothesized that cell types would share more similar gene expression profiles between
236 human and marmoset than between either primate and mouse because primates share a more
237 recent common ancestor. Indeed, we found that between primates, transcriptomic profiles of
238 consensus cell types were more correlated and had 25-50% fewer differentially expressed (DE)
239 genes than between primates and mouse (**Fig. 2d,e**). Three non-neuronal types had greater
240 spearman correlations of overall gene expression (**Fig. 2d**, right columns) between marmoset and
241 mouse likely because non-neuronal cells were undersampled in human M1 resulting in fewer
242 rare types [48]. Robust conservation of cell types across mammals, including types with known
243 specificity in electrical properties and connectivity such as chandelier cells and long-range
244 projecting *Sst*-expressing cells (*Sst Chodl*), is strong evidence for the functional significance of
245 these types.
246
247 Glutamatergic subclasses expressed many marker genes (using Seurat's FindAllMarkers function
248 with test.use set to 'roc', >0.7 classification power) compared to other subclasses, and the
249 majority of markers were species-specific (**Fig. 2f,g**). The evolutionary divergence of marker
250 gene expression may reflect species adaptations or relaxed constraints on genes that can be
251 substituted with others for related cellular functions. Subclasses also had a core set of marker
252 genes that were conserved across all three species (**Fig. 2g**); these genes are candidates for
253 consistent labeling of consensus cell types and for determining the conserved features of those
254 cells that are central to their function. GABAergic consensus types also had conserved markers
255 with similar absolute expression levels across species (**Fig. 2h**, bar plots) and relatively specific
256 expressions compared to other cell types (**Fig. 2h**, heatmap). Marker genes of Lamp5_2 (*NFIX*)
257 and Pvalb_1 (*TMEM132C*) GABAergic neurons showed evidence for cell-type-specific
258 enhancers located in regions of open chromatin and DNA hypomethylation in both human (**Fig.
259 2i**) and mouse (**Fig. 2j**).
260
261 In summary, the multi-omic approach reveals a unified molecular genetic landscape of cortical
262 cell types that integrates gene expression, chromatin state and DNA methylation maps and yields
263 a robust molecular classification of cell types that is consistent between transcriptomic and
264 epigenomic analyses. These studies further allow the identification of putative regulatory
265 elements associated with cell type identity. Cell types are generally conserved between primates
266 and rodents, and have a small number of conserved marker genes that are candidates for
267 consistent labeling of conserved cell types.
268
269

A multimodal cell census and atlas of the mammalian primary motor cortex

270 **Figure 2. MOp consensus cell type taxonomy. a**, Integrated transcriptomic and epigenomic
271 datasets using SCF show consistent molecular cell-type signatures as revealed by a low-
272 dimensional embedding in mouse MOp. Each Uniform Manifold Approximation and Projection
273 (UMAP) plot represents one dataset. Colors indicate different subclasses. **b,** Dendrogram of
274 integrated human, marmoset, and mouse cell types based on single nucleus RNA-seq datasets
275 (10x Chromium v3). **c,** Number of within-species clusters that are included in each cross-species
276 cluster. **d-e,** For each consensus cluster, correlations (d) and differentially expressed (DE;
277 Wilcoxon test, FDR < 0.01 and $\log_e$ fold-change > 2) genes (e) between pairs of species.
278 Asterisks denote non-neuronal populations that were under-sampled in human. **f,** Venn diagrams
279 of shared DE genes between species for L2/3 IT and L5 ET glutamatergic neuron subclasses. **g,**
280 Conserved and species-specific DE genes for all glutamatergic subclasses. Heatmap shows gene
281 expression normalized by the maximum for each gene for up to 50 randomly sampled nuclei
282 from each subclass and species. **h,** Conserved markers of GABAergic neuron types across three
283 species. **i-j,** Genome browser showing transcriptomic and epigenetic signatures for gene markers
284 of Lamp5_2 (*NFIX*) and Pvalb_1 (*TMEM132C*) GABAergic neurons in human (i) and mouse (j).
285 Yellow bars highlight sites of open chromatin and DNA hypomethylation in the cell type with
286 corresponding marker expression.
287
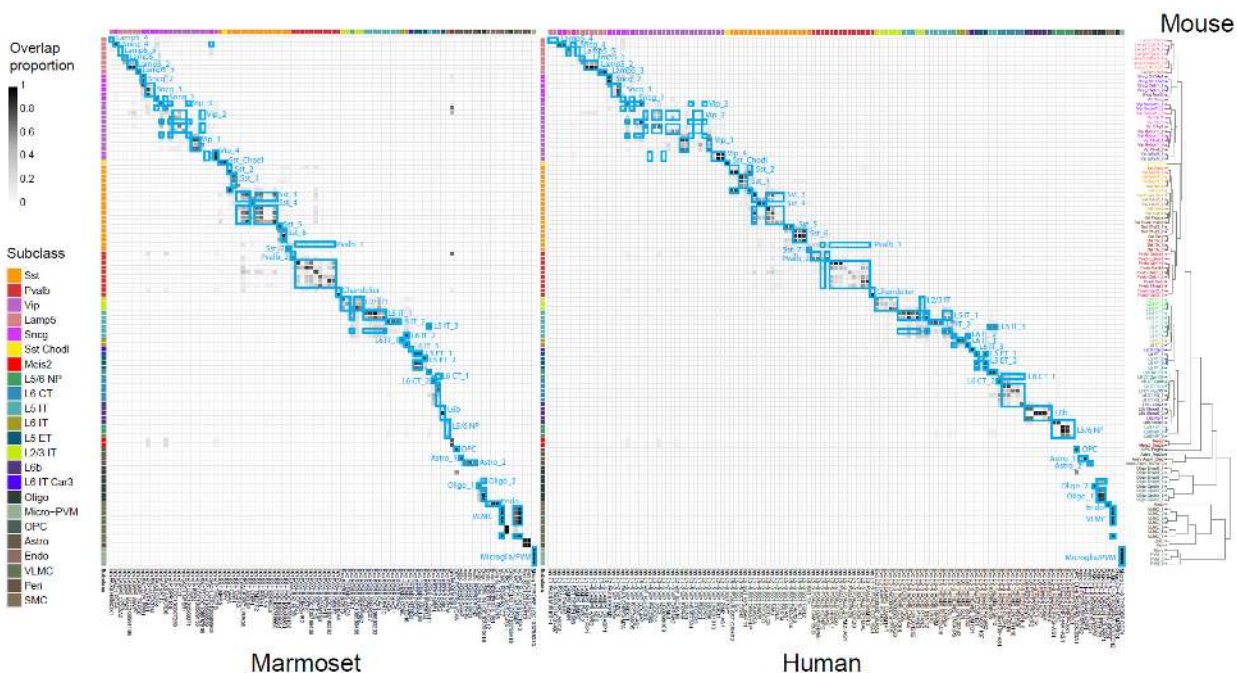
288
289
290 **Extended Data Figure 1.** Cluster overlap heatmap showing the proportion of nuclei in each pair
291 of species clusters that are mixed in the cross-species integrated space. Cross-species consensus
292 clusters are indicated by labeled blue boxes. Mouse clusters (rows) are ordered by the mouse
293 MOp transcriptomic taxonomy dendrogram reproduced from [45]. Marmoset (left columns) and
294 human (right columns) transcriptomic clusters (reproduced from [48] are ordered to align with

295   mouse clusters. Color bars at top and left indicate subclasses of within-species clusters.

296

297

298   **Spatially resolved cell atlas of the mouse MOp by MERFISH**

299   Sequencing-based single-cell methods require dissociation of cells from tissues, and hence the

300   spatial organization of neuronal and non-neuronal cells, which is critical for brain function, is

301   lost. To obtain a spatially resolved cell atlas of the mouse MOp region, we used MERFISH, a

302   single-cell transcriptome imaging method [28,29], to identify cell types *in situ* and map their spatial

303   organization. We selected a panel of 258 genes (254 of which passed quality control) to image

304   by MERFISH, on the basis of both prior knowledge of marker genes for major subclasses of

305   cells in the cortex and marker genes differentially expressed in the neuronal clusters identified by

306   the sn/scRNA-seq experiments, and we imaged ~300,000 individual cells across the MOp and its

307   vicinity (companion paper [54]).

308

309   Clustering analysis of the MERFISH-derived single-cell expression profiles resulted in a total of

310   95 cell clusters in MOp, including 42 GABAergic, 39 glutamatergic, and 14 non-neuronal

311   clusters (**Fig. 3a,b**), as well as four distinct cell clusters observed exclusively outside the MOp

312   (in striatum or lateral ventricle). These 95 clusters showed excellent correspondence with the 116

313   cell clusters identified by the sn/scRNA-seq datasets [54]. MERFISH analysis also revealed

314   clusters not identified by scRNA-seq and vice versa, mostly in the form of refined splitting of

315   clusters [54].

316

317   The spatial distribution of the clusters derived from MERFISH showed a complex, laminar

318   organization of cells in the MOp (**Fig. 3c**). MERFISH data divided glutamatergic neurons into

319   IT, ET, NP, CT, and L6b subclasses, each of which were further divided into finer clusters.

320   Many of these clusters adopted narrow distributions along the cortical depth direction that

321   subdivided individual cortical layers, though often without discrete boundaries [54]. Notably, IT

322   cells, the largest branch of neurons in the MOp, formed a largely continuous spectrum of cells

323   with gradual changes both in their expression profiles and in their cortical depth positions, in a

324   highly correlated manner [54] (**Fig. 3d**). The five major subclasses of GABAergic neurons (Lamp5,

325   Sncg, Vip, Sst and Pvalb) were also divided into finer clusters. Interestingly, many individual

326   GABAergic clusters showed layered distribution as well, preferentially residing within one or

327   two cortical layers [54]. Among the non-neuronal cell clusters, VLMCs formed the out-most layer

328   of cells of the cortex, mature oligodendrocytes and some astrocytes were enriched in white

329   matter, whereas other major subclasses of non-neuronal cells were largely dispersed across all

330   layers. In addition to the laminar organization, MERFISH analysis also revealed interesting

331   spatial distributions of cell types along the medial-lateral and anterior-posterior axes [54]. Overall,

332   the 95 neuronal and non-neuronal cell clusters in the MOp form a complex spatial organization

333   refining traditionally defined cortical layers.

334

335 Integration of retrograde tracing with MERFISH (Retro-MERFISH) further allowed us to map
336 the projection targets of different neuronal cell types in the MOp. By injecting retrograde tracers
337 into several different cortical areas (secondary motor cortex, primary somatosensory cortex, and
338 temporal association area) and imaging retrograde labels together with the MERFISH gene panel
339 in the MOp (**Fig. 3e**), we observed that all three examined target regions received inputs from
340 multiple cell clusters in the MOp, primarily from IT cells. In addition, each IT cluster projected
341 to multiple regions, with each region receiving input from a different composition of IT clusters
342 [54]. Overall, the projection of MOp neurons does not follow a simple "one cell type to one target
343 region" pattern, but rather forms a complex many-to-many network.
344
345 In summary, these MERFISH measurements revealed the spatial organization of neuronal and
346 non-neuronal cell types in the MOp with an unprecedented resolution and granularity.
347 Integration of MERFISH with retrograde tracing further allowed determination of both gene
348 expression profiles and projection targets with single-cell resolution, revealing the compositions
349 and spatial distributions of MOp neurons that project to several cortical regions.
350
351
352 **Figure 3. In situ cell-type identification, spatial mapping and projection mapping of**
353 **individual cells in the MOp by MERFISH. a,** UMAP of the ~300,000 cells in the mouse MOp
354 imaged by MERFISH. Cell clusters are grouped into 23 subclasses, and all cells in the same
355 subclass are plotted in the same color. **b,** Dendrogram showing the hierarchical relationship
356 among the 39 glutamatergic, 42 GABAergic, and 14 non-neuronal clusters in the mouse MOp
357 identified by MERFISH, colored by the subclass that each cluster belongs to. **c,** Left: Spatial map
358 of the cell clusters identified in a coronal slice (Bregma +0.90), with cells colored by their cluster
359 identity as shown in the color index. Top right: Zoom-in map of the boxed region of the left
360 panel. Bottom right: Spatial localization of individual RNA molecules in the boxed region of the
361 top right panel, colored by their gene identity. The segmented cell boundaries are colored
362 according to the cell clusters they belong to. **d,** The IT neurons in the same coronal slice as
363 shown in c. The IT neurons are colored by their cluster identity, as shown in the color index,
364 together with L6b cells in dark blue to mark the bottom border of the cortex. All other cells are
365 shown in grey. **e,** Neuronal cluster identities of the cells projecting into three other regions of the
366 brain, secondary motor cortex (MOs), primary somatosensory cortex (SSp), and temporal
367 association area (TEa). Dye-labeled cholera toxin b (CTb) are used as retrograde tracers, and the
368 CTb signals and the MERFISH gene panel are imaged in the MOp to determine both the cell
369 cluster identities (left panel) and projection targets (right panel) of individual cells. Only clusters
370 with 3 or more cells labeled by CTb are shown in color and the remaining cells are shown in
371 grey.
372
373
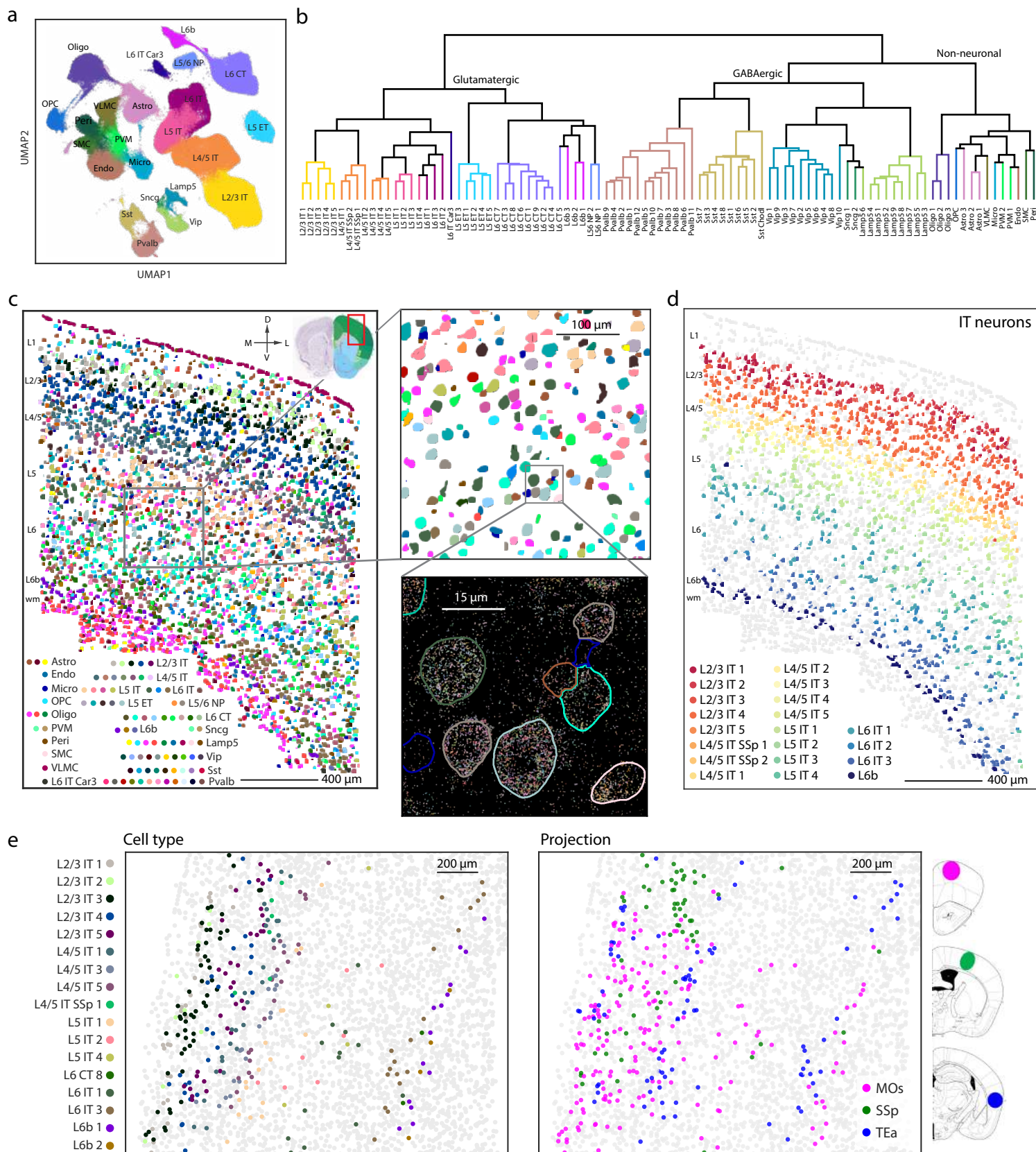
374   **Multimodal analysis of cell types with Patch-seq**
375   To characterize the electrophysiological and morphological phenotypes and laminar location of
376   the transcriptomically identified cell types, i.e., the t-types, we used the recently developed
377   Patch-seq technique [30,62]. We patched >1,300 neurons in MOp of adult mice, recorded their
378   electrophysiological responses to a set of current steps, filled them with biocytin to recover their
379   morphology (~50% of the cells) and obtained their transcriptomes using Smart-seq2 sequencing
380   (companion paper [64]). We mapped these cells to the mouse MOp transcriptomic taxonomy [45].
381   Our dataset covered all major subclasses of glutamatergic and GABAergic neurons, with cells
382   assigned to 77 t-types (**Fig. 4a**). This allowed us to describe the electrophysiological and
383   morphological phenotypes of most t-types (see examples in **Fig. 4b,c**).
384
385   We found that the measured morpho-electrical (me) phenotype of a neuron was largely
386   determined by its transcriptomic subclass, with different subclasses having distinct phenotypes.
387   For example, Sst interneurons were often characterized by large membrane time constants,
388   pronounced hyperpolarization sag, and rebound firing after stimulation offset. However, within
389   each subclass, there was substantial variation in electrophysiological and morphological
390   properties between t-types. This variation was not random but organized such that
391   transcriptomically similar t-types had more similar morpho-electric properties than distant t-
392   types. For example, excitatory t-types from the IT subclasses with more similar transcriptomes
393   were located also at adjacent cortical depths, suggesting that distances in t-space co-varied with
394   distances in the me-space, even within a layer (**Fig. 4g**). Likewise, the electrophysiological
395   properties of Sst interneurons varied continuously across the transcriptomic landscape [64].
396
397   At the level of single t-types, we found that some t-types showed layer-adapting morphologies
398   across layers (**Fig. 4e,f**) or even considerable within-type morpho-electric variability within a
399   layer. For example, Vip Mybpc1_2 neurons had variable rebound firing strength after stimulation
400   offset. Surprisingly few t-types were entirely homogeneous with regard to the measured morpho-
401   electric properties (**Fig. 4d**).
402
403   In summary, we found that the morpho-electric phenotype of a neuron in MOp was primarily
404   determined by the major subclass of neurons it belonged to, with different subclasses being
405   transcriptomically as well as morpho-electrically distinct. Within each subclass, variation in
406   electrophysiological and morphological properties often appeared to be continuous across the
407   transcriptomic landscape, without clear-cut boundaries between neighbouring t-types.
408
409   Patch-seq also permits direct comparison of the morpho-electric properties of homologous cell
410   types across species [48]. Here we focused our analysis on one of the most recognizable
411   mammalian neuron types, the gigantocellular Betz cells found in M1 of primates and large
412   carnivores. These neurons are predicted to be in the layer 5 ET (L5 ET) subclass [48], which also
**413**   contains the homologous corticospinal projecting neurons in the mouse. To allow cross-species

414   analysis of primate Betz cells and mouse ET neurons, we first created a joint embedding of
415   excitatory neurons in mouse, macaque and human, which showed strong homology across all
416   three species for the L5 ET subclass (**Fig. 4h**). Patch-seq recordings were made from L5 neurons
417   in acute and cultured slice preparations of mouse MOp and macaque M1. We also capitalized on
418   a unique opportunity to record from neurosurgical tissue excised from the human premotor
419   cortex, which also contains Betz cells, during an epilepsy treatment surgery. To permit
420   visualization of cells in heavily myelinated macaque M1 and human premotor cortex, AAV
421   viruses were used to drive fluorophore expression in glutamatergic neurons in slice culture.
422
423   Patch-seq cells in each species that mapped to the L5 ET subclass (**Fig. 4h**) were all large layer 5
424   neurons that sent apical dendrites to the pial surface (**Fig. 4i**, note truncation in human Betz cell).
425   However, macaque and human L5 ET neurons were much larger, and had long "tap root" basal
426   dendrites that are a canonical hallmark of Betz cells [83]. Subthreshold membrane properties were
427   relatively well conserved across species. For example, L5 ET neurons in all three species had a
428   low input resistance, although it was exceptionally low in macaque and human (**Fig. 4j**).
429   Conversely, suprathreshold properties of macaque and human Betz/ET neurons were highly
430   specialized. Most notably, human and macaque neurons responded to prolonged suprathreshold
431   current injections with a biphasic firing pattern in which a pause in firing early in the sweep was
432   followed by a dramatic increase in firing late in the sweep (**Fig. 4k**). Intriguingly, we identified
433   several genes encoding ion channels that were enriched in macaque and human L5 ET neurons
434   compared with mouse (**Fig. 4l**). These primate specific ion channels may contribute to the
435   distinctive suprathreshold properties of primate ET neurons. Together this indicates that primate
436   Betz cells are homologous to mouse thick-tufted L5 ET neurons, but display phenotypic
437   differences in their morphology, physiology and gene expression. Similar to transcriptomics,
438   these results indicate strong conservation of cell subclasses but with significant species
439   specializations in anatomical and functional properties.
440
441

A multimodal cell census and atlas of the mammalian primary motor cortex
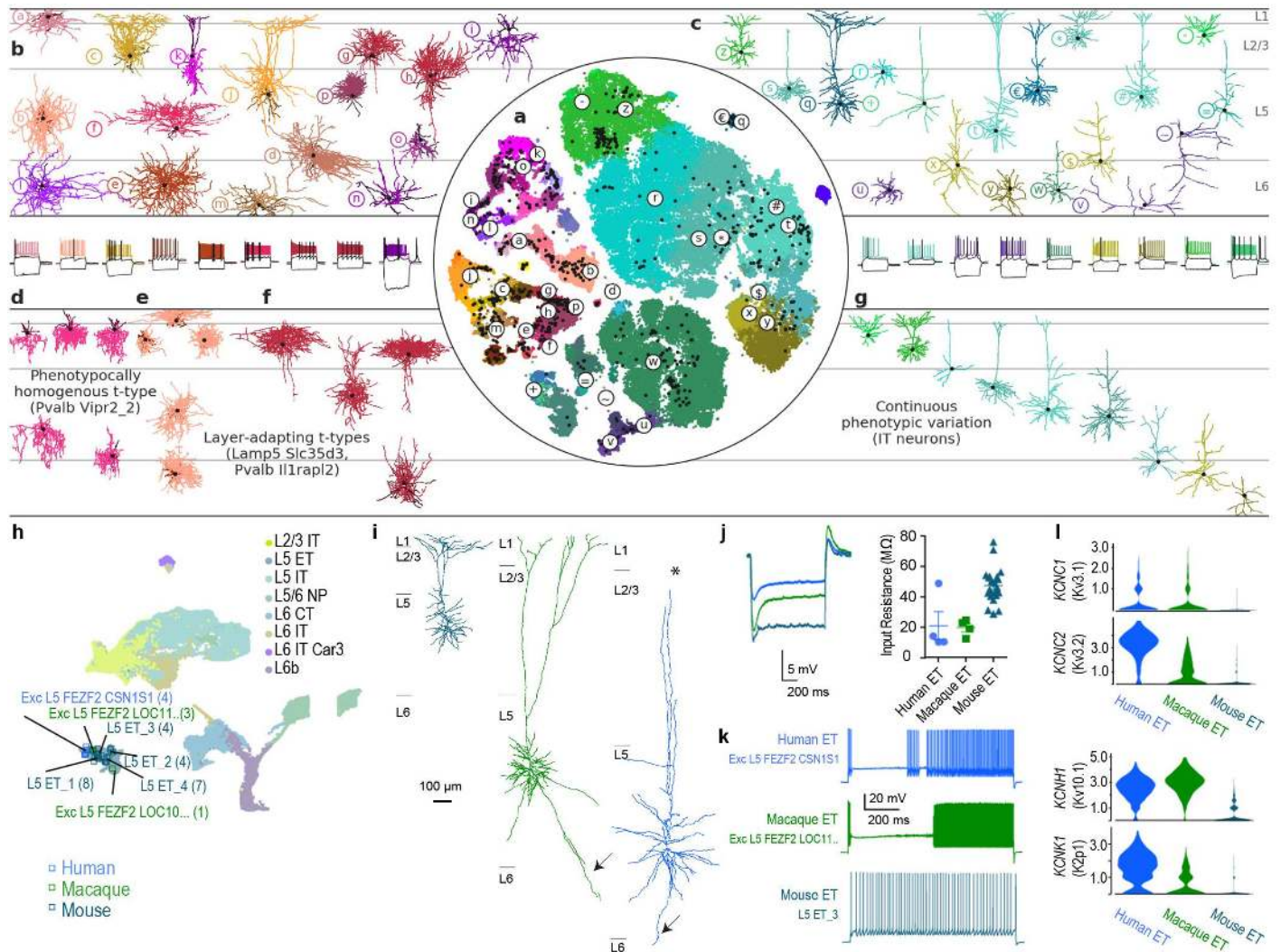
442

Figure 4. Correspondence between transcriptomic and morpho-electrical properties of mouse MOp neurons by Patch-seq, and cross-species comparison of L5 ET neurons. a, t-SNE of the scRNA-seq 10x v2 dataset [45] with the Patch-seq neurons (black dots) positioned on top of it [84]. **b**, Examples of GABAergic interneuron morphologies and electrophysiological recordings (below). Letters refer to cells marked in a. **c**, Examples of glutamatergic excitatory neuron morphologies and electrophysiological recordings. **d**, Example of a phenotypically homogenous t-type (Pvalb Vipr2_2, chandelier neurons). **e-f**, Two examples of t-types showing layer-adapting morphologies (e, Lamp5 Slc35d3, neurogliaform cells; f, Pvalb Il1rapl2, fast-spiking basket cells). **g**, Example of a transcriptomic subclass (excitatory IT neurons) that shows continuous within-subclass co-variation between distances in transcriptomic space and morphological space (compare the color ordering in **a** (right) with the color ordering in **g**. **h**, UMAP visualization of cross-species integration of snRNA-seq data for glutamatergic neurons isolated from mouse, macaque and human, with colors corresponding to cell subclass. Patch-seq samples mapping to various ET neuron types are denoted by squares, color-coded by species. **i**,

A multimodal cell census and atlas of the mammalian primary motor cortex

457   Dendritic reconstructions of L5 ET neurons. The human (Exc L5 FEZF2 CSN1S1) and macaque
458   (Exc L5 FEZF2 LOC114676463) neurons display classical Betz cell features, including taproot
459   dendrites (arrows). Note, the human neuron is truncated (asterisk) before reaching the pial
460   surface. **j**, Voltage response of mouse, macaque and human ET neurons to a 1 s, -300 pA current
461   injection (left). Input resistance is low in all species, but exceptionally low in human and
462   macaque Betz cells. Error bars represent SEM (right; macaque n=4, human n=4, mouse n=22;
463   FDR corrected two-sided Wilcoxon ranked sum test (human vs mouse W=12, p = 0.31, d=2.09;
464   human vs monkey W = 5, p = .49, d=.08; monkey v mouse W = 0 p = .0004., d =2.5). **k**,
465   Example spike trains in response to a 10s suprathreshold current injection. Macaque and human
466   L5 ET neurons tended to respond with a distinctive, biphasic firing pattern. **l**, Violin plots of
467   enriched potassium channel gene expression in human and macaque compared to mouse L5 ET
468   neurons.
469
470
471   **Multimodal correspondence by Epi-Retro-Seq**
472   To obtain a comprehensive view of the molecular diversity among projection neurons in MOp,
473   we developed Epi-Retro-Seq (companion paper [79]) and applied it to mouse MOp neurons that
474   project to each of the 8 selected brain regions that receive inputs from MOp (**Fig. 5a**). The target
475   regions included two cortical areas, SSp and anterior cingulate area (ACA), and six subcortical
476   areas, striatum (STR), thalamus (TH), superior colliculus (SC), ventral tegmental area and
477   substantia nigra (VTA+SN), pons, and medulla (MY). Specifically, we injected the retrograde
478   tracer rAAV2-retro-Cre [77] into the target region in INTACT mice [85], which turned on Cre-
479   dependent GFP expression in the nuclei of MOp neurons projecting to the injected target region.
480   Individual GFP-labeled nuclei of MOp projection neurons were then isolated using fluorescence-
481   activated nucleus sorting (FANS). Single-nucleus methylcytosine sequencing (snmC-Seq2) [49]
482   was performed to profile the DNA methylation (mC) of each single nucleus.
483
484   After removing low-quality cells, potential doublets, and non-neuronal cells, we obtained high-
485   quality methylomes for 2,111 MOp projection neurons. When co-clustering them with MOp
486   neurons collected without enrichment of specific projections, we observed a precise agreement
487   among all of the major cell subclasses (**Fig. 5b,c**), demonstrating the robustness of Epi-Retro-
488   Seq to classify cell types. Although neurons projecting to different target regions were not
489   completely separated on t-SNE, we observed the explicit enrichment of cortico-cortical and
490   cortico-striatal projecting neurons in IT subclasses (L2/3, L4, L5 IT, L6 IT, and L6 IT Car3), and
491   cortico-subcerebral projecting neurons in L5 ET. Many cortico-thalamic projecting neurons were
492   also observed in L6 CT subclass (**Fig. 5d**). These observations are consistent with the known
493   laminar distribution of the cortico-cortical and cortical-subcortical projection neurons [81],
494   reflecting the high quality of retrograde-labeling of neuronal nuclei in our Epi-Retro-Seq dataset.
495

A multimodal cell census and atlas of the mammalian primary motor cortex

496   The enrichment of L5 ET neurons in the Epi-Retro-Seq data (40.2% vs. 5.62% in unbiased
497   profiling of MOp using snmC-seq2) allowed a more detailed investigation of the subtypes of L5
498   ET neurons which are known to project to multiple subcortical targets in TH, VTA+SN, pons
499   and MY [81]. The 848 L5 ET neurons further segregated into 6 clusters (**Fig. 5e,f**). MY-projecting
500   neurons showed a clear enrichment in L5 ET cluster 0 (**Fig. 5f,g**), in agreement with scRNA-Seq
501   data for anterolateral motor cortex (ALM), part of MOs [15,86]. We used gene body non-CG
502   methylation (mCH) levels to integrate the L5 ET Epi-Retro-Seq data with the ALM Retro-seq
503   data and also observed the enrichment of MY-projecting cells in the same cluster [79].
504
505   A major advantage of DNA methylation profiling of neurons is its ability to obtain information
506   for both genes and cis-regulatory elements. Specifically, mCH at gene bodies is strongly anti-
507   correlated with gene expression in neurons, while promoter-distal differentially CG-methylated
508   regions (CG-DMRs) are reliable markers of regulatory elements such as enhancers [20]. We thus
509   identified 511 differentially CH-methylated genes (CH-DMGs) and 58,680 CG-DMRs across the
510   L5 ET clusters (**Fig. 5h**). We also inferred transcription factors (TFs) that may contribute to
511   defining the cell subclusters by identifying enriched TF-binding DNA sequence motifs within
512   CG-DMRs (**Fig. 5i**). For example, *Ascl1* is a transcription factor whose motif was significantly
513   enriched in the MY-projecting cluster. Previous studies had shown its necessity for neuronal
514   differentiation and specification in multiple regions of the nervous system [87,88]. In addition, 230
515   hypo-CH-DMGs were identified between the MY-projecting cluster and other projection
516   neurons. Interestingly, one of the most differentially methylated genes is *Ptprg* (**Fig. 5j**), which
517   interacts with contactin proteins to mediate neural projection development [89].
518
519   In summary, Epi-Retro-Seq mapping data for MOp revealed specific enrichment of MY-
520   projecting neurons in one of the molecularly-defined subpopulations of MOp L5 ET neurons,
521   allowing identification of regulatory elements for this unique cell type. In addition to MOp, we
522   have performed 63 Epi-Retro-Seq mapping experiments for 7 cortical regions, comprising 26
523   cortico-cortical projections and 37 cortico-subcortical projections [79]. Together, these epigenomic
524   mapping data for projection neurons facilitates the understanding of gene regulation in
525   establishing neuronal identity and connectivity, by discovering projection-specific gene
526   regulatory elements which can be used to target specific types of projection neurons.
527
528
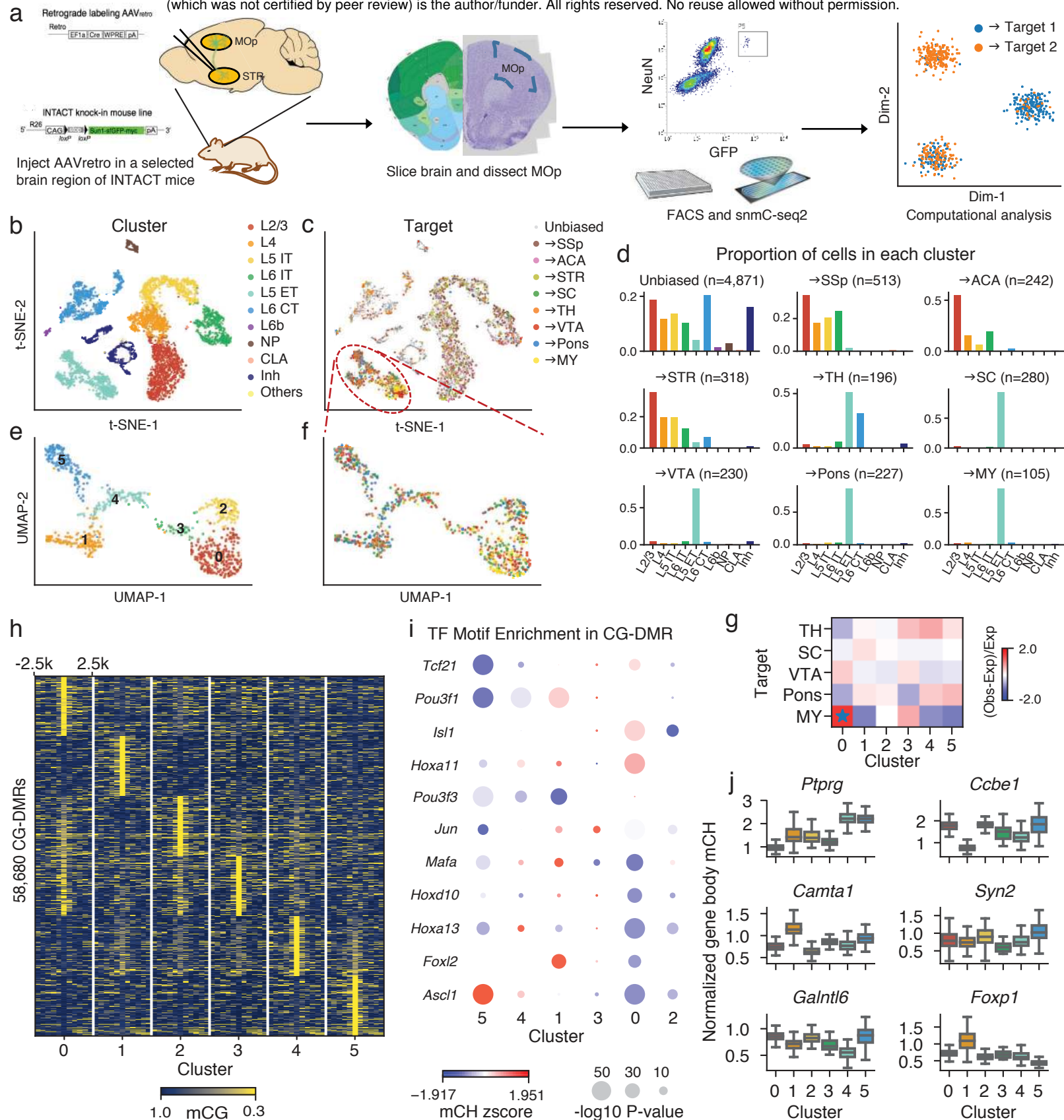
529 **Figure 5. Epi-Retro-Seq links molecular cell type with distal projection targets. a,**
530 Workflow of Epi-Retro-Seq. **b, c,** UMAP embedding of MOp cells profiled by Epi-Retro-Seq
531 (n=2,115) and unbiased snmC-Seq2 (n=4,871) computed with 100kb-bin-level mCH, colored by
532 subclasses (**b**) or projection targets (**c**). **d,** Distribution across subclasses of neurons from
533 unbiased snmC-Seq2 and neurons projecting to each target. **e, f,** UMAP embedding of L5 ET
534 cells in MOp profiled by Epi-Retro-Seq (n=848) computed with 100kb-bin-level mCH, colored
535 by clusters (**e**) or projection targets (**f**). **g,** Enrichment of L5 ET neurons projecting to each target
536 in each cluster. * represents FDR<0.05. **h,** mCG levels at CG-DMRs identified between the six
537 clusters and their flanking 2.5k regions. Top 100 DMRs in each cluster were shown. **i,** TF motif
538 enrichment in CG-DMRs in each cluster. Color represents z-scored gene-body mCH level of the
539 TFs, and size represents -log10 *P* value of motif enrichment in the CG-DMRs. **j,** Boxplots of
540 normalized mCH levels at gene-bodies of example CH-DMGs in the six clusters. Numbers of
541 cells represented by the boxes are 242, 165, 118, 42, 119, and 162 for the six clusters. The
542 elements of boxplots are defined as: center line, median; box limits, first and third quartiles;
543 whiskers, 1.5× interquartile range.

544
545

546 **MOp projection neuron types and input-output wiring diagram**
547 Building upon the molecularly defined and spatially resolved cell atlas (**Fig. 3**) and the multi-
548 modal correspondence between gene expression and morpho-electric properties of MOp
549 neurons (**Fig. 4**), we next describe a comprehensive cellular resolution input-output MOp
550 wiring diagram. To achieve this, we combined classic tracers, genetic viral labeling in Cre
551 driver lines and single neuron reconstructions with high-resolution, brain-wide imaging,
552 precise 3D registration to CCF, and computational analyses (companion paper [69]).

553

554 First, we systematically characterized the global inputs and outputs of MOp upper limb (MOp-
555 ul) region using classic anterograde (PHAL) and retrograde (CTb) tract tracing [69] (**Fig. 6a**). At
556 the macro-scale, MOp-ul projects to more than 110 gray matter regions and cervical spinal
557 cord, and ~60 structures in the cerebral cortex and thalamus project back to MOp-ul.

558

559 Next, we generated a fine-grained areal and laminar distribution map of multiple MOp-ul
560 projection neuron populations using retrograde pathway-tracing. Accordingly, we identified 25
561 distinct neuron projection types based on their unique combinations of axonal targets and
562 laminar somatic distributions (**Fig. 6b**, top; for details see [69]). For example, IT cells (e.g. TEa-
563 targeting or contralateral MOp-targeting) are distributed throughout L2-L6b; ET cells (pons- or
564 medulla-targeting) are distributed primarily in L5b and most CT (posterior thalamic nucleus-
565 targeting) neurons are distributed in L6a.

566

567 In parallel with these tracer-labeled, projection- and layer-defined cell types, we quantitatively
568 characterized the distribution patterns of neuronal subpopulations in the MOp-ul labeled in 28

569 Cre-expressing "driver" lines (**Fig. 6b**, bottom). These lines selectively label neurons from
570 different IT (e.g. Cux2, Plxnd1, Tlx3), L5 ET (Rbp4, Sim1, Fezf2), and CT (Ntsr1, Tle4)
571 subpopulations with distinct laminar distributions [75,90,91].

573 Subsequently, we used viral tracers to systematically examine MOp-ul cell-type-specific inputs
574 and outputs (**Fig. 6c**). First, neurons projecting *to* Cre-defined starter cells were labeled using
575 transsynaptic rabies viral tracing methods; an example from the *Tlx3* L5 IT line is shown in
576 **Fig. 6c** (upper left, red). Projections *from* MOp were labeled following AAV-GFP injections
577 into C57BL6/J mice, revealing patterns consistent with PHAL tracing results (**Fig. 6a**).
578 Projections from L2/3 IT, L4 IT, L5 IT, L5 ET, and L6 CT cells were mapped following
579 injections of Cre-dependent viral tracers into Cre lines selective for these laminar- and
580 projection- cell subclasses [71]. Most Cre line anterograde tracing experiments revealed a
581 component of the overall output pathway (**Fig. 6c**). For example, the L6 *Ntsr1* line revealed a
582 typical CT projection pattern with dense projections specific to thalamic nuclei. This result is
583 consistent with labeling from retrograde injections in various thalamic nuclei (PO, VAL, VM)
584 and cortical areas such as MOs and SSp (**Fig. 6b**, top). Further characterization of the
585 distinctive projection patterns of several IT, L5 ET, and CT driver lines is provided in the
586 anatomy companion paper [69].

588 To further refine the projection neuron characterization, we carried out single cell analysis by
589 combining sparse labeling, high-resolution whole-brain imaging, complete axonal
590 reconstruction and quantitative analysis (companion papers [68,69]); additional analysis was also
591 conducted using BARseq [69], a high-throughput projection mapping technique based on *in situ*
592 sequencing [67]. We augmented the full morphology reconstruction dataset with publicly
593 available single cell reconstructions in MOp from the Janelia Mouselight project [26]. We
594 systematically characterized axonal projections of 151 single MOp pyramidal neurons. This
595 analysis revealed a rich diversity of projection patterns within the IT, ET and CT subclasses
596 (**Fig. 6c,d**). For example, individual L6 neurons display several distinct axonal arborization
597 targets that likely contribute to the composite subpopulation output described for the *Ntsr1* and
598 *Tle4* diver lines (**Fig. 6d**). Confirming and extending previous reports [86], we characterized
599 detailed axonal trajectories and terminations of two major types of L5b ET cells, namely
600 medulla-projecting and non-medulla projecting neurons; both types may collateralize in the
601 thalamus and terminate in the midbrain (**Fig. 6d**). Individual IT cells across L2-L6 also
602 generate richly diverse axonal trajectories (detailed in [68,69]. Further analyses of complete single
603 neuron morphologies, precisely registered in the CCF, will provide the ultimate resolution
604 toward defining anatomical cell types and clarify the anatomical heterogeneity described at the
605 subpopulation level.

607 In summary, combining multiple approaches complementary in their coverage, throughput, and
608 resolution, we provide a comprehensive identification of major projection neuron types with
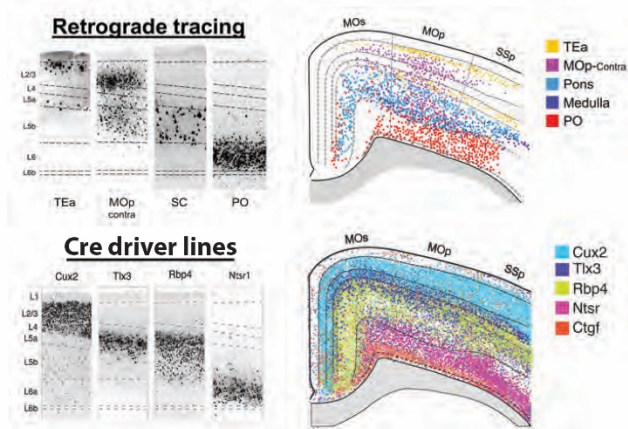
A multimodal cell census and atlas of the mammalian primary motor cortex

609   correspondence to molecular markers. We further delineate their input-output patterns at the
610   subpopulation level and describe projection patterns at single-cell resolution, deriving the first
611   multi-scale wiring diagram of MOp. A major future goal is to link these anatomic and
612   especially projection types with transcriptomic types (**Fig. 2b**), with precise registration to a
613   spatial atlas (e.g. **Fig. 3e**).
614

a. Global inputs and outputs of the MOp-ul
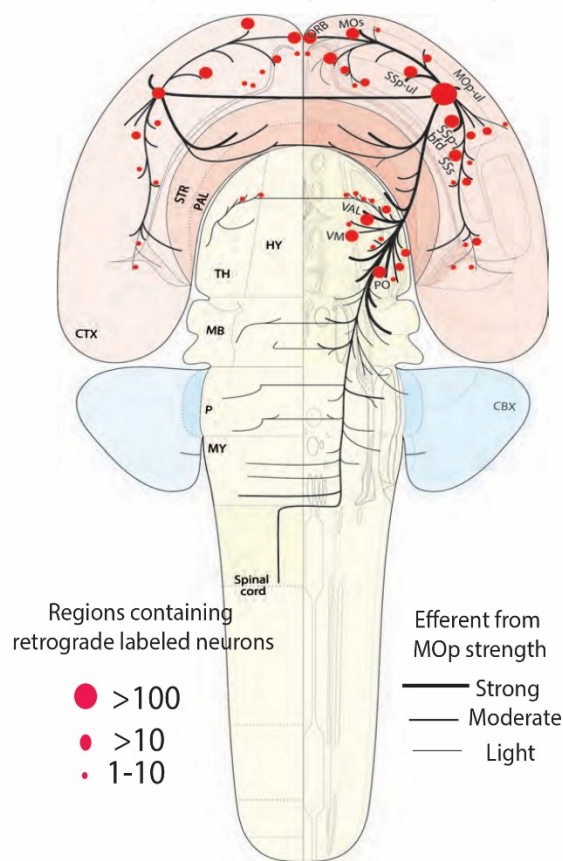
b. Characterization of MOp neuron types

c. 3D view of multi-scale projection pathways
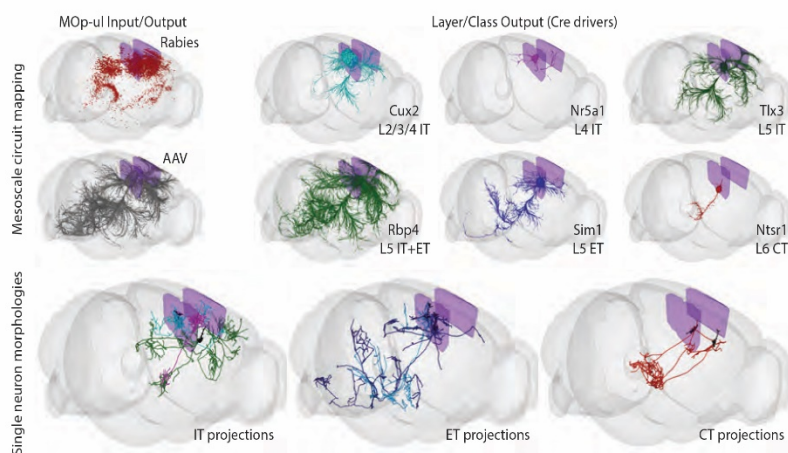
d. MOp neuron type schema
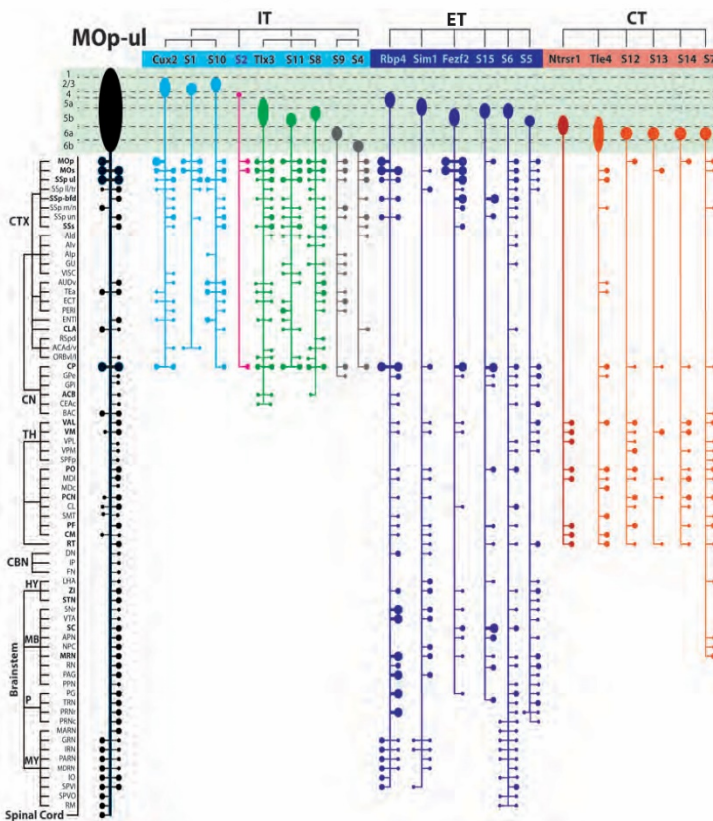
Figure 6. Global wiring diagram and anatomical characterization of MOp-ul neuron types. a, Flatmap representation of the MOp-ul input/output wiring diagram. Black lines and red dots indicate axonal projections (outputs) and retrograde labeling sources (inputs), respectively, with line thickness and dot sizes representing relative connection strengths. Most MOp-ul projection targets in the cortex and thalamus also contain input sources, suggesting bi-directional connections. The flatmap is adapted from the Swanson Brainmap 4.0 and the mouse brain flatmap [92]. **b,** MOp-ul neurons classified by projection targets or transgenic *Cre* expression. (Top) Retrograde tracing using CTB revealed layer-specific distributions of MOp-ul neurons with respect to their major projection targets. Representative images (left) show neurons labeled by CTB injections into cortical areas (TEa, contralateral MOp), superior colliculus (SC) in the midbrain, and posterior complex (PO) of the thalamus. Detected cells were pseudo-colored and overlaid onto a schematic coronal section near the center of MOp-ul (right). MOp neurons that project to TEa are distributed in L2 and L5 (yellow), to the contralateral MOp in L2-L6b (purple), to targets in the pons and medulla in L5b (blue), and to thalamus in L6a (red). (Bottom) The distribution of neurons labeled in 28 transgenic *Cre* lines was mapped in MOp and across the whole cortex. Images (left) show laminar patterns of Cre+ nuclei in MOp-ul from four driver lines (*Cux2*, *Tlx3*, *Rbp4*, and *Ntsr1*). Detected nuclei from these lines, plus the *Ctgf-Cre* line, were pseudo-colored and overlaid onto a schematic coronal section near the center of MOp-ul (right). Cre+ nuclei are found in L2-4 in Cux2; L5a and superficial L5b in Tlx3; L5a and L5b in Rbp4; L6a in Ntsr1, and L6b in Ctgf. **c,** 3D views show brain-wide MOp input-output patterns at the population and single cell resolution. (Top left) Regional MOp inputs and outputs were mapped using retrograde (in red, example shows rabies tracing from the *Tlx3-Cre* driver line) and anterograde (in black, example shows AAV-EGFP) tracing methods. (Top right) Whole-brain axonal trajectories from 6 Cre line-defined subpopulations labeled with Cre-dependent AAV tracer injections at the same MOp-ul location. (Bottom) Individual projection neurons were fully reconstructed following high-resolution whole-brain imaging of sparsely labeled cells. Representative examples of IT, ET, and CT neurons are shown in each panel. The two ET examples represent distinct projection-types; medulla (dark blue)- and non-medulla-projecting (light blue). 3D renderings were generated following registration of projection and reconstruction data into CCFv3 using BrainRender [93]. **d,** Projection patterns arising from major cell types, IT, ET and CT, with corresponding Cre-line assignment and somatic laminar location, compared with the overall projection pattern from the MOp-ul region (left, black). Along each vertical output pathway, horizontal bars on the right and left sides represent ipsilateral and contralateral collaterals, respectively, with dot sizes indicating the strength of axonal terminals in different targets. Brain structure nomenclature adopted from ARA [94].

Cell Type Targeting Tools

A multimodal cell census and atlas of the mammalian primary motor cortex

654   The identification and classification of MOp cell types based on single-cell integration of
655   transcriptomes and epigenomes (**Fig. 2**), spatially resolved single-cell transcriptomics (**Fig. 3**)
656   and anatomical and physiological analysis (**Fig. 4-6**) provides deep insights into the molecular
657   basis of cellular diversity. In addition to establishing a principled basis for a taxonomy of brain
658   cell types, knowledge of cellular gene expression also provides information to create mouse
659   models in which genetically encoded reporters and actuators are targeted to these molecularly
660   defined cell types [33].

661

662   As an embodiment of this approach, we used CRISPR/Cas-9-mediated homologous
663   recombination in ES cells to generate genetically modified mice (Stafford, Daigle, Chance et
664   al., companion manuscript in preparation) in which sequences encoding FlpO and Cre
665   recombinases were targeted respectively to *Npnt* and *Slco2a1*, genes whose differential
666   expression discriminates between two types of L5 ET neurons with distinct subcortical
667   projection target specificities [15,86]. Confirming the assignment of *Npnt*- and *Slco2a1*-expressing
668   cells to subsets of L2/3 IT and L5 ET neurons in the consensus transcriptomic taxonomy (**Fig.
669   7a**), FlpO- and Cre-dependent tdTomato reporter expression in *Npnt-P2A-FlpO;Ai65F* and
670   *Slco2a1-P2A-Cre;Ai14* mice localized to these cortical cell layers in MOp (**Fig. 7b**). In *Npnt*
671   mice, both L2/3 and L5 neurons were labeled. In *Slco2a1* mice, predominantly L5 neurons
672   were labeled. It is noteworthy that *Slco2a1* labeled cells occupying a deeper sub-lamina of L5
673   than those targeted by *Npnt*, in accord with a previous report describing the two types of L5 ET
674   neurons [86] (see also **Fig. 9** below). To test the projection specificity of neurons labeled by these
675   novel genetic tools, we injected a recombinant AAV encoding a Cre-dependent EGFP reporter
676   into deep L5 in MOp of a *Slco2a1-P2A-Cre* mouse (**Fig. 7c**). Consistent with previous studies
677   [86] as well as those described in **Figures 5, 6 and 9** (below), GFP-labeled axon terminals were
678   found in pontine gray and medulla, indicating that this mouse line labels the medulla-
679   projecting L5 ET cell type.

680

681   To expand on cell type driver lines, we further built a genetic toolkit for cortical pyramidal
682   neurons (PyNs) with more comprehensive coverage of projection types and with combinatorial
683   strategies for improved specificity (companion paper [75]). First, we generated and characterized
684   a set of 15 Cre and Flp gene knockin mouse driver lines for targeting major PyN
685   subpopulations and progenitor types, guided by knowledge in their gene expression as well as
686   developmental genetic programs (**Fig. 7d,e**). These include the broad CT (*Tbr1, Tle4, Foxp2*),
687   ET (*Fezf2, Adcyap1, Tcerg1l, Sema3e*) and IT (*Plxnd1, Cux1,* and *Tbr1* late embryonic
688   inductions) subclasses as well as subpopulations within these subclasses. When crossed with
689   reporter alleles, these driver lines activated reporter expression that precisely recapitulated
690   endogenous expression patterns highlighted here with 4 representative lines (**Fig. 7f**): L2/3 and
691   L5a for IT-*Plxnd1* (IT^{Plxnd1}), L5b and L6 for ET- *Fezf2* (ET^{Fezf2}), L6 for CT-*Tle4* and CT-
692   *Foxp2* (CT^{Tle4}, CT^{Foxp2}). To examine the projection pattern of these driver-defined
693   subpopulations, we converted inducible CreER expression to constitutive Flp expression

694 followed by MOp injection of a Flp-dependent AAV reporter vector (**Fig. 7g-i**). Largely as
695 expected, IT^{Plxnd1} projected to multiple ipsi- and contra-lateral cortical areas and the
696 striatum/caudate putamen; ET^{Fezf2} projected robustly to several ipsi-lateral cortical sites,
697 striatum, and numerous subcortical targets including thalamus, medulla and the corticospinal
698 tract; CT^{Tle4} projected to a set of highly specific thalamic nuclei. Surprisingly, CT^{Foxp2}
699 projected to a set of specific thalamic nuclei as well as to midbrain, brainstem and corticospinal
700 tract. Further characterization of this set of new driver lines (**Fig. 7e**) is presented in [75].
701
702 Together, these tools and strategies establish an experimental approach for accessing
703 hierarchically organized neuronal cell types at progressively finer resolution. Such genetic
704 access will enable an integrated multi-modal analysis to further validate and characterize these
705 cell populations as well as to explore their multi-faceted function in neural circuit operation
706 and behavior.--
707
708
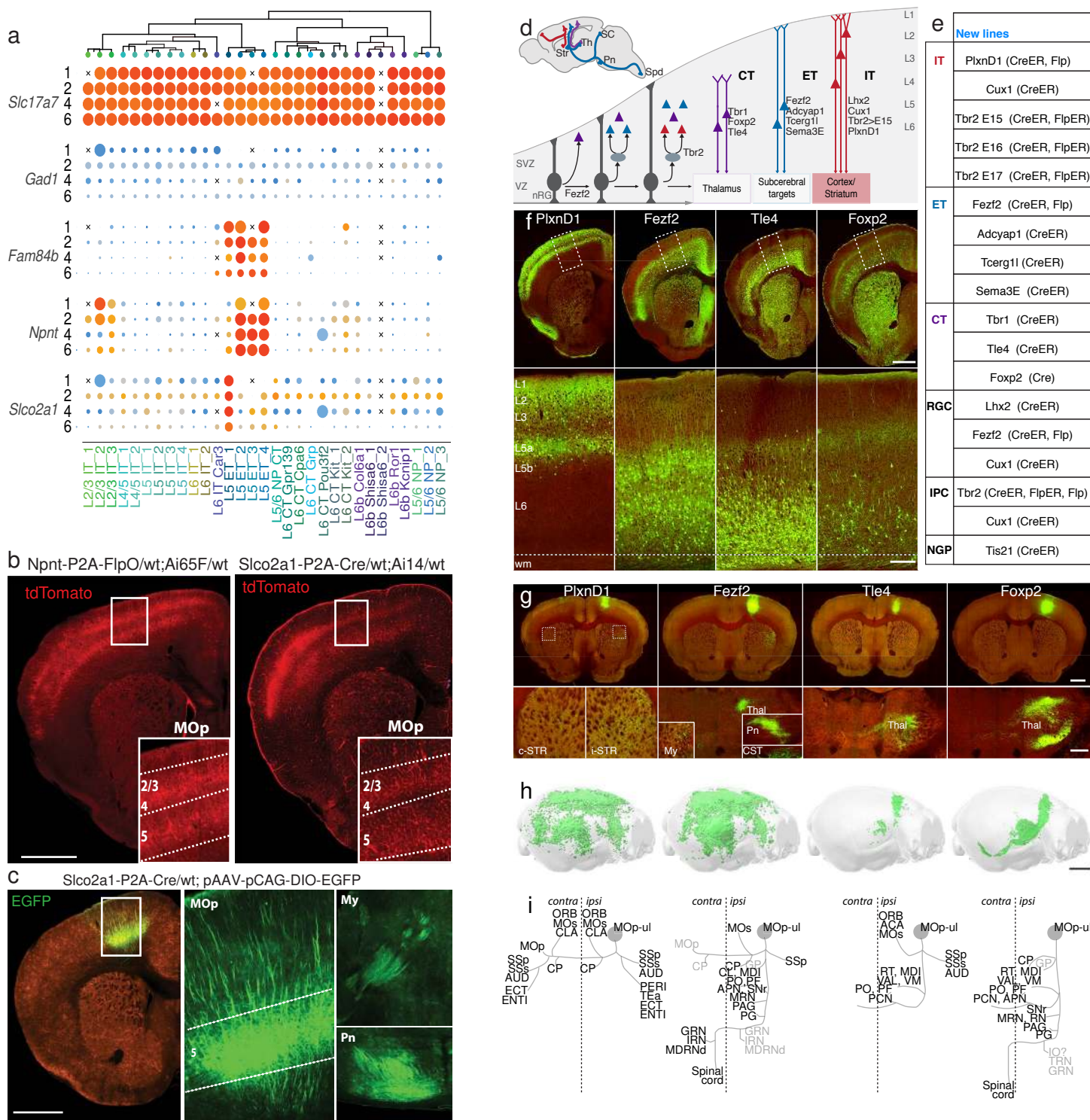
709 **Figure 7. Genetic tools for targeting cortical glutamatergic projection neuron types. a,**
710 Dendrogram of excitatory cells types within the MOp transcriptomic taxonomy followed by the
711 proportion of cells (dot size) expressing pan-excitatory or -inhibitory marker genes (*Slc17a7* or
712 *Gad1*), or L5 ET marker genes (*Fam84b*, *Npnt*, and *Slco2a1*). Expression results from four
713 different single cell RNA-seq platforms are shown: 1: scRNA-seq SMART-Seq; 2: scRNA-seq
714 10X v3 A; 4: snRNA-seq SMART-Seq; 6: snRNA-seq 10X v3 B [45]. **b,** Representative images of
715 native tdTomato fluorescence from MOp of *Npnt-P2A-FlpO;Ai65F* and *Slco2a1-P2A-Cre;Ai14*
716 animals. Reporter expression was observed in L2/3 and L5 neurons with both driver lines and in
717 the vasculature with only the Slco2a1 line. **c,** Representative images of native EGFP
718 fluorescence from MOp, My (medulla), and Pn (pons) in the brain of an *Slco2a1-P2A-Cre*
719 animal injected in MOp with a Cre-dependent reporter AAV (*pCAG-FLEX-EGFP-WPRE*).
720 Robust reporter expression in L5 neurons was observed at the injection site (MOp) and in fibers
721 terminating in My and Pn. **d,** Schematic (upper left panel) depicting several major pyramidal
722 neuron (PyN) projection classes that mediate intra-telencephalic streams (IT-red; cortical and
723 striatal) and cortical output channels (ET-blue, CT-purple). Str, striatum; Th, thalamus; SC,
724 superior colliculus; Spd, spinal cord. Developmental trajectory of PyNs (lower panel) depicting
725 lineage progression from progenitors to mature PyNs across major laminar and projection types.
726 Genes used to target progenitor and PyN subpopulations are listed according to their cellular
727 expression patterns. VZ, ventricular zone; SVZ, subventricular zone. **e,** Table presenting new
728 gene knockin driver mouse lines targeting PyN progenitors and projection types. RGC, radial
729 glia cell; IPC, intermediate progenitor cell; NGP, neurogenic progenitor. **f,** Cre recombination
730 patterns visualized through reporter expression (green) and background autofluorescence (red)
731 from four driver/reporter lines *PlexinD1-2A-CreER* (*PlxnD1*);*Snap25-LSL-EGFP, Fezf2-2A-*
732 *CreER* (*Fezf2*)*;Ai14, Tle4-2A-CreER* (*Tle4*)*;Snap25-LSL-EGFP* and *Foxp2-IRES-Cre*
733 (*Foxp2*)*;AAV9-CAG-FLEX-EGFP* (systemic injection). Top row: coronal hemisections
734 containing MOp. Bottom row: a segment of MOp (dashed lines, top row) with laminar
735 delineations. *CreER* Tamoxifen (TM) inductions were at P21 and P28. **g,** Anterograde tracing
736 from PyN subpopulations in MOp. CreER drivers were crossed with a *Rosa26-CAG-LSL-Flp*
737 mouse, and postnatal TM induction to convert to constitutive Flp expression for anterograde
738 tracing with a Flp-dependent AAV vector expressing EGFP (*AAV8-CAG-fDIO-TVA-EGFP*).
739 Representative images of native EGFP fluorescence from the MOp injection site (top row) from
740 cell-type-specific viral vector (green) and background autofluorescence (red) at selected
741 subcortical projection targets for four driver lines: Th; Str; cerebral peduncle (cp), Pn, My and
742 corticospinal tract (CST). **h,** Whole-brain three dimensional renderings of axon projections
743 registered to the CCFv3 for each PyN subpopulation in the MOp cortex (parasagittal view). **i,**
744 Schematics of main projection targets for each PyN subpopulation. Vertical dashed line indicates
745 midline; filled circle indicates MOp injection site. Scale bars: hemisections (f & g) and h, 1mm;
746 bottom row in f, 200μm; bottom row in g, 500μm; h, 2 mm.
747
748

A multimodal cell census and atlas of the mammalian primary motor cortex

749 **Integrated multimodal characterization reveals L4 IT neurons in MOp**
750 To investigate if our collective multimodal characterization can lead to an integrated
751 understanding of cell types in MOp, we selected two case studies to demonstrate convergence of
752 multiple corresponding properties onto specific cell types.
753
754 Traditionally MOp has been considered an agranular cortical area, defined by the lack of a
755 cytoarchitectonic layer 4 which usually contains spiny stellate or star pyramid excitatory
756 neurons. However, a previous study challenged this view and presented evidence that L4 neurons
757 similar to those typically found in sensory cortical areas also are present in MOp [95]. Here as the
758 first case study, we used multimodal evidence to confirm the presence of L4-like neurons in
759 mouse MOp and possibly in primate M1 as well (**Fig. 8**).
760
761 We first performed a joint clustering (see Methods) and UMAP embedding of all IT cells
762 (excluding the highly distinct L6 IT Car3 cells) from 11 different mouse molecular datasets,
763 including 6 sc/snRNA-seq datasets, and the snmC-Seq2, snATAC-Seq, Epi-Retro-Seq,
764 MERFISH and Patch-seq data (**Fig. 8a**). This resulted in 5 joint clusters, mostly along a
765 continuous variation axis moving from L2/3 to L4/5 to L5 to L6. The joint clustering enabled
766 linkage of the cells independently profiled by each individual modality into types -
767 transcriptomic, epigenomic, spatially resolved transcriptomic, and morpho-electric-
768 transcriptomic, and cross-correlation of these disparate properties. Consequently, we identified
769 epigenomic peaks linked to cluster-specific marker genes - *Cux2* for L2/3 IT and L4/5 IT (1),
770 *Rspo1* for L4/5 IT (1), *Htr2c* for L4/5 IT (2-3), and *Rorb* for L4/5 IT and L5 IT (**Fig. 8b**, cluster
771 names from SingleCellFusion). MERFISH data also showed that L4/5 IT and L5 IT cells
772 occupied distinct layers, and the L4/5 IT type expressed *Rspo1* (**Fig. 8c**), a L4 cell type marker in
773 sensory cortical areas identified in previous studies [15]. Transcriptomic IT types from mouse
774 corresponded well with those from human and marmoset, but such correspondence was mostly at
775 main branches or subclass level while significant confusions existed at single cluster level (**Fig.
776 8d**), likely due to the substantial gene expression variation between rodents and primates (**Fig.
777 2**). In particular, mouse L4/5 IT 1 and 2 transcriptomic clusters together corresponded to a set of
778 5-7 L3-5 IT clusters in human and marmoset.
779
780 We further compared the L4-like cells in mouse MOp with those from mouse primary visual
781 cortex (VISp) [15] after co-clustering all the SMART-Seq glutamatergic neurons from both regions
782 (**Fig. 8e**). In the UMAP representation, L4/5 IT cells in MOp occupied a subspace of the L4 IT
783 co-cluster defined by the intersection of marker genes *Cux2* and *Rorb*, suggesting that L4-like
784 cells in MOp are similar to a subset of L4 cells in VISp while the L4 cells in VISp have
**785** additional diversity and specificity.
786
787 L4-like IT cells in MOp also exhibited morphological features characteristic of traditionally
788 defined L4 excitatory neurons. From the Patch-seq study [64], cells from the L4/5 IT_1 type had no

789   or minimal apical dendrites without tufts in layer 1, in contrast to cells from the neighboring
790   L2/3 IT, L4/5 IT_2 and L5 IT types which had tufted apical dendrites (**Fig. 8f**). We also obtained
791   full morphological reconstructions of excitatory neurons with their somas located in L2, L3 or
792   L4 in MOp or the neighboring secondary motor area (MOs) from fMOST imaging of *Cux2-*
793   *CreERT2;Ai166* mice [68,80]. As shown in **Fig. 8b**, *Cux2* is a specific marker gene for L2/3 IT and
794   L4/5 IT_1 types. These full reconstructions allowed us to examine, in addition to dendritic
795   morphologies, the full extent of both local and long-range axon projections. The MOp/MOs
796   neurons with somas in putative L4 (between L2/3 and L5) exhibited two local morphological
797   features characteristic of L4 neurons found in sensory cortical areas (**Fig. 8g**). First, the dendrites
798   of the L4 neurons were simple and untufted whereas those of the L2 and L3 neurons all had
799   extensive tufts. Second, the local axons of L4 neurons mostly projected upward into L2/3 in
800   addition to collateral projections; on the contrary, the local axons of L2 and L3 neurons mostly
801   projected downward, reaching into L5. These local projection patterns are consistent with the
802   canonical feedforward projections within a cortical column observed in somatosensory and
803   visual cortices, with the first feedforward step being from L4 to L2/3 and the second feedforward
804   step from L2/3 to L5 [96]. We also found that the MOp/MOs L4 neurons had intracortical long-
805   range projections like the L2 and L3 neurons (**Fig. 6d**).
806
807   Taken together, our multimodal characterization demonstrates that mouse MOp indeed has
808   excitatory neurons with L4 characteristics, namely, occupying a specific layer between L2/3 and
809   L5, having simple and untufted dendrites and upward-projecting local axons, belonging to a
810   transcriptomic type (L4/5 IT_1) marked by a L4-specific gene *Rspo1* as well as the intersection
811   of a L2/3/4-specific gene *Cux2* and a L4/5-specific gene *Rorb*, and having corresponding
812   epigenomic regulatory elements. L4-like neurons may also exist in human and marmoset M1.
813
814
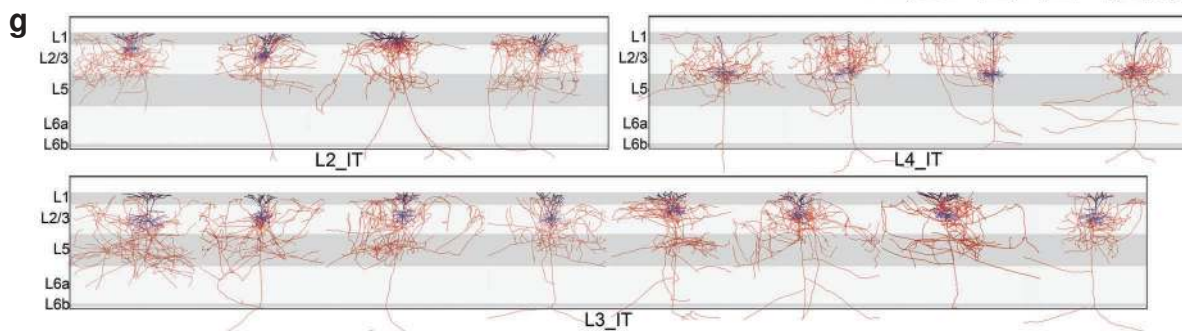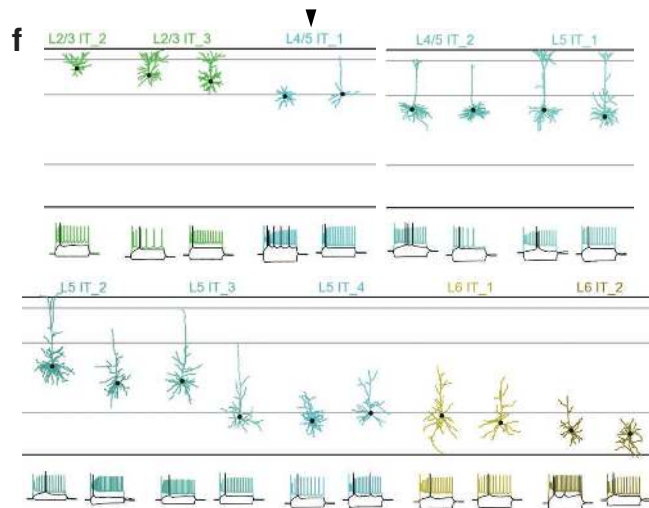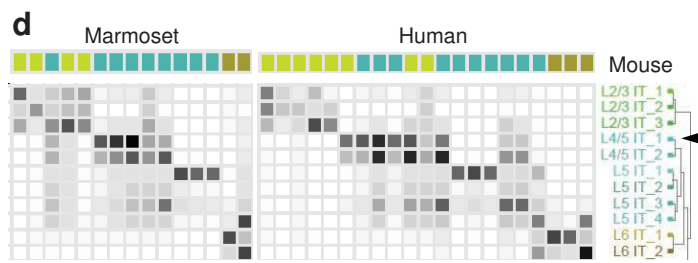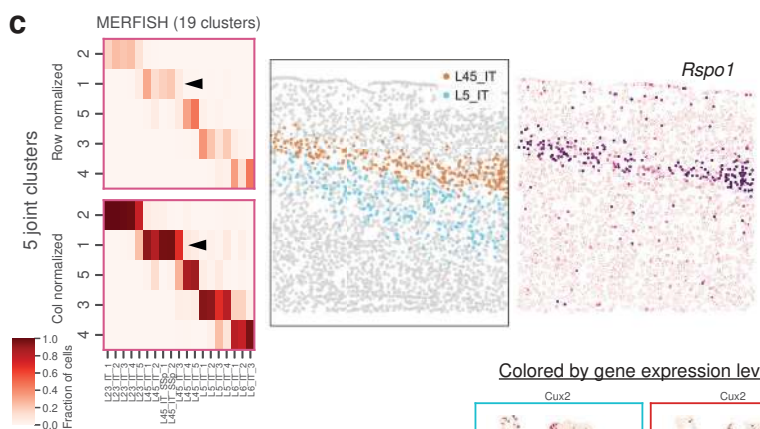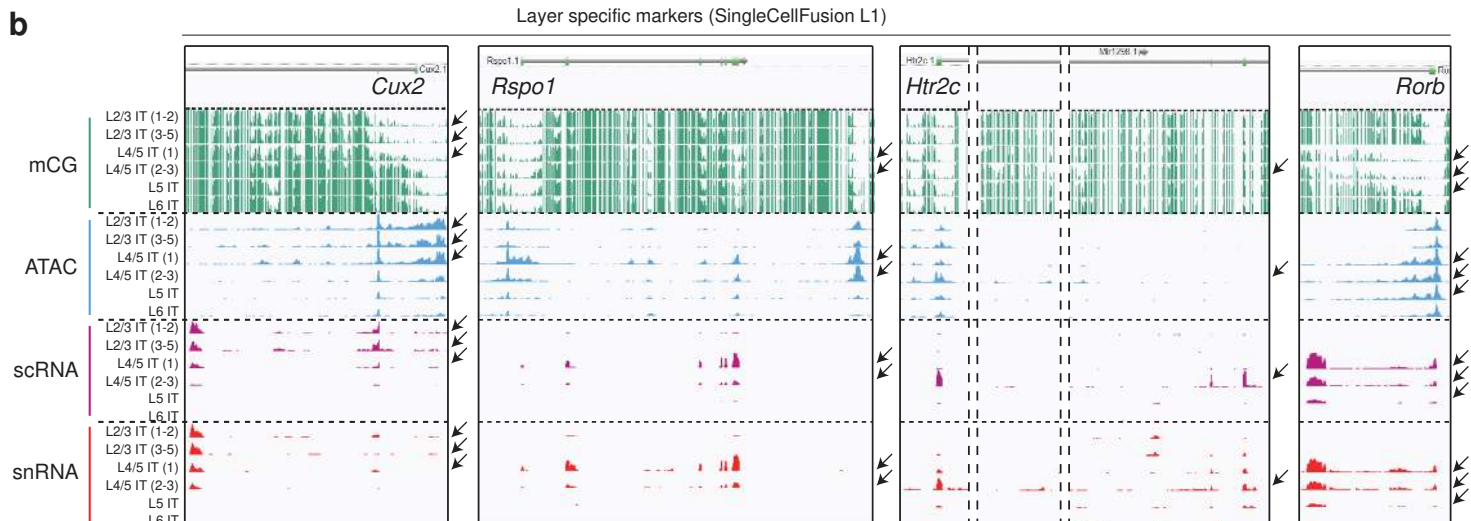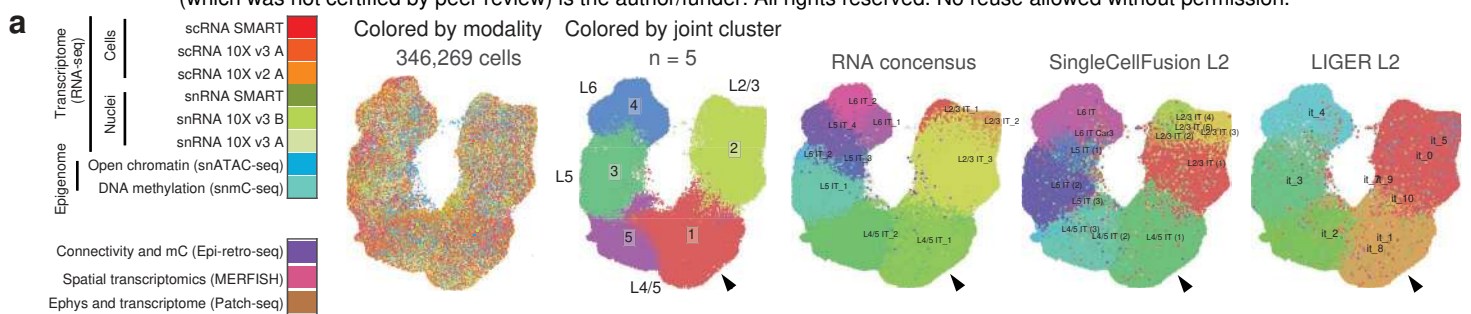
815 **Figure 8. Existence of L4 excitatory neurons in MOp. a**, UMAP embedding of IT cells from
816 11 datasets. Cells are colored by modalities, by cluster identities from the 11-dataset joint
817 clustering, and by cluster identities generated from other consensus clustering methods in [45]. **b**,
818 Genome browser of layer-specific gene markers - from L2/3 to L5 - across IT cell types as
819 defined in [45] (SingleCellFusion L1). Arrows indicate cell types with correlated transcription and
820 epigenomic signatures of the specific marker gene. **c**, MERFISH IT clusters correspond well
821 with the above joint clusters from **a** (confusion matrices, left panel) and reveal a L4 specific
822 cluster (L45_IT) distinctly separated from the L5_IT cluster (middle panel) and marked by gene
823 *Rspo1* (right panel). **d**, Correspondence between mouse and human or marmoset transcriptomic
824 IT types. **e**, UMAP embedding of excitatory cells from MOp (scRNA_SMART) [45] and VISp [15]
825 show that L4 excitatory cells from MOp correspond to a subset of L4 excitatory cells from VISp.
826 Cells are colored by brain regions (MOp, red; VISp, blue), by cell types, and by expression
827 levels (log10(TPM+1)) of marker genes *Cux2* and *Rorb*. **f**, Dendritic morphologies and spiking
828 patterns of representative Patch-seq cells from all IT types (L2/3 to L6). Arrow heads in a, c, d
829 and f indicate the L4/5 IT 1 type. **g**, Local dendritic and axonal morphologies of fully
830 reconstructed IT neurons with somas located in L2, L3 and L4. Black, apical dendrites. Blue,
831 basal dendrites. Red, axons.
832
833
834 **Integrated multimodal characterization of two L5 ET projection neuron types in MOp**
835 Previous studies had shown that in the mouse anterolateral motor (ALM) cortex, part of MOs, L5
836 ET neurons have two transcriptomically distinct projection types that may carry out different
837 motor-control functions; the thalamus projecting type may be involved in movement planning
838 whereas the medulla (MY) projecting type may be involved in initiation of the movement [15,86].
839 Here as the second case study, through integrated multimodal characterization we demonstrate
840 that L5 ET neurons in MOp can also be divided into MY-projecting and non-MY-projecting
841 types.
842
843 As shown in the companion paper [45], we mapped the mouse MOp L5 ET transcriptomic types to
844 the previous VISp-ALM transcriptomic taxonomy [15]. From this mapping we found that the MOp
845 L5 ET 1 type corresponded to the ALM MY-projecting type marked by *Slco2a1*, whereas MOp
846 L5 ET 2-4 types corresponded to the ALM thalamus-projecting types with L5 ET 2/3 marked by
847 *Hpgd* and L5 ET 4 by *Npsr1*. Here we show such distinction is consistent across all molecular
848 datasets (**Fig. 9a-b**). Mouse transcriptomic type L5 ET 1 corresponded well with both integrated
849 molecular type SCF L5 ET (1) and MERFISH clusters L5_ET_5, as well as with a L5 ET
850 transcriptomic type from human and marmoset. Mouse transcriptomic types L5 ET 2-4
851 corresponded with integrated molecular types SCF L5 ET (2-3), MERFISH clusters L5_ET_1-4,
852 and two L5 ET transcriptomic types from human and marmoset. The laminar distribution of
853 these two groups was revealed by MERFISH, with cells in L5_ET_1-4 clusters intermingled in
854 the upper part of L5 and cells in L5_ET_5 located distinctly in lower L5 (**Fig. 9c**). The two
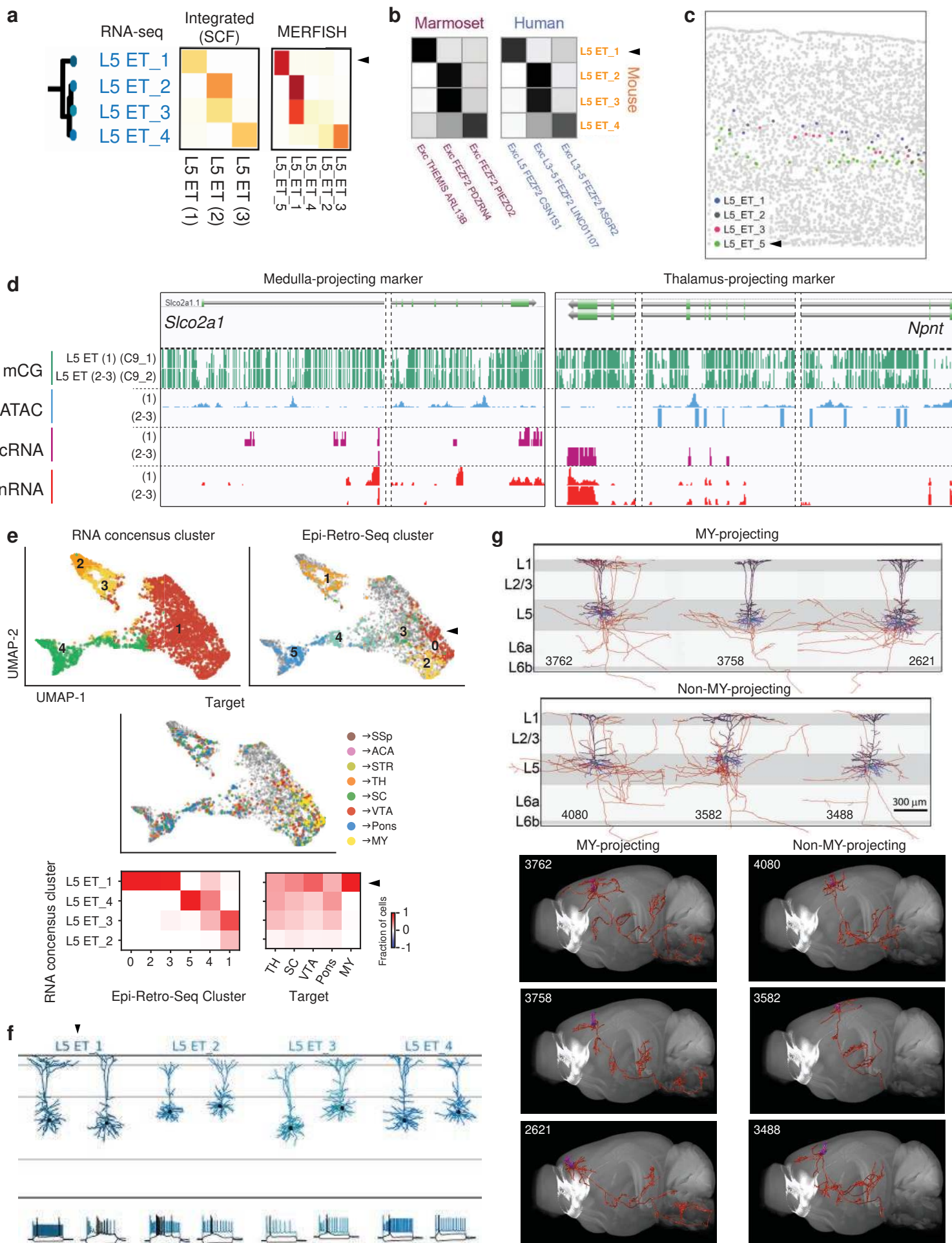
855 groups were further distinguished by epigenomic peaks associated with specific marker genes,
856 *Slco2a1* for SCF L5 ET (1) type and *Npnt* for SCF L5 ET (2-3) types (**Fig. 9d**), providing
857 validity to the two novel transgenic driver lines we generated, *Slco2a1-P2A-Cre* and *Npnt-P2A-*
858 *FlpO* (**Fig. 7**).
859
860 Epi-Retro-Seq study (see above) revealed more complex long-range projection patterns among
861 the 6 epigenetic L5-ET clusters identified, with cluster 0 predominantly projecting to MY while
862 other clusters having variable and less specific projection patterns (clusters 2 and 3 also
863 containing MY-projecting cells) (**Fig. 5g**). We co-clustered the L5 ET cells from the Epi-Retro-
864 Seq data and the snRNA-seq 10x v3 B data [45] to investigate the correspondence of Epi-Retro-Seq
865 clusters and projection targets with transcriptomic clusters (**Fig. 9e**). We found that the
866 consensus transcriptomic cluster L5 ET 1 corresponds to Epi-Retro-Seq clusters 0, 2 and 3, all of
867 which contain MY-projecting neurons. On the other hand, transcriptomic clusters L5 ET 2-4
868 correspond to Epi-Retro-Seq clusters 1, 4 and 5, which do not contain MY-projecting neurons.
869 Thus, all MY-projecting neurons are mapped to transcriptomic type L5 ET 1, while neurons in
870 the L5 ET 2-4 types do not project to MY.
871
872 Anterograde tracing in *Slco2a1-P2A-Cre* mice demonstrated predominant projection from
873 *Slco2a1*-labeled neurons in MOp to MY (**Fig. 7**). We identified multiple full morphology
874 reconstructions of MOp L5 ET neurons from fMOST imaging of *Fezf2-CreER;Ai166* and *Pvalb-*
875 *T2A-CreERT2;Ai166* transgenic mice [68]. These reconstructions could be clearly separated into a
876 MY-projecting group and a non-MY-projecting group (**Fig. 9g**), though they were not directly
877 linked to transcriptomic types yet. Both groups of cells had thick-tufted dendrites that were
878 similar to each other (**Fig. 9g**). Consistent with this, Patch-seq cells corresponding to
879 transcriptomic types L5 ET 1-4 also were indistinguishable from each other by their dendritic
880 morphologies (**Fig. 9f**).
881
882 Altogether, our integrated multimodal characterization identified two major types of mouse L5
883 ET projection neurons, MY-projecting and non-MY-projecting, with distinct gene markers,
884 epigenomic elements, laminar distribution, and corresponding types in human and marmoset.
885
886

a RNA-seq, Integrated (SCF), MERFISH — L5 ET_1, L5 ET_2, L5 ET_3, L5 ET_4

b Marmoset, Human, Mouse — L5 ET_1, L5 ET_2, L5 ET_3, L5 ET_4

c L5_ET_1, L5_ET_2, L5_ET_3, L5_ET_5

d Medulla-projecting marker (Slco2a1); Thalamus-projecting marker (Npnt)
mCG: L5 ET (1) (C9_1), L5 ET (2-3) (C9_2); ATAC (1), (2-3); scRNA (1), (2-3); snRNA (1), (2-3)

e RNA concensus cluster; Epi-Retro-Seq cluster; Target (→SSp, →ACA, →STR, →TH, →SC, →VTA, →Pons, →MY); Fraction of cells

f L5 ET 1, L5 ET 2, L5 ET 3, L5 ET 4

g MY-projecting (3762, 3758, 2621); Non-MY-projecting (4080, 3582, 3488)

A multimodal cell census and atlas of the mammalian primary motor cortex

887 **Figure 9. Two distinct L5 ET projection neuron types in MOp. a**, Within the mouse L5 ET
888 subclass, good correspondence is observed between the 4 transcriptomic clusters and the 3
889 integrated molecular clusters (SingleCellFusion) or the 5 MERFISH clusters. **b**, Within the L5
890 ET subclass, the 4 mouse transcriptomic clusters correspond well with the 3 transcriptomic
891 clusters in either human or marmoset. **c**, In MERFISH, cells belonging to the L5_ET_1-4
892 clusters co-occupy the upper L5, whereas L5_ET_5 cells are distinctly located in lower L5. **d**,
893 Genome browser of gene markers between the MY-projecting (*Slco2a1*) and the non-MY-
894 projecting (*Npnt*) L5 ET neurons. **e**, Integration panels between L5 ET Epi-Retro-Seq clusters
895 and consensus transcriptomic clusters. The transcriptomic dataset used here is snRNA 10x v3 B,
896 which has the largest number of L5 ET cells (>4k). Top panels, UMAP plots colored by
897 consensus transcriptomic clusters, Epi-Retro-Seq clusters and projection targets (retrograde
898 tracer injection sites). Bottom panels, confusion matrices between consensus transcriptomic
899 clusters and Epi-Retro-Seq clusters or major projection targets. The heatmaps are column-wise
900 normalized rather than row-wise to avoid misleading interpretation, since the number of cells
901 sampled from each projection may differ a lot in Epi-Retro-Seq. **f**, Dendritic morphologies and
902 spiking patterns of representative Patch-seq cells corresponding to the 4 mouse transcriptomic L5
903 ET types. **g**, Local dendritic and axonal morphologies (upper panels) and brain-wide axon
904 projections (lower panels) of representative fully-reconstructed L5 ET neurons, separated into
905 MY-projecting and non-MY-projecting types. Black, apical dendrites. Blue, basal dendrites. Red,
906 axons.
907
908
909 **An integrated synthesis of multimodal features of cell types in the primary motor cortex**
910 As the conclusion of this series of studies from BICCN, we present an overview and integrated
911 synthesis of the knowledge gained in constructing a multimodal census and atlas of cell types in
912 the primary motor cortex of mouse, marmoset and human (**Fig. 10**). A critical aspect of our
913 studies is that this synthesis is only made possible by the systematic integrative computational
914 analyses across multiple transcriptomic and epigenomic data types that connect a diverse range
915 of cellular features together at cell subclass or type level to allow mutual correlation.
916
917 This integrated synthesis uses the mouse MOp consensus transcriptomic taxonomy (containing
918 18 subclasses and 116 clusters/types) [45] as the anchor (**Fig. 10**) because it was derived from the
919 largest and deepest datasets and was the reference taxonomy for nearly all the cross-modality and
920 cross-species comparisons. Correspondence matrices between the different molecular modalities
921 show that the mouse MERFISH-based spatial transcriptomic taxonomy (95 clusters) [54], the
922 integrated mouse molecular taxonomies combining transcriptomic and epigenomic data using
923 either SingleCellFusion (SCF, 56 neuronal clusters) or LIGER (71 clusters) approach [45], and the
924 human and marmoset transcriptomic taxonomies (127 and 94 clusters, respectively) [48] all aligned
925 largely consistently with the mouse consensus transcriptomic taxonomy. Such alignment
926 convincingly demonstrates that cell types in a given brain region can be consistently described

A multimodal cell census and atlas of the mammalian primary motor cortex

927   by different types of characterization. At the same time, it should also be noted that the
928   alignments are not perfect and disagreements do exist at the individual cluster level (which is
929   most pronounced in cross-species comparisons), suggesting that differential variations exist in
930   different data types and consistency, in particular that across species, may be more appropriately
931   described at an intermediate level of granularity.
932
933   In this integrated synthesis, we can further assign additional attributes to the molecularly defined
934   cell types (**Fig. 10**). Based on Patch-seq [64], Retro-seq (e.g. Epi-Retro-Seq [79]), Retro-MERFISH
935   [54], and axon projection [68,69] studies, we relate many transcriptomic neuronal types or subclasses
936   to cortical neuron types traditionally defined by electrophysiological, morphological and
937   connectional properties, thus bridging our cell type census with historical and community
938   knowledge. We provide the relative proportion of each cell type within the mouse MOp using
939   either snRNA-seq or MERFISH data. The MERFISH data also identify the spatial distribution
940   pattern of each cell type [54]. For example, we found that excitatory or inhibitory neuron types are
941   distributed along the cortical depth, with many individual types adopting narrow cortical-depth
942   distributions, often occupying predominantly a single layer or a sublayer, and related types (e.g.
943   the L2/3-6 IT excitatory types) can display a gradual transitioning across cortical depths/layers.
944   On the other hand, non-neuronal cell types are either distributed across all layers or specific to
945   layer 1 or the white matter (WM). Patch-seq data also provided the cortical depth positions of a
946   variety of neuronal cell types.
947
948   Finally, we demonstrate the possibility to elucidate gene regulatory mechanisms by discovering
949   candidate cis-regulatory elements (cCREs) as well as master transcription factors (TFs) specific
950   to neuronal subclasses by mining the combined mouse MOp transcriptomic and epigenomic
951   datasets to access both RNA expression of genes and accessibility or DNA methylation of
952   cCREs from the same cell clusters.
953
954   For example, we found 7,245 distal cCRE (>1 kbp from transcriptional start site)-gene pairs in
955   neuronal cells in MOp that showed positive correlation between accessibility at the 6,280 cCREs
956   and expression levels of 2,490 putative target genes (see Methods, and companion papers [45,52]).
957   We grouped these putative enhancers into modules based on accessibility across cell clusters
958   (**Extended Data Fig. 2**). 76% of putative enhancers showed remarkable sub-type specific
959   chromatin accessibility and were enriched for lineage-specific transcription factors, while 24% of
960   putative enhancers (1,527) were widely accessible and linked to genes expressed across neuronal
961   cell clusters with highest expression levels in Glutamatergic neurons (module M1, **Extended
962   Data Fig. 2b**). Putative enhancers in this module showed enrichment of sequence motifs
963   recognized by transcription factors CTCF, MEF2 indicating a more general rule of these factors
964   in establishing neuronal gene regulatory programs (**Extended Data Fig. 2c**). Meanwhile, other
965   modules (M2 to M14) of enhancer-gene pairs were active in a subclass-specific manner
966   (**Extended Data Fig. 2b-d**). Thus, using this approach we have identified a large number of

967 enhancer-gene pairs for each subclass of neurons (**Fig. 10**). These enhancers can be potentially
968 used to generate cell type-targeting viral tools.
969
970 Similarly, we identified transcription factors showing cell-type specificity supported by both
971 RNA expression and DNA binding motif enrichment in hypo-CG-DMR of MOp subclasses (see
972 Methods, and companion papers [45,50]) (**Extended Data Fig. 3**). Combining these two
973 orthologous pieces of evidence identified many well-studied TFs in embryonic precursors, such
974 as the Dlx family members for pan-inhibitory neurons, and Lhx6 and Mafb for MGE derived
975 inhibitory neurons. We further identified many additional TFs with more restricted patterns in
976 specific subclasses, such as Rfx3 and Rreb1 (in L2/3 IT), Atoh7 and Rorb (in L4/5 IT), Pou3
977 family members (in L5 ET), Etv1 (in L5/6 NP), Esrr family members (in Pvalb), and Arid5a (in
978 Lamp5). The agreement of these two modalities suggests a requirement of TF regulatory activity
979 in mature neurons to maintain aspects of cell phenotypes and identity.
980
981 In summary, our comprehensive multimodal characterization of cell types from the MOp region
982 demonstrates that transcriptomic, epigenomic, spatial, physiological, morphological and
983 connectional properties can be all correlated and integrated together, to reveal organizational
984 principles of brain cell types and bridge molecular, structural and functional studies in different
985 modalities and across species.
986
987
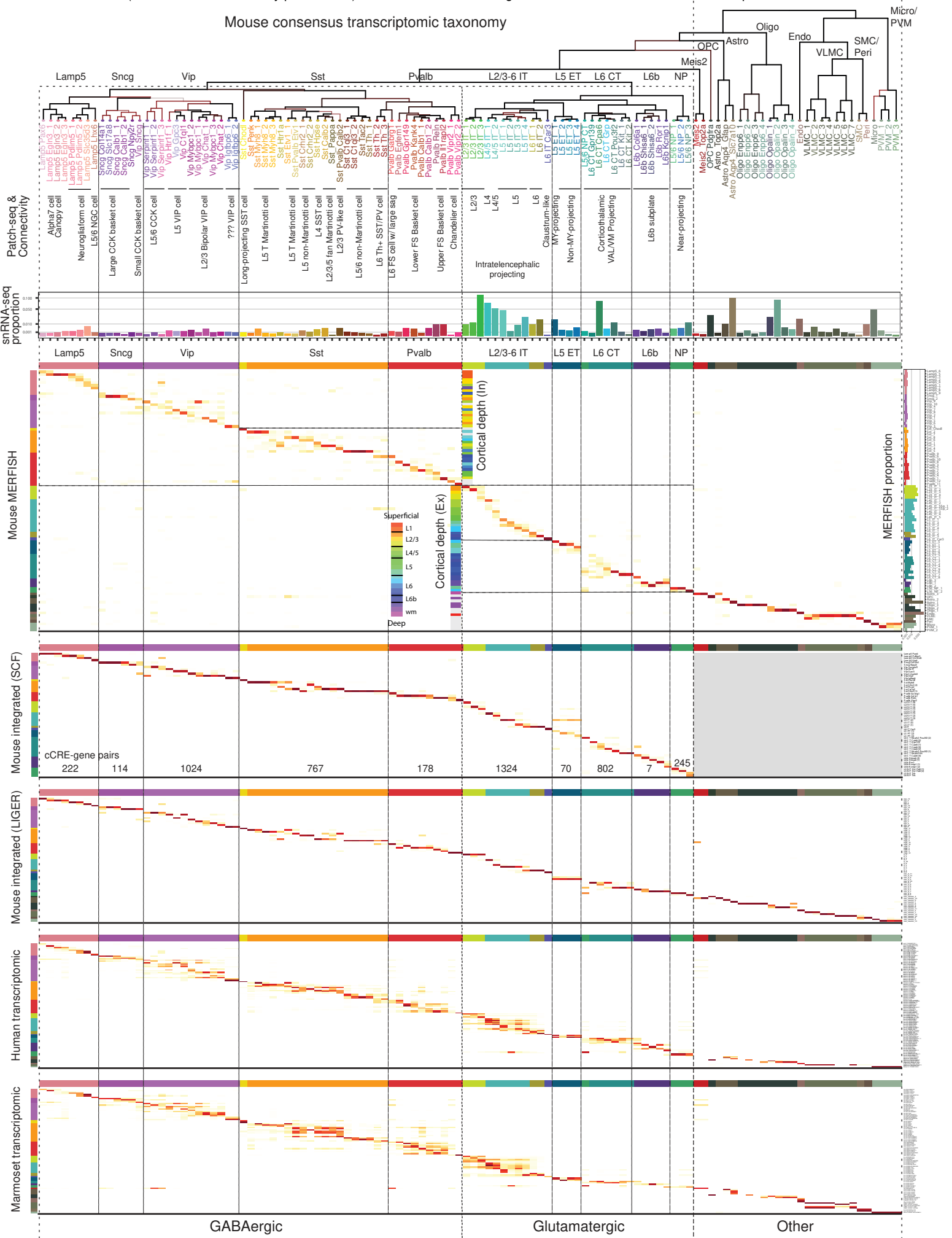
A multimodal cell census and atlas of the mammalian primary motor cortex

988 **Figure 10. An integrated multimodal census and atlas of cell types in the primary motor**
989 **cortex of mouse, marmoset and human.** The mouse MOp consensus transcriptomic taxonomy
990 at the top is used to anchor cell type features in all the other modalities. Subclass labels are
991 shown above major branches and cluster labels are shown below each leaf node. Confusion
992 matrices show the correspondence between the mouse MOp transcriptomic taxonomy with those
993 derived from other molecular datasets, including mouse MERFISH, the integrated mouse
994 molecular taxonomies by SingleCellFusion (SCF) or LIGER, and the human and marmoset
995 transcriptomic taxonomies. Using Patch-seq and connectivity studies, many transcriptomic
996 neuronal types or subclasses are annotated and correlated with known cortical neuron types
997 traditionally defined by electrophysiological, morphological and connectional properties. (Note:
998 no Patch-seq data were collected for the Vip cells labeled by question marks.) The relative
999 proportions of all cell types within the mouse MOp are calculated from either the snRNA-seq
1000 10x v3 B (horizontal bar graph) or MERFISH (vertical bar graph to the right of the MERFISH
1001 matrix) dataset. Median cortical depth position of each cell type derived from MERFISH is
1002 shown as color-coded bar graphs at the center of the MERFISH matrix, colored according to the
1003 rainbow scheme from superficial (red) to deep (purple) layers as shown on the left. Cell types
1004 with dispersed distributions spanning relatively large ranges of cortical depth are colored in grey.
1005 The numbers of cCRE-gene pairs in modules corresponding to neuronal subclasses identified by
1006 Cicero from the scRNA-seq and snATAC-seq datasets are shown at the bottom of the SCF
1007 matrix.
1008
1009
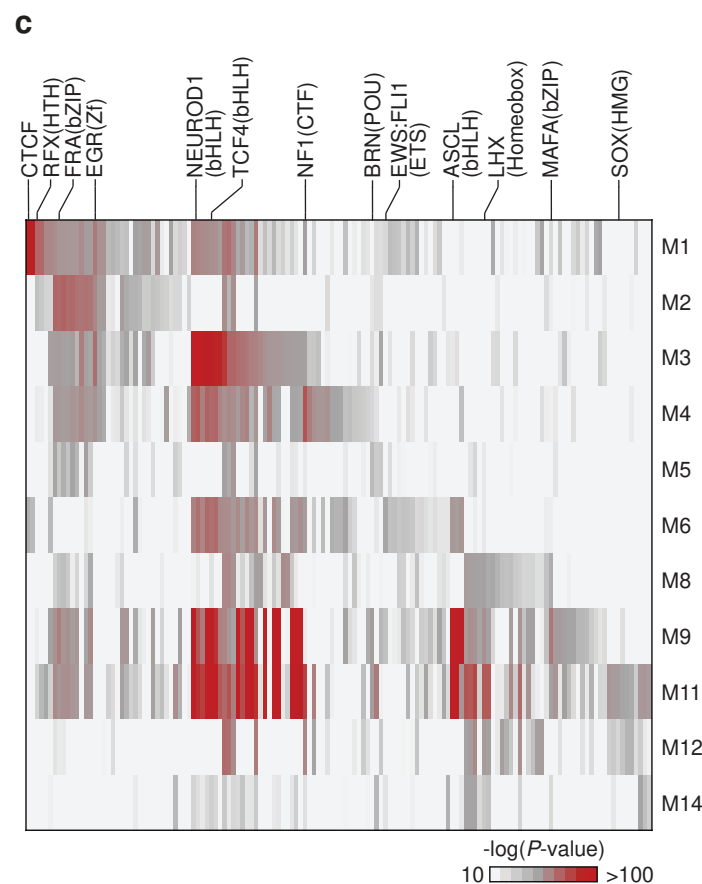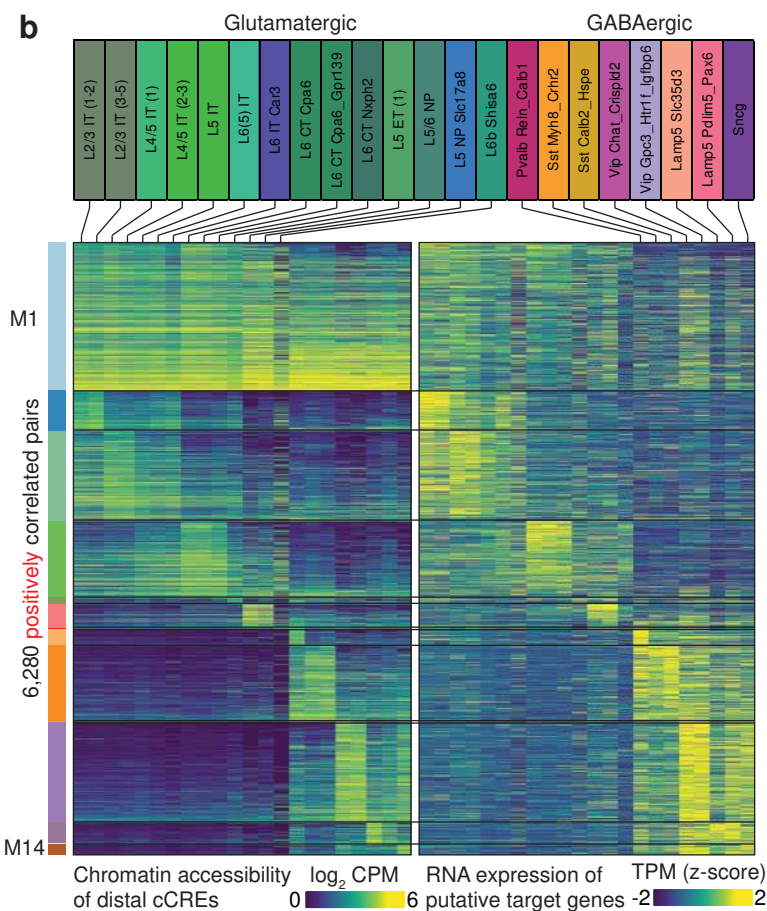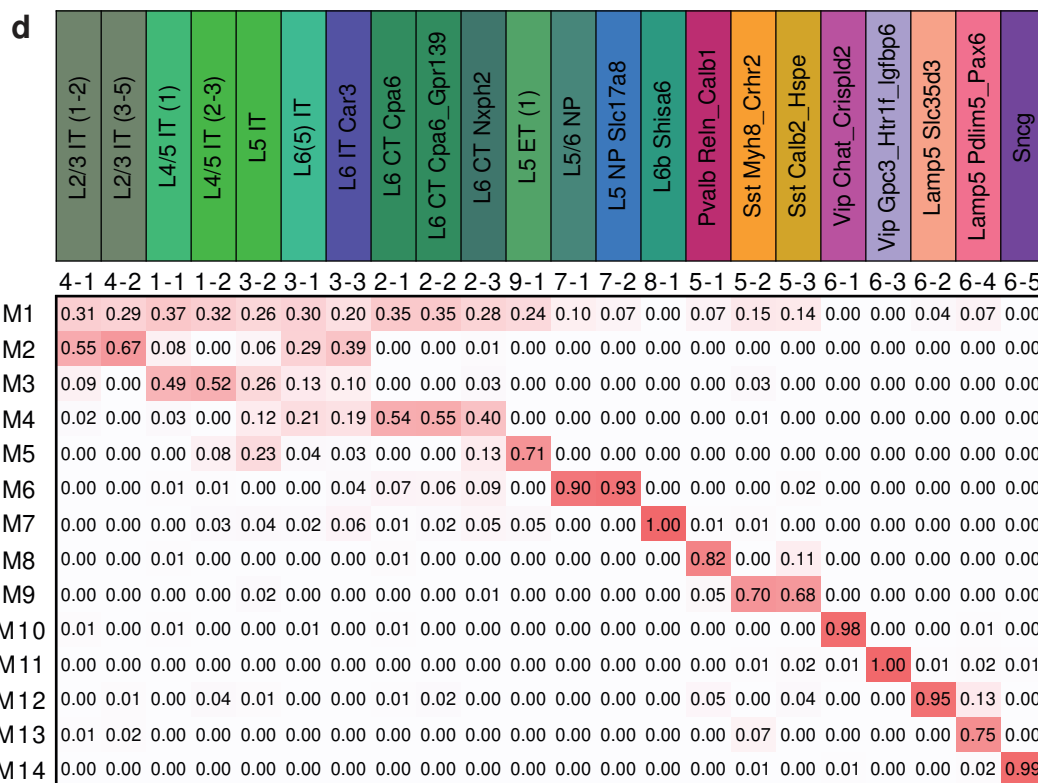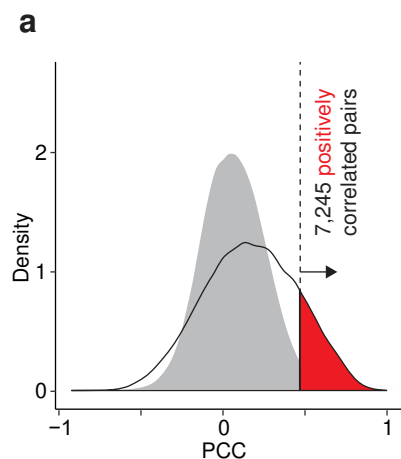
A multimodal cell census and atlas of the mammalian primary motor cortex

1010   **Extended Data Figure 2.** Characterization of putative enhancer-gene pairs. **a,** Detection of
1011   putative enhancer-gene pairs. 7,245 pairs of positively correlated cCRE and genes (highlighted in
1012   red) were identified using an empirically defined significance threshold of FDR<0.01. Grey
1013   filled curve shows the distribution of PCC for randomly shuffled cCRE-gene pairs. **b,** Heatmap
1014   of chromatin accessibility of 6,280 putative enhancers across joint cell clusters (left) and
1015   expression of 2,490 target genes (right). Note genes are displayed for each putative enhancer
1016   separately. CPM: counts per million, TPM: transcripts per million. **c,** Enrichment of known
1017   transcription factor motifs in distinct enhancer-gene modules. Displayed are known motifs from
1018   HOMER with enrichment -log p-value >5. In module M1, de novo motif analysis of putative
1019   enhancers in this module showed enrichment of sequence motif recognized by transcription
1020   factors CTCF, MEF2. CTCF is a widely expressed DNA binding protein with a well-established
1021   role in transcriptional insulation and chromatin organization, but recently it was also reported
1022   that CTCF can promote neurogenesis by binding to promoters and enhancers of related genes. In
1023   the L2/3 IT selective module M2, the putative enhancers were enriched for the binding motif for
1024   Zinc-finger transcription factor EGR, a known master transcriptional regulator of excitatory
1025   neurons [97]. In the Pvalb selective module M8, the putative enhancers were enriched for sequence
1026   motifs recognized by the MADS factor MEF2, which is associated with regulating cortical
1027   inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders [98]. **d,**
1028   Heatmap shows the weights of each joint cell cluster in each module, which were derived from
1029   the coefficient matrix. The values of each column are scaled (0–1).
1030
1031
1032   **Extended Data Figure 3.** Dot plot illustrating the RNA expression levels (red) and hypo-CG-
1033   DMR motif enrichments (blue) of transcription factors (TFs) in mouse MOp subclasses. The size
1034   and color of red dots indicate the proportion of expressing cells and the average expression level
1035   in each subclass, respectively. The size and color of blue dots indicate adjusted P-value and
1036   log2(Odds Ratio) of motif enrichment analysis, respectively.
1037
1038

Transcription Factors RNA Expression And
Motif Enrichment in MOp Subclasses
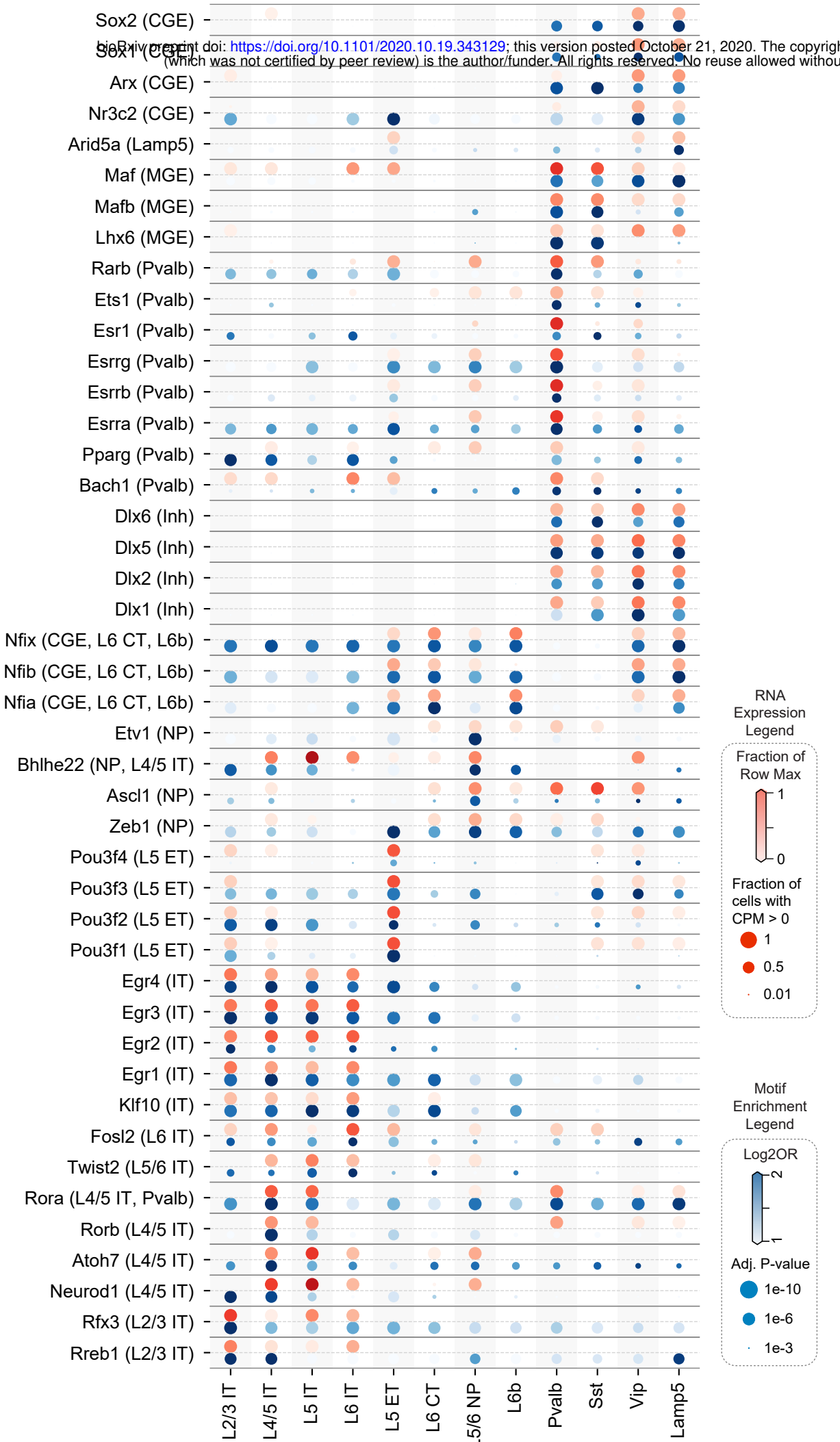
1039 **DISCUSSION**

1040

1041 *A cell census and atlas of motor cortex*
1042 Understanding the principles of brain circuit organization requires a detailed understanding of its
1043 basic components, but the cellular diversity and complexity of the brain, including the neocortex,
1044 have defied a comprehensive and quantitative description. Single-cell transcriptomics and
1045 epigenomics, as well as spatially resolved single-cell transcriptomics, are accelerating efforts to
1046 classify all molecular cell types in many organ systems [40,99], including the brain [5,6,100]. The
1047 current effort combines these technologies to derive a robust and comprehensive molecular cell
1048 type classification and census of the primary motor cortex of mouse, marmoset and human,
1049 coupled with a spatial atlas of cell types and an anatomical input/output wiring diagram in
1050 mouse. We demonstrate the robustness and validity of this classification through strong
1051 correlations across cellular phenotypes, and strong conservation across species. Together these
1052 data comprise a cell atlas of primary motor cortex that encompasses a comprehensive reference
1053 catalog of cell types, their proportions, spatial distributions and anatomical and physiological
1054 characteristics, and molecular genetic profiles, registered into a Common Coordinate Framework
1055 [41]. This cell atlas establishes a foundation for an integrative study of the architecture and function
1056 of cortical circuits akin to reference genomes for studying gene function and genome regulatory
1057 architecture. Furthermore, it provides a comprehensive map of the genes that contribute to
1058 cellular phenotypes and their epigenetic regulation. These data resources and associated tools
1059 enabling genetic access for manipulative experimentation are publicly available and provide a
1060 roadmap for exploring cellular diversity and organization across brain regions, organ systems,
1061 and species.

1062

1063 The molecular classification presented here is overall consistent with prior literature and
1064 synthesizes a wide body of existing and new information into a coherent quantitative framework
1065 that provides metrics for the robustness of, and the similarities and distinctions between, cell
1066 types. For motor cortex, as for other cortical regions [15,18], this cellular organization is
1067 hierarchical, with different branches comprising major cell classes, subclasses, and types
1068 representing the finest resolution clusters afforded by each method. This classification provides
1069 strong evidence for the existence of hitherto poorly studied but molecularly distinct subclasses
1070 such as the near-projecting (NP) pyramidal neurons, and many more novel cell types. At the
1071 level of cell class and subclass (and some highly distinctive types like chandelier cells and long-
1072 range projecting Sst Chodl interneurons), we find remarkable concordance across
1073 transcriptomics, epigenomics, spatial patterning, physiology and connectivity, as well as strong
1074 homology across species. The class and subclass branches clearly represent different
1075 developmental programs, such as GABAergic neuron derivatives of different zones of the
1076 ganglionic eminences [101,102] or the layer-selective glutamatergic neurons derived sequentially
1077 from progenitors of the cortical plate, and the hierarchical organization generates new
1078 hypotheses about developmental origins of highly distinctive cell types. This quantitative

A multimodal cell census and atlas of the mammalian primary motor cortex

1079   hierarchy also challenges well-established nomenclature systems. For example, the term glia is
1080   typically used to encapsulate astrocytes, oligodendrocytes and OPCs, and microglia. However,
1081   microglia are not closely related to these neuroectoderm-derived populations based on
1082   transcriptomics or developmental origins [103] and should be grouped with other more similar non-
1083   neuronal cell types such as endothelial cells, VLMCs and pericytes. Substantial challenges
1084   remain for redefining data-driven cell ontologies and nomenclature systems [100,104].

1085

1086   Comparisons of the MOp results described here to other regions also help to understand what
1087   makes the motor cortex functionally distinct. Previous transcriptomic studies suggested that
1088   GABAergic interneuron types are shared among cortical regions whereas glutamatergic
1089   projection neuron types exhibit gradient-like distribution across the cortical sheet and are more
1090   distinct between distant regions but more similar between neighboring regions [15,44]. Thus the
1091   projection neurons in MOp are more similar to those of nearby regions, yet our anatomical
1092   tracing study defines a MOp-specific input-output wiring diagram. This result suggests that
1093   differential axonal projections of similar molecular types among different cortical areas may be
1094   the major feature defining regional functional specificity. We also find substantial variation in
1095   the proportion of specific cell types between cortical areas. For example, we identify two
1096   glutamatergic neuron types that distinguish MOp from its neighboring primary somatosensory
1097   (SSp) region, the L4 IT neurons that are present in MOp at lower abundance level than in SSp
1098   and the *Slco2a1*-expressing, medulla-projecting L5 ET neurons that are more abundant in MOp
1099   than in SSp [54,68]. These regional differences in cellular makeup may contribute to the functional
1100   specialization of MOp as well.

1101

1102   ***Cell type discreteness, variation and phenotypic concordance***
1103   The concordance of transcriptomic and epigenomic results and their overall correlation with
1104   other cellular phenotypes, including spatial distributions, morphological properties,
1105   electrophysiological properties, and projection/connectivity, strongly argues for a unifying
1106   molecular genetic framework for understanding cortical cell types, particularly at the level of
1107   subclasses and distinctive cell types. At the same time, substantial multimodal variations at finer
1108   granularity appear to preclude a fully discretized representation of cell types with consistency
1109   across all cellular phenotypes. One source of variation is differences in granularity with different
1110   molecular data modalities, with transcriptomics providing the highest granularity at present.
1111   This may reflect true biology or differences in technological information content, for example
1112   sparse genome coverage in epigenetic methods. A second source involves continuous rather than
1113   discrete variation. For example, while some highly granular cell types are highly distinct from
1114   others (e.g. L6 IT Car3, Sst Chodl and Pvalb chandelier cells), many other types exhibit
1115   continuous variation in their properties both within types and among closely related types with
1116   no clear boundaries between them. However, even at this fine-grained level of continuous
1117   variation, spatial, morphological and physiological properties often co-vary with transcriptomic
1118   profiles as shown by MERFISH and Patch-seq. Similar findings on continuous as well as unitary

1119 variations have been reported for hippocampal interneurons [16]. These results suggest that
1120 continuous phenotypic variation may represent a general organizing principle underlying the
1121 diversification of brain cell types.
1122
1123 As shown in our mouse Epi-Retro-Seq, MERFISH, and single-neuron full morphology and
1124 projection studies there is a strong correlation between molecular phenotype and axonal target
1125 specificity at the subclass level (e.g., IT, L5 ET, L6 CT). This was also the case for medulla-
1126 projecting L5 ET type. However, a strict correlation between molecular cell types and specific
1127 axonal projection targets was not generally observed. It is possible that axon pathfinding during
1128 development involves stochastic decisions and subsequent activity-dependent pruning that
1129 mature cell transcriptomes do not represent. Furthermore, individual projection neurons typically
1130 have collaterals to many different target regions which complicates understanding these
1131 relationships. Comprehensive datasets on the complete axonal projections of individual neurons
1132 whose molecular identity is clearly established will be needed to address this issue.
1133
1134 ***Cell type conservation and divergence***
1135 Evolutionary conservation is strong evidence of functional significance. The demonstrated
1136 conservation of cell types from mouse, marmoset, macaque and human strongly suggests that
1137 these conserved types play important roles in cortical circuitry and contribute to a common
1138 blueprint essential for cortical function in mammals and even more distantly related species. We
1139 also find that similarity of cell types varies as a function of evolutionary distance, with
1140 substantial species differences that either represent adaptive specializations or genetic drift. For
1141 the most part species specializations tend to appear at the finer branches or leaves of the
1142 hierarchical taxonomy. This result is consistent with a recent hypothesis in which cell types are
1143 defined by common evolutionary descent and evolve independently, such that new cell types are
1144 generally derived from existing genetic programs and appear as specializations at the finer levels
1145 of the taxonomic tree [105].
1146
1147 A surprising finding across all homologous cell types was the relatively high degree of
1148 divergence for genes with highly cell type-specific expression in a given species. This
1149 observation provides a clear path to identify the core conserved genes underlying the canonical
1150 identity and features of those cell types. Furthermore, it highlights the need to understand species
1151 adaptations superimposed on the conserved program, as many specific cellular phenotypes may
1152 vary across species including gene expression, epigenetic regulation, morphology and
1153 connectivity, and physiological functional properties. As we illustrate in the Betz cells, there is
1154 clear homology across species in the layer 5 ET subclass, but variation in many measurable
1155 properties across species.
1156
1157 ***A framework for linking model organisms to human biology and disease***

1158   The results presented have major utility and implications for the consideration of model
1159   organisms to understand human brain function and disease. Despite major investments, animal
1160   models of neuropsychiatric disorders have often been characterized by "loss of translation,"
1161   fueling heated debates about the utility of model organisms in the search for therapeutic targets
1162   for treating human diseases. The molecular genetic framework of cell type organization
1163   established by the current study will provide a robust cellular metric system for cross-species
1164   translation of knowledge and insight that bridges levels of organization based on their inherent
1165   biological and evolutionary relationships. For example, the characterization of cell types and
1166   their properties shown in **Figure 10** can be used to infer the main characteristics of homologous
1167   cell types in humans and other mammalian species, despite the often extreme difficulty in
1168   measuring their specific properties in those species. On the other hand, they also reveal the
1169   potential limitations of model organisms and the necessity to study human and closely related
1170   primate species to understand the specific features of cell types as they contribute to human brain
1171   function and susceptibility to human-specific diseases. Having cell census information aligned
1172   across species as illustrated here should be highly valuable for making rational choices about the
1173   best models for each disease and therapeutic target. This reductionist dissection of the cellular
1174   components provides a foundation for understanding the general principles of neural circuit
1175   organization and computation that underlie mental activities and brain disorders.
1176
1177   ***Future directions***
1178   The success of the current strategy to systematically and comprehensively dissect cell types and
1179   generate a cell census and atlas opens up numerous avenues for future work.   This census and
1180   atlas form the foundation for the larger community to study specific features of cell types and
1181   aggregate information about cell types across species much as genomic databases aggregate
1182   information about genes. Classification of cell types and description of their molecular, spatial,
1183   and connectional properties in the adult sets the stage for developmental studies to understand
1184   the molecular genetic programs underlying cell type specification, maturation and circuit
1185   connectivity. The molecular classification and the utility of combined single cell transcriptomics
1186   and epigenomics to identify functional enhancers promises to deliver tools for genetic access to
1187   the great majority of brain cell types via transgenic and viral strategies. A combination of some
1188   of the approaches, such as imaging-based single-cell transcriptomics, with behavior stimulation
1189   and functional imaging can further elucidate the functional roles of distinct cell types in circuit
1190   computation. This systematic, multi-modal strategy described here is extensible to the whole
1191   brain, and major efforts are underway in the BICCN to generate a brain-wide cell census and
1192   atlas in the mouse with increasing coverage of human and non-human primates.
1193
1194
1195   **METHODS**
1196
1197   ***Integrating 10x v3 snRNA-seq datasets across species***

A multimodal cell census and atlas of the mammalian primary motor cortex

1198 To identify homologous cell types across species, human, marmoset, and mouse 10x v3 snRNA-
1199 seq datasets were integrated using Seurat's SCTransform workflow. Each major cell class
1200 (glutamatergic, GABAergic, and non-neuronal cells) was integrated separately across species.
1201 Expression matrices were reduced to 14,870 one-to-one orthologs across the three species (NCBI
1202 Homologene, 11/22/2019). Nuclei were downsampled to have approximately equivalent
1203 numbers at the subclass level across species. Marker genes were identified for each species'
1204 cluster and used as input to guide alignment and anchor-finding during integration steps. For full
1205 methods see [48].
1206
1207 *Estimation of cell type homology*
1208 To establish a robust cross-species cell type taxonomy, we applied a tree-based clustering
1209 method on integrated class-level datasets (https://github.com/AllenInstitute/BICCN_M1_Evo).
1210 The integrated space (from the above mentioned Seurat integration) was over-clustering into
1211 small sets of highly similar nuclei for each class (~500 clusters per class). Clusters were
1212 aggregated into metacells, then hierarchical clustering was performed based on the metacell gene
1213 expression matrix using Ward's method. Hierarchical trees were then assessed for cluster size,
1214 species mixing, and branch stability by subsampling the dataset 100 times with 95% of nuclei.
1215 Finally, we recursively searched every node of the tree, and if certain heuristic criteria were not
1216 sufficed for a node below the upper node, all nodes below the upper node were pruned and nuclei
1217 belonging to this subtree were merged into one homologous group. We identified 24
1218 GABAergic, 13 glutamatergic, and 8 non-neuronal cross-species consensus clusters that were
1219 highly mixed across species and robust. For full methods see [48].
1220
1221 *Cross-species differential gene expression and correlations*
1222 Expression matrices for each species, for each major cell class (GABAergic, glutamatergic, and
1223 non-neuronal cells) were normalized using Seurat's SCTransform function with default
1224 parameters to generate a 'corrected UMI' matrix and remove technical variation within each
1225 species. SCTransform normalized counts matrices were then counts per 100,000 UMI (CP100K)
1226 normalized to account for variable sequencing depths between species. CP100K normalization
1227 was performed by multiplying each value in the 'corrected UMI' (SCTransform normalized)
1228 matrix by 100,000 and dividing by the column sums (total UMIs from each nuclei).
1229 SCTransform-CP100K normalized matrices were then used to find DE genes and correlations
1230 between species for each cross-species cluster.
1231
1232 DE gene analysis was performed with Seurat's FindAllMarkers function, using the Wilcoxon
1233 rank sum test, between each pair of species for a given cross-species cluster (e.g. human
1234 Lamp5_1 vs. marmoset Lamp5_1, human Lamp5_1 vs. mouse Lamp5_1, and marmoset
1235 Lamp5_1 vs. mouse Lamp5_1). Marker genes (FDR < 0.01, log fold-change > 2, expressed in at
1236 least 10% of nuclei) from each pairwise species comparison were identified for each cross-
1237 species cluster. We report the sum of marker genes between each species comparison as a

1238   heatmap in Figure 2e and show that human and marmoset have fewer DE genes between each
1239   other than with mouse across all cross-species clusters.
1240
1241   To visualize the correspondence of a given cross-species cluster between each pair of species, we
1242   first found the average SCTransform-CP100K expression for each cross-species cluster for each
1243   species. Average expression was then log-transformed and the spearman correlations between
1244   each species pair were identified and reported in the Figure 2d heatmap, which shows human and
1245   marmoset have higher correlations than either primate with mouse for all clusters except Endo,
1246   VLMC, and Microglia/PVM clusters (likely due to differences in sampling).
1247
1248   ***Integrating mouse transcriptomic, spatially resolved transcriptomic, and epigenomic datasets***
1249   To integrate IT cell types from different mouse datasets, we first take all cells that are labeled as
1250   IT, except for L6_IT_Car3, from the 11 datasets as listed in Figure 8a. These cell labels are
1251   either from dataset-specific analyses [54,64,79], or from the integrated clustering of multiple datasets
1252   [45]. The integrated clustering and embedding of the 11 datasets are then generated by projecting
1253   all datasets into the 10x v2 scRNA-seq dataset using SingleCellFusion [45,59]. Genome browser
1254   views of IT and ET cell types (Figure 8b and Figure 9d) are taken from the corresponding cell
1255   types of the brainome portal (brainome.ucsd.edu/BICCN_MOp) [45].
1256
1257   ***Integration of L5 ET cells from Epi-Retro-Seq and 10x snRNA-Seq***
1258   For snRNA-Seq, the 4,515 cells from 10x v3 B dataset labeled as L5 ET by SCF were selected
1259   [45]. The read counts were normalized by the total read counts per cell and log transformed. Top
1260   5,000 highly variable genes were identified with Scanpy [106] and z-score scaled across all the
1261   cells. For Epi-Retro-Seq, the posterior methylation levels of 12,261 genes in the 848 L5 ET cells
1262   were computed [79]. Top 5,000 highly variable genes were identified with AllCools [59] and z-score
1263   scaled across all the cells. The 1,512 genes as the intersection between the two highly variable
1264   gene lists were used in Scanorama [107] to integrate the z-scored expression matrix and minus z-
1265   scored methylation matrix with sigma equal to 100.
1266
1267   ***Identification of candidate cis-regulatory elements***
1268   For peak calling in the snATAC-seq data, we extracted all the fragments for each cluster, and
1269   then performed peak calling on each aggregate profile using MACS2 [108] with parameter: "--
1270   nomodel --shift −100 --ext 200 --qval 1e-2 –B --SPMR". First, we extended peak summits by
1271   250 bp on either side to a final width of 501 bp. Then, to account for differences in performance
1272   of MACS2 based on read depth and/or number of nuclei in individual clusters, we converted
1273   MACS2 peak scores (-log10(q-value)) to "score per million" [109]. Next, a union peak set was
1274   obtained by applying an iterative overlap peak merging procedure, which avoids daisy-chaining
1275   and still allows for use of fixed-width peaks. Finally, we filtered peaks by choosing a "score per
1276   million" cut-off of 5 as candidate cis-regulatory elements (cCREs) for downstream analysis.
1277

A multimodal cell census and atlas of the mammalian primary motor cortex

1278 *Predicting enhancer-promoter interactions*
1279 First, co-accessible cCREs are identified for all open regions in all neurons types (cell clusters
1280 with less than 100 nuclei from snATAC-seq are excluded), using Cicero [110] with following
1281 parameters: aggregation k = 50, window size = 500 kb, distance constraint = 250 kb. In order to
1282 find an optimal co-accessibility threshold, we generated a random shuffled cCRE-by-cell matrix
1283 as background and calculated co-accessible scores from this shuffled matrix. We fitted the
1284 distribution of co-accessibility scores from random shuffled background into a normal
1285 distribution model by using R package fitdistrplus [111]. Next, we tested every co-accessible cCRE
1286 pair and set the cut-off at co-accessibility score with an empirically defined significance
1287 threshold of FDR<0.01. The cCREs outside of $\pm$ 1 kb of transcriptional start sites (TSS) in
1288 GENCODE mm10 (v16) were considered distal. Next, we assigned co-accessibility pairs to three
1289 groups: proximal-to-proximal, distal-to-distal, and distal-to-proximal. In this study, we focus
1290 only on distal-to-proximal pairs. We calculated the Pearson's correlation coefficient (PCC)
1291 between gene expression (scRNA SMART-seq) and cCRE accessibility across the joint clusters
1292 to examine the relationships between the distal cCREs and target genes as predicted by the co-
1293 accessibility pairs. To do so, we first aggregated all nuclei/cells from scRNA-seq and snATAC-
1294 seq for every joint cluster to calculate accessibility scores (log2 CPM) and relative expression
1295 levels (log2 TPM). Then, PCC was calculated for every gene-cCRE pair within a 1 Mbp window
1296 centered on the TSS for every gene. We also generated a set of background pairs by randomly
1297 selecting regions from different chromosomes and shuffling of cluster labels. Finally, we fit a
1298 normal distribution model and defined a cut-off at PCC score with an empirically defined
1299 significance threshold of FDR<0.01, in order to select significant positively correlated cCRE-
1300 gene pairs.
1301
1302 *Identification of cis-regulatory modules*
1303 We used Nonnegative Matrix Factorization (NMF) to group cCREs into cis-regulatory modules
1304 based on their relative accessibility across cell clusters. We adapted NMF (Python package:
1305 sklearn) to decompose the cell-by-cCRE matrix V (N×M, N rows: cCRE, M columns: cell
1306 clusters) into a coefficient matrix H (R×M, R rows: number of modules) and a basis matrix W
1307 (N×R), with a given rank R:
1308 $$V \approx WH,$$
1309 The basis matrix defines module related accessible cCREs, and the coefficient matrix defines the
1310 cell cluster components and their weights in each module. The key issue to decompose the
1311 occupancy profile matrix was to find a reasonable value for the rank R (i.e., the number of
1312 modules). Several criteria have been proposed to decide whether a given rank R decomposes the
1313 occupancy profile matrix into meaningful clusters. Here we applied a measurement called
1314 "Sparseness" [112] to evaluate the clustering result. Median values were calculated from 100 times
1315 for NMF runs at each given rank with a random seed, which will ensure the measurements are
1316 stable. Next, we used the coefficient matrix to associate modules with distinct cell clusters. In the
1317 coefficient matrix, each row represents a module and each column represents a cell cluster. The

1318 values in the matrix indicate the weights of clusters in their corresponding module. The
1319 coefficient matrix was then scaled by column (cluster) from 0 to 1. Subsequently, we used a
1320 coefficient > 0.1 (~95th percentile of the whole matrix) as a threshold to associate a cluster with
1321 a module. Similarly, we associated each module with accessible elements using the basis matrix.
1322 For each element and each module, we derived a basis coefficient score, which represents the
1323 accessible signal contributed by all clusters in the defined module.
1324
1325 ***Identification of subclass-selective TFs by both RNA expression and motif enrichment***
1326 All analyses for this section were at the subclass level. For RNA expression, we used the sc
1327 SMART-seq dataset and compared each subclass with the rest of the population through a one-
1328 tailed Wilcoxon test and FDR correction to select significantly differentially-expressed
1329 transcription factors (adjusted P-value < 0.05, cluster average fold change > 2). To perform the
1330 motif enrichment analysis, we used known motifs from the JASPAR 2020 database [113] and the
1331 subclass specific hypo-CG-DMR identified in Yao et al [45]. The AME software from the MEME
1332 suite (v5.1.1) [114] was used to identify significant motif enrichment (adjusted P-value < 1e-3, odds
1333 ratio > 1.3) using default parameters and the same background region set as described in Yao et
1334 al [45]. All genes in **Extended Data Figure 3** were both significantly expressed and had their motif
1335 enriched in at least one of the subclasses.
1336
1337 ***Generation and use of transgenic mouse lines***
1338 Npnt-P2A-FlpO and Slco2a1-P2A-Cre mouse driver lines were generated by CRISPR/Cas9-
1339 mediated homologous recombination (Stafford et al., BICCN companion manuscript in
1340 preparation). Details are provided in the Supplementary Methods.
1341
1342 All experimental procedures were approved by the Institutional Animal Care and Use
1343 Committees (IACUC) of Cold Spring Harbor Laboratory, University of California, Berkeley and
1344 Allen Institute, in accordance with NIH guidelines. Mouse knock-in driver lines are being
1345 deposited at the Jackson Laboratory for wide distribution.
1346
1347
1348 ***Data and code availability***
1349
1350 **Figure 1. Summary of experimental and computational approaches taken as well as**
1351 **community resources generated by the BICCN**
1352
1353 All primary data available through the BICCN portal, data archives, and supporting tools.
1354   ● Brain Cell Data Center (BCDC), www.biccn.org
1355   ● Neuroscience Multi-Omics Archive (NeMO), www.nemoarchive.org
1356   ● Brain Image Library (BIL), www.brainimagelibrary.org
1357   ● Neurophysiology (DANDI), dandiarchive.org

A multimodal cell census and atlas of the mammalian primary motor cortex

1358   • Allen Transcriptomics Explorer, https://portal.brain-map.org/atlases-and-data/rnaseq
1359   • NeMO Analytics, www.nemoanalytics.org
1360   • Morphological reconstructions, NeuroMorpho, www.neuromorpho.org
1361

1362 **Figure 2. MOp consensus cell type taxonomy**

1363

1364 **Primary Data**

Panel a		http://data.nemoarchive.org/publication_release/MOp_MiniAtlas_2020/
Panels b-g	10x V3 macaque	http://data.nemoarchive.org/biccn/lab/lein/lein/transcriptomic/sncell/raw/
	10x V3 human (10X159-1 through 10x160-8)	http://data.nemoarchive.org/biccn/lab/linnarsson/transcriptome/sncell/10X/raw/10X159-1/
	10x V3 marmoset (bi005_m1, bi006_m1)	http://data.nemoarchive.org/biccn/lab/feng/transcriptome/sncell/raw/
	10x V3 mouse broad data (files with prefix pBICCNsMMrMOp)	http://data.nemoarchive.org/biccn/grant/huang/macosko/transcriptome/sncell/raw
Panel h:		

1365

1366 **Intermediate analyses**

Panel a:	https://github.com/mukamel-lab/BICCN-Mouse-MOp/tree/master/flagship_fig2a
Panel b:	dendrogram from companion paper (Bakken et al. 2020)
Panel c:	http://data.nemoarchive.org/brain/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/flagship/dendrogram_barplots
Panels d, e:	http://data.nemoarchive.org/brain/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/flagship/dendrogram_heatmaps
Panels f, g:	from companion paper (Bakken et al. 2020)
Panel h:	http://data.nemoarchive.org/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/flagship/cross_species_heatmap/GABAergic_avg_CP100K_expr.csv.gz
Panel i:	Custom UCSC browser of all M1 tracks https://genome.ucsc.edu/s/sarojas/hg38-mop-dense

1367

1368 **Extended Data**

Panel j:	Browser https://brainome.ucsd.edu/annoj/BICCN_MOp/
Extended Data Figure 1:	http://data.nemoarchive.org/brain/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/flagship/cluster_overlap_plots

1369

A multimodal cell census and atlas of the mammalian primary motor cortex

1370 **Figure 3: In situ cell-type identification, spatial mapping and projection mapping of**
1371 **individual cells in the MOp by MERFISH**
1372
1373 **Primary Data**
1374 ftp://download.brainimagelibrary.org:8811/02/26/02265ddb0dae51de/
1375
1376 **Figure 4. Correspondence between transcriptomic and morpho-electrical properties of**
1377 **mouse MOp neurons by Patch-seq, and cross-species comparison of L5 ET neurons.**
1378
1379 **Primary Data**

Panel b,c,j,k	Electrophysiology data	https://dandiarchive.org/dandiset/000008
Panel h:	10x V3 macaque	http://data.nemoarchive.org/biccn/lab/lein/lein/transcriptomic/sncell/raw/
	10x V3 human (10X159-1 through 10x160-8)	http://data.nemoarchive.org/biccn/lab/linnarsson/transcriptome/sncell/10X/raw/10X159-1/
	10x V3 marmoset (bi005_m1, bi006_m1)	http://data.nemoarchive.org/biccn/lab/feng/transcriptome/sncell/raw/
	10x V3 mouse broad data (files with prefix pBICCNsMMrMOp)	http://data.nemoarchive.org/biccn/grant/huang/macosko/transcriptome/sncell/raw

1380
1381 **Intermediate analyses**

	10x 4 species integration	http://data.nemoarchive.org/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/cross_species_integration/sample.combined_exc_4_species_integration.RDS
Panel h:		
Panel h:	Code:	http://data.nemoarchive.org/brain/biccn/lab/lein/2020_M1_study_analysis/Transcriptomics/flagship/projecting_patch_seq_onto_umap

1382
1383 **Figure 5: Epi-Retro-Seq links molecular cell type with distal projection targets**
1384
1385 **Intermediate analyses**
1386 **https://github.com/zhoujt1994/BICCN2020Flagship.git**
1387
1388 **Figure 6: Global wiring diagram and anatomical characterization of MOp-ul neuron types**
1389
1390 **Primary Data**

Label in Fig	Link to registered swc (single cells) or 25 um grid file (tracer)	high resolution image data

A multimodal cell census and atlas of the mammalian primary motor cortex

Rabies	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	ftp://download.brainimagelibrary.org:8811/74/02/7402741313727c9b/tissuecyte_data/0500370607/
AAV	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	http://connectivity.brain-map.org/projection/experiment/127084296
Cux2 L2/3/4 IT	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	n/a
Nr5a1 L4 IT	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	n/a
Tlx3 L5 IT	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	n/a
Rbp4 L5 IT+ET	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	n/a
Sim1 L5 ET	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	http://connectivity.brain-map.org/projection/experiment/297711339
Ntsr1 L6 CT	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Viral_Tracer_Data_in_MOp_25_um_nrrd/	http://connectivity.brain-map.org/projection/experiment/159651060
IT projections	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Single_Cell_Reconstructions_in_MOp/	/bil/proj/u19zeng/fMOST_raw_data/mouseID_374712-18453_CH2 (pending upload)
IT projections	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Single_Cell_Reconstructions_in_MOp/	n/a
IT projections	http://ml-neuronbrowser.janelia.org/	n/a
ET projections	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Single_Cell_Reconstructions_in_MOp/	/bil/data/2b/da/2bdaf9e66a246844/mouseID_405429-182725 (pending upload)
ET projections	http://download.alleninstitute.org/publications/cellular_anatomy_of_the_mouse_primary_motor_cortex/Single_Cell_Reconstructions_in_MOp/	/bil/data/2b/da/2bdaf9e66a246844/mouseID_405429-182725 (pending upload)

A multimodal cell census and atlas of the mammalian primary motor cortex

CT projections	http://ml-neuronbrowser.janelia.org/	n/a
CT projections	http://ml-neuronbrowser.janelia.org/	n/a

1391

Panels d-i				
Label in Fig	**Full Descriptive ID**	**experiment id**	**Originating Lab**	**Link to Brain Architecture viewer**
PlxnD1	PlxnD1-CreER;LSL-Flp	180722	Huang	http://brainarchitecture.org/viewer4/mouse/map/8401F
PlxnD1	PlxnD1-CreER;LSL-Flp	180730	Huang	http://brainarchitecture.org/viewer4/mouse/map/28819F
Tle4	Tle4-CreER;LSL-Flp	180605	Huang	http://brainarchitecture.org/viewer4/mouse/map/28814F
Tle4	Tle4-CreER;LSL-Flp	180816	Huang	http://brainarchitecture.org/viewer4/mouse/map/8421F

1392
1393   **Intermediate analyses**

panel c	https://github.com/AllenInstitute/MOp_anatomy_rendering	code to reproduce rendering of registered data in 3D

1394
1395   **Figure 7: Genetic tools for targeting cortical glutamatergic projection neuron types**

1396
1397   **Primary Data**

Panels d-i					
Label in Fig	**Full Descriptive ID**	**experiment id**	**Originating Lab**	**Link to Brain Architecture viewer**	**BIL link**
PlxnD1	PlxnD1-CreER; LSL-Flp	180722	Huang	http://brainarchitecture.org/viewer4/mouse/map/8401F	https://download.brainimagelibrary.org/84/aa/84aa97d12a6c17ba/180722_WG_PlxnD1lslFlpCFA1female_processed/
PlxnD1	PlxnD1-CreER; LSL-Flp	180730	Huang	http://brainarchitecture.org/viewer4/mouse/map/28819F	https://download.brainimagelibrary.org/e9/2a/e92aa2dc0e14ad4d/180730_WG010_PlxnD1_CFA_female_processed/
Fezf2	Fezf2-	180830	Huang	http://brainarchitectu	https://download.brainimagelibrary.org

A multimodal cell census and atlas of the mammalian primary motor cortex

	CreER; LSL-Flp			re.org/viewer4/mouse/map/28827F	/db/b8/dbb827c84942c557/180830_JH_WG_Fezf2LSLflp_CFA_female/
Fezf2	Fezf2-CreER; LSL-Flp	190903	Huang	http://brainarchitecture.org/viewer4/mouse/map/28917F	https://download.brainimagelibrary.org/2b/6e/2b6e48dc425d16db/190903_JH_WG0006_Fezf2LSLflp_MOp_CFA_female_processed/
Tle4	Tle4-CreER; LSL-Flp	180605	Huang	http://brainarchitecture.org/viewer4/mouse/map/28814F	https://download.brainimagelibrary.org/84/aa/84aa97d12a6c17ba/180605_WG_Tle4lslFlpRPCFA_female_processed/
Tle4	Tle4-CreER; LSL-Flp	180816	Huang	http://brainarchitecture.org/viewer4/mouse/map/8421F	https://download.brainimagelibrary.org/c8/1f/c81fe306a97b33e8/180816_JH_WG_Tle4LSLFlpNPCfa_female/

1398

Figure 8: Existence of L4 excitatory neurons in MOp.

1400

Intermediate analyses

Panels a,c,e	https://github.com/mukamel-lab/BICCN-Mouse-MOp/tree/master/flagship_fig8
Panel b	https://brainome.ucsd.edu/annoj/BICCN_MOp/

1402

Figure 9: Two distinct L5 ET projection neuron types in MOp

1404

Intermediate analyses

Panel d	https://brainome.ucsd.edu/annoj/BICCN_MOp/
Panel e	https://github.com/zhoujt1994/BICCN2020Flagship.git

1406

Figure 10: An integrated multimodal census and atlas of cell types in the primary motor cortex of mouse, marmoset and human.

1409

Intermediate analyses
https://github.com/yal054/snATACutils

1412

Extended Data

https://github.com/yal054/snATACutils
https://github.com/lhqing/flagship_tf_figure (code and data for Extended data figure 3)

1414

1415

1416

SUPPLEMENTARY NOTES

1418

Nomenclature of the L5 ET subclass of glutamatergic neurons

In this manuscript we have adopted a nomenclature for major subclasses of cortical glutamatergic excitatory neurons, which have long-range projections both within and outside of the cortex, following a long tradition of naming conventions that often classify neurons based on their projection targets. This nomenclature is based on our *de novo* transcriptomic taxonomy (**Fig. 10**) that organizes cell types hierarchically and validates the naming of the primary branches of glutamatergic neurons by their major long-range projection targets. At these levels, glutamatergic neurons are clearly divided into several subclasses, the cortico-cortical and cortico-striatal projecting intratelencephalic (IT) neurons that are distributed across nearly all layers (L2/3 IT, L4/5 IT, L5 IT, L6 IT and L6 IT Car3), the layer 5 neurons projecting to extratelencephalic targets (L5 ET), the cortico-thalamic (CT) projecting neurons in layer 6 (L6 CT), the near-projecting (NP) neurons found in layers 5 and 6, and the L6b neurons whose projection patterns remain largely unknown.

While the IT, CT, NP and L6b neurons have been consistently labeled as such in the field, the L5 ET neurons have not been named consistently in the literature, largely due to their large variety of projection targets and other phenotypic features that vary depending on cortical areas and species. Here we use the term L5 ET (layer 5 extratelencephalic) to refer to this prominent and distinct subclass of neurons as a standard name that can be accurately used across cortical regions and across species, and we provide our rationale below.

It has long been appreciated that cortical layer 5 contains two distinct populations of neurons that can be distinguished, not only based on the presence or absence of projections to ET targets (ET and IT cells), but also based on their predominant soma locations, dendritic morphologies and intrinsic physiology [81]. Accordingly, various names incorporating these features have been adopted to refer to L5 ET versus L5 IT cells, such as L5b versus L5a, thick- versus thin-tufted and burst-firing versus regular-firing. And the most common term used to refer to L5 ET cells residing in motor cortical areas has been PT, which refers to neurons projecting to the pyramidal tract. As accurately stated in Wikipedia, "The **pyramidal tracts** include both the corticobulbar tract and the corticospinal tract. These are aggregations of efferent nerve fibers from the upper motor neurons that travel from the cerebral cortex and terminate either in the brainstem (*corticobulbar*) or spinal cord (*corticospinal*) and are involved in the control of motor functions of the body."

Due to the past wide use of the term PT, we do not take the decision to use L5 ET rather than PT lightly. However, in the face of multiple lines of evidence that have accumulated over the last several years [115,116] and prominently highlighted in this manuscript, it is now clear that PT represents only a subset of L5 ET cells and is thus unable to accurately encompass the entire L5 ET subclass. This realization is informed by comparisons across species and cortical areas, and

A multimodal cell census and atlas of the mammalian primary motor cortex

1459 by single-cell transcriptomics and descriptions of the projections of single neurons, as well as
1460 studies linking transcriptional clusters to projection targets.
1461
1462 As noted above, the overall transcriptomic relationships between cortical neurons are well-
1463 described by a hierarchical tree that closely matches developmental lineage relationships as
1464 neurons become progressively restricted in their adult fates [45,48] (**Fig. 10**). The cortical excitatory
1465 neurons are a major branch, distinct from inhibitory, glial, and epithelial cells. Subsequent
1466 splitting of the excitatory neurons reveals several major excitatory neuron subclasses – IT, L5
1467 ET, L6 CT, NP and L6b. These major subclasses are conserved across mammalian species [15,18],
1468 as well as across all cortical areas as shown in mouse [44]. It is therefore clear that names are
1469 needed that both accurately incorporate and accurately distinguish between neurons in these
1470 subclasses, and which are applicable across all cortical areas.
1471
1472 Also as noted above, a widely used alternative to L5 ET is PT. Further, this term is traditionally
1473 used along with CT to distinguish between cells with these different projections. The two main
1474 observations that make these alternative nomenclatures untenable are: 1) PT refers to motor
1475 neurons that project into medulla or spinal cord, but in many cortical areas (e.g. visual and
1476 auditory areas) none of the L5 ET cells are motor neurons; and 2) even in the motor cortex many
1477 cells in the L5 ET subclass do not project to the pyramidal tract and instead project solely to the
1478 thalamus (or to thalamus and other non-PT targets). This is revealed by single neuron
1479 reconstructions [26,68,86] (**Fig. 6 and 9**), BARseq [67], projections from neuron populations with
1480 known gene expression and anatomical position in mouse lines [71], and studies directly linking
1481 projections to transcriptomics [15,54] and epigenetics [79] (**Fig. 5 and 9**). The term PT therefore fails
1482 to be inclusive of the entire L5 ET subclass. Furthermore, the L5 CT cells within the L5 ET
1483 subclass are largely continuous with PT cells (or "PT-like" cells), not only genetically but also
1484 anatomically [54,64] (**Fig. 3-4**), as a majority of L5 ET cells project to multiple targets, typically
1485 including both the thalamus and the PT structures (e.g., medulla and spinal cord), as well as the
1486 midbrain (**Fig. 6 and 9**) [68]. Thus, the L5 ET subclass should neither be split into PT and CT, nor
1487 should the CT-only cells be omitted by use of the term PT. These facts also inform us that it is
1488 important to maintain a distinction between L5 CT (a type of L5 ET) and L6 CT (a major
1489 subclass of cortical excitatory neurons that is highly distinct from L5 ET, despite the presence of
1490 some L6 CT cells at the bottom of layer 5) [54]. CT can be accurately used as a generic term, but
1491 CT neurons do not belong to a single subclass of cortical excitatory neurons.
1492
1493 We recognize that another name that has been used to describe L5 ET cells is SCPN (subcerebral
1494 projection neuron) [82]. Given that the telencephalon is equivalent to the cerebrum, ET and
1495 subcerebral have the same meaning and the term L5-SCPN would be an accurate and equivalent
1496 alternative. But the "L5" qualifier is crucial in either case in order to distinguish these cells from
1497 the L6 CT subclass. We favor the use of ET because SCPN has not been widely adopted and due
1498 to symmetry with the widely used "IT" nomenclature. Alternatively, given their evidence that

1499 "unlike pyramidal tract neurons in the motor cortex, these neurons in the auditory cortex do not

1500 project to the spinal cord", Chen et al [67] used the term "pyramidal tract-like (PT- l)." We also

1501 favor L5 ET over L5 PT-l which clings to an inaccurate and now outdated nomenclature.

1502

1503

1504 **Supplementary Methods**

1505

1506 ***Generation of Npnt-P2A-FlpO and Slco2a1-P2A-Cre mouse lines***

1507 To generate lines bearing in-frame genomic insertions of *P2A-FlpO* or *P2A-Cre*, we engineered

1508 double-strand breaks at the stop codons of *Npnt* and *Slco2a1*, respectively, using

1509 ribonucleoprotein (RNP) complexes composed of SpCas9-NLS protein and in vitro transcribed

1510 sgRNA (Npnt: *GATGATGTGAGCTTGAAAAG* and Slco2a1: *CAGTCTGCAGGAGAATGCCT*).

1511 These RNP complexes were nucleofected into 10^6 v6.5 mouse embryonic stem cells

1512 (C57/BL6;129/sv; a gift from R. Jaenisch) along with repair constructs in which *P2A-FlpO* or

1513 *P2A-Cre* was flanked with the following sequences homologous to the target site, thereby

1514 enabling homology-directed repair.

1515

1516 Npnt-P2A-FlpO:

1517 TGGCCCTTGAGCTCTAGTGTTCCCACTTGCCATAGAAATCTGATCTTCGGTTTGGGGG

1518 AAGGGTTGCCTTACCATGCTCCATGAGTGAGCACTGGGAAAAGGGGCAGAGGAGGC

1519 CTGACCAGTGTATACGTTCTCTCCCTAGGTCATCTTCAAAGGTGAAAAAAGGCGTGG

1520 TCACACGGGGGAGATTGGATTGGATGATGTGAGCTTGAAGCGCGGAAGATGTGGAA

1521 GCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCT

1522 GGACCTATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTG

1523 CAAGACCCCGCCGAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAGGCCCA

1524 GCGGCGAAAAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATC

1525 ACCCACAACGGCACCGCGATCAAGAGGGCCACCTTCATGAGTTATAACACCATCAT

1526 CAGCAACAGCCTGAGTTTTGACATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGA

1527 CCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCATGGGAG

1528 TTCACGATTATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGT

1529 GTCCAGCCTGCAGCTGCAGTTCGAAAGCAGCGAGGAGGCCGACAAGGGGAATAGCC

1530 ACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAAGGCGAGAGCATCTGGGAGATT

1531 ACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGATTTACCAAAACGAAGAC

1532 CCTGTACCAGTTCCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACAT

1533 CAAGAACGTGGACCCGAAGAGCTTCAAGCTCGTCCAGAACAAGTATCTGGGCGTGA

1534 TCATTCAGTGCCTGGTCACGGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTT

1535 TCAGCGCCAGAGGCAGGATCGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGAAC

1536 AGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGC

1537 AGGAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAG

1538 AAGAACGCCCCCTACCCCATCTTCGCTATTAAAAACGGCCCTAAGAGCCACATCGGC

A multimodal cell census and atlas of the mammalian primary motor cortex

```
1539   AGGCACCTGATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGT
1540   GGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCC
1541   ACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCT
1542   ACGACCCCATCAGTAAGGAGATGATCGCCCTGAAGGACGAGACCAACCCCATCGAG
1543   GAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGATACC
1544   CCGCCTGGAACGGCATTATAAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATC
1545   AACAGGCGGATCTGAAAGAGGTCGCTGCTGAGAAGACCCCTGGCAGCTCCCGAGCT
1546   AGCAGTGAATTTGTCGCTCTCCCTCATTTCCCAATGCTTGCCCTCTTGTCTCCCTCTTA
1547   TCAGGCCTAGGGCAGGAGTGGGTCAGGAGGAAGGTTGCTTGGTGACTCGGGTCTCG
1548   GTGGCCTGTTTTGGTGCAATCCCAGTGAACAGTGACACTCTCGAAGTACAGGAGCAT
1549   CTGGAGACACCTCCGGGCCCTTCTG
1550
1551   Slco2a1-P2A-Cre:
1552   TGCCCCTGGGCCTCACCATACCTGTCTCTTCCTGCCTCATAGGTACCTGGGCCTACAG
1553   GTAATCTACAAGGTCTTGGGCACACTGCTGCTCTTCTTCATCAGCTGGAGGGTGAAG
1554   AAGAACAGGGAATACAGTCTGCAGGAGAATGCTTCCGGATTGATTGGAAGCGGAGC
1555   TACTAACTTCTCCCTGTTGAAACAAGCAGGGGATGTCGAAGAGAATCCTGGACCTAT
1556   GGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTGCAAGACTC
1557   CTCCAAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAGGCCCAGCGGCGAG
1558   AAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAAC
1559   GGCACCGCCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAG
1560   CCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGG
1561   CCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATC
1562   ATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGTGTCCAGCCT
1563   GCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAGCAAG
1564   AAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGAA
1565   GATCCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACC
1566   AGTTCCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACG
1567   TGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAG
1568   TGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCC
1569   AGAGGCAGGATCGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCC
1570   CGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGCAGGAGTAC
1571   CAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAACGC
1572   CCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCT
1573   GATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCA
1574   ACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATC
1575   ACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCC
1576   ATCAGCAAGGAGATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCA
1577   GCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGA
1578   ACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGG
```

A multimodal cell census and atlas of the mammalian primary motor cortex

1579 ATCTGACCTTCAGCTGGGACTACTGCCCTGCCCCAGAGACTGGATATCCTACCCCTC
1580 CACACCTACCTATATTAACTAATGTTAGCATGCCTTCCTCCTCCTTCC
1581
1582 Transfected cells were cultured and resulting colonies directly screened by PCR for correct
1583 integration using the following genotyping primers:
1584 **Genotyping primers**

	Flanking Primer	Internal recombinase Primer
Npnt-P2A-FlpO Left homology arm	ATGCATTGCTTCATGCCATA	CCTTCAGCAGCTGGTACTCC
Npnt-P2A-FlpO right homology arm	GATTGAGGTCAGGCCAGAAG	TCGACATCGTGAACAAGAGC
Slco2a1-P2A-Cre Left homology arm	CTGGTGAAAGGGGAACTCTTGCT	GATCCCTGAACATGTCCATCAGG
Slco2a1-P2A-Cre Right homology arm	TACAGCATCCCTGACAAACACCA	TAGCACCGCAGGTGTAGAGAAGG

1585
1586 The inserted transgenes were fully sequenced and candidate lines were analyzed for normal
1587 karyotype. Lines passing quality control were aggregated with albino morulae and implanted
1588 into pseudopregnant females, producing germline-competent chimeric founders which in turn
1589 were crossed with the appropriate reporter lines on the C57/BL6 background.
1590
1591
1592 **ACKNOWLEDGEMENTS**
1593

A multimodal cell census and atlas of the mammalian primary motor cortex

1632

1633 **AUTHOR INFORMATION**

1634

1635 BRAIN Initiative Cell Census Network (BICCN)

1636

1637 **Corresponding authors**

1638 Edward M. Callaway[1], Hong-Wei Dong[2,†], Joseph R. Ecker[3,4], Mike Hawrylycz[5], Z. Josh

1639 Huang[6,†], Ed S. Lein[5], John Ngai[7,8,†], Pavel Osten[6], Bing Ren[9,10], Andreas Savas Tolias[11,12],

1640 Owen White[13], Hongkui Zeng[5], Xiaowei Zhuang[14]

1641

1642 **BICCN Contributing PI**

A multimodal cell census and atlas of the mammalian primary motor cortex

1643   Giorgio A. Ascoli[15,16], M. Margarita Behrens[17], Edward M. Callaway[1], Jerold Chun[18], Hong-Wei
1644   Dong[2,†], Joseph R. Ecker[3,4], Guoping Feng[19,20,21], James C. Gee[22], Satrajit S. Ghosh[23], Yaroslav
1645   O. Halchenko[24], Mike Hawrylycz[5], Ronna Hertzano[13,25], Z. Josh Huang[6,†], Ed S. Lein[5],
1646   Byungkook Lim[26], Maryann E. Martone[27,28], Lydia Ng[5], John Ngai[7,8,†], Pavel Osten[6], Lior
1647   Pachter[29], Alexander J Ropelewski[30], Timothy L. Tickle[31], Andreas Savas Tolias[11,12], Owen
1648   White[13], X. William Yang[32,33], Hongkui Zeng[5], Kun Zhang[34], Xiaowei Zhuang[14]
1649

1650   **Principal manuscript editors**
1651   Z. Josh Huang[6,†], Ed S. Lein[5], Hongkui Zeng[5]
1652

1653   **Manuscript writing and figure generation**
1654   Giorgio A. Ascoli[15,16], Trygve E. Bakken[5], Philipp Berens[35,36,37,38], Edward M. Callaway[1], Tanya
1655   L. Daigle[5], Hong-Wei Dong[2,†], Joseph R. Ecker[3,4], Julie A. Harris[5,†], Mike Hawrylycz[5], Z. Josh
1656   Huang[6,†], Nikolas L. Jorstad[5], Brian E. Kalmbach[5,39], Dmitry Kobak[35], Ed S. Lein[5], Yang Eric
1657   Li[10], Hanqing Liu[3,40], Katherine S. Matho[6], Eran A. Mukamel[41], Maitham Naeemi[5], John
1658   Ngai[7,8,†], Pavel Osten[6], Bing Ren[9,10], Federico Scala[11,12], Pengcheng Tan[3,42], Jonathan T. Ting[5,39],
1659   Andreas Savas Tolias[11,12], Fangming Xie[43], Hongkui Zeng[5], Meng Zhang[14], Zhuzhu Zhang[3],
1660   Jingtian Zhou[3,44], Xiaowei Zhuang[14], Brian Zingg[2,†]
1661

1662   **Analysis coordination**
1663   Trygve E. Bakken[5], Edward M. Callaway[1], Hong-Wei Dong[2,†], Joseph R. Ecker[3,4], Julie A.
1664   Harris[5,†], Mike Hawrylycz[5], Z. Josh Huang[6,†], Ed S. Lein[5], Eran A. Mukamel[41], John Ngai[7,8,†],
1665   Pavel Osten[6], Bing Ren[9,10], Andreas Savas Tolias[11,12], Hongkui Zeng[5], Xiaowei Zhuang[14]
1666

1667   **Integrated data analysis**
1668   Ethan Armand[41], Trygve E. Bakken[5], Philipp Berens[35,36,37,38], Hong-Wei Dong[2,†], Joseph R.
1669   Ecker[3,4], Julie A. Harris[5,†], Z. Josh Huang[6,†], Nikolas L. Jorstad[5], Brian E. Kalmbach[5,39], Dmitry
1670   Kobak[35], Ed S. Lein[5], Yang Eric Li[10], Hanqing Liu[3,40], Eran A. Mukamel[41], Pavel Osten[6], Bing
1671   Ren[9,10], Federico Scala[11,12], Pengcheng Tan[3,42], Andreas Savas Tolias[11,12], Fangming Xie[43],
1672   Zizhen Yao[5], Hongkui Zeng[5], Meng Zhang[14], Zhuzhu Zhang[3], Jingtian Zhou[3,44], Xiaowei
1673   Zhuang[14]
1674

1675   **Sc/snRNA-seq data generation and processing**
1676   Darren Bertagnolli[5], Tamara Casper[5], Jerold Chun[18], Kirsten Crichton[5], Nick Dee[5], Dinh Diep[34],
1677   Song-Lin Ding[5], Weixiu Dong[34], Elizabeth L. Dougherty[45], Guoping Feng[19,20,21], Olivia Fong[5],
1678   Melissa Goldman[46], Jeff Goldy[5], Rebecca D. Hodge[5], Lijuan Hu[47], C. Dirk Keene[48], Fenna M.
1679   Krienen[46], Matthew Kroll[5], Blue B. Lake[34], Kanan Lathia[5], Ed S. Lein[5], Sten Linnarsson[47],
1680   Christine S. Liu[18,49], Evan Z. Macosko[45], Steven A. McCarroll[45,46], Delissa McMillen[5], Naeem
1681   M. Nadaf[45], Thuc Nghi Nguyen[5,†], Carter R. Palmer[18,49], Thanh Pham[5], Nongluk
1682   Plongthongkum[34], Nora M. Reed[46], Aviv Regev[45,50,†], Christine Rimorin[5], William J.

A multimodal cell census and atlas of the mammalian primary motor cortex

1683 Romanow[18,49], Steven Savoia[6], Kimberly Siletti[47], Kimberly Smith[5], Josef Sulc[5], Bosiljka Tasic[5],
1684 Michael Tieu[5], Amy Torkelson[5], Herman Tung[5], Cindy T.J. van Velthoven[5], Charles R.
1685 Vanderburg[45], Anna Marie Yanny[5], Hongkui Zeng[5], Kun Zhang[34]
1686

ATAC-seq data generation and processing
1688 M. Margarita Behrens[17], Jerold Chun[18], Dinh Diep[34], Weixiu Dong[34], Rongxin Fang[44],
1689 Xiaomeng Hou[9], Blue B. Lake[34], Yang Eric Li[10], Christine S. Liu[18,49], Jacinta D. Lucero[17], Julia
1690 K. Osteen[17], Carter R. Palmer[18,49], Antonio Pinto-Duarte[17], Nongluk Plongthongkum[34], Olivier
1691 Poirion[9], Sebastian Preissl[9], Bing Ren[9,10], William J. Romanow[18,49], Xinxin Wang[9,†], Kun
1692 Zhang[34]
1693

Methylcytosine data production and analysis
1695 Andrew I. Aldridge[3], Anna Bartlett[3], M. Margarita Behrens[17], Lara Boggeman[51], Carolyn
1696 O'Connor[51], Rosa G. Castanon[3], Huaming Chen[3], Joseph R. Ecker[3,4], Conor Fitzpatrick[51],
1697 Hanqing Liu[3,40], Jacinta D. Lucero[17], Chongyuan Luo[3,4,†], Joseph R. Nery[3], Michael Nunn[3], Julia
1698 K. Osteen[17], Antonio Pinto-Duarte[17], Angeline C. Rivkin[3], Wei Tian[3], Jingtian Zhou[3,44]
1699

Epi-Retro-Seq data generation and processing
1701 Anna Bartlett[3], M. Margarita Behrens[17], Lara Boggeman[51], Edward M. Callaway[1], Carolyn
1702 O'Connor[51], Rosa G. Castanon[3], Bertha Dominguez[52], Joseph R. Ecker[3,4], Conor Fitzpatrick[51],
1703 Tony Ito-Cole[1], Matthew Jacobs[1], Xin Jin[53], Cheng-Ta Lee[52], Kuo-Fen Lee[52], Paula Assakura
1704 Miyazaki[1], Eran A. Mukamel[41], Joseph R. Nery[3], Michael Nunn[3], Yan Pang[1], Antonio Pinto-
1705 Duarte[17], Mohammad Rashid[1], Angeline C. Rivkin[3], Jared B. Smith[53], Pengcheng Tan[3,42], Minh
1706 Vu[1], Elora Williams[53], Zhuzhu Zhang[3], Jingtian Zhou[3,44]
1707

Omics data analysis
1709 Ethan Armand[41], Trygve E. Bakken[5], Tommaso Biancalani[45], A. Sina Booeshaghi[29], Megan
1710 Crow[54], Dinh Diep[34], Sandrine Dudoit[55], Joseph R. Ecker[3,4], Rongxin Fang[44], Stephan Fischer[54],
1711 Olivia Fong[5], Jesse Gillis[54], Jeff Goldy[5], Qiwen Hu[56], Nikolas L. Jorstad[5], Peter V.
1712 Kharchenko[56], Fenna M. Krienen[46], Blue B. Lake[34], Ed S. Lein[5], Yang Eric Li[10], Sten
1713 Linnarsson[47], Hanqing Liu[3,40], Evan Z. Macosko[45], Eran A. Mukamel[41], John Ngai[7,8,†], Sheng-
1714 Yong Niu[3], Vasilis Ntranos[57], Lior Pachter[29], Olivier Poirion[9], Elizabeth Purdom[58], Aviv
1715 Regev[45,50,†], Davide Risso[59], Hector Roux de Bézieux[60], Kimberly Siletti[47], Kimberly Smith[5],
1716 Saroja Somasundaram[5], Kelly Street[61], Valentine Svensson[29], Bosiljka Tasic[5], Wei Tian[3], Eeshit
1717 Dhaval Vaishnav[45], Koen Van den Berge[58,62], Cindy T.J. van Velthoven[5], Joshua D. Welch[63],
1718 Fangming Xie[43], Zizhen Yao[5], Hongkui Zeng[5], Jingtian Zhou[3,44]
1719

Tracing and connectivity data generation
1721 Xu An[6], Helen S. Bateup[7,8,64], Ian Bowman[2,†], Rebecca K. Chance[7], Hong-Wei Dong[2,†], Nicholas
1722 N. Foster[2,†], William Galbavy[6,66], Hui Gong[65,67], Lin Gou[2,†], Julie A. Harris[5,†], Joshua T.

A multimodal cell census and atlas of the mammalian primary motor cortex

1723 Hatfield[6], Houri Hintiryan[2,†], Karla E. Hirokawa[5,†], Z. Josh Huang[6,†], Gukhan Kim[6], Daniel J.
1724 Kramer[7], Anan Li[65,67], Xiangning Li[65], Byungkook Lim[26], Qingming Luo[67,68], Katherine S.
1725 Matho[6], Lydia Ng[5], John Ngai[7,8,†], Rodrigo Muñoz-Castañeda[6], David A. Stafford[7], Hongkui
1726 Zeng[5], Brian Zingg[2,†]

1727

1728 **Morphology data generation and reconstruction**
1729 Tanya L. Daigle[5], Hong-Wei Dong[2,†], Zhao Feng[65], Hui Gong[65,67], Julie A. Harris[5,†], Karla E.
1730 Hirokawa[5,†], Z. Josh Huang[6,†], Xueyan Jia[67], Shengdian Jiang[69], Tao Jiang[67], Xiuli Kuang[70],
1731 Rachael Larsen[5], Phil Lesnar[5], Anan Li[65,67], Xiangning Li[65], Yaoyao Li[70], Yuanyuan Li[71], Lijuan
1732 Liu[69], Qingming Luo[67,68], Hanchuan Peng[69], Lei Qu[72], Miao Ren[68], Zongcai Ruan[69], Elise Shen[5],
1733 Yuanyuan Song[69], Wayne Wakeman[5], Peng Wang[73], Yimin Wang[74], Yun Wang[5], Lulu Yin[69],
1734 Jing Yuan[65,67], Hongkui Zeng[5], Sujun Zhao[69], Xuan Zhao[69]

1735

1736 **OLST/STPT and other data generation**
1737 Xu An[6], William Galbavy[6,66], Joshua T. Hatfield[6], Z. Josh Huang[6,†], Gukhan Kim[6], Katherine S.
1738 Matho[6], Arun Narasimhan[6], Pavel Osten[6], Ramesh Palaniswamy[6], Rodrigo Muñoz-Castañeda[6]

1739

1740 **Morphology, connectivity and imaging analysis**
1741 Xu An[6], Giorgio A. Ascoli[15,16], Samik Banerjee[6], Liya Ding[69], Hong-Wei Dong[2,†], Zhao Feng[65],
1742 Nicholas N. Foster[2,†], William Galbavy[6,66], Hui Gong[65,67], Julie A. Harris[5,†], Joshua T. Hatfield[6],
1743 Z. Josh Huang[6,†], Bingxing Huo[6], Xueyan Jia[67], Gukhan Kim[6], Hsien-Chi Kuo[5], Sophie
1744 Laturnus[35], Anan Li[65,67], Xu Li[6], Katherine S. Matho[6], Partha P Mitra[6], Judith Mizrachi[6],
1745 Maitham Naeemi[5], Arun Narasimhan[6], Lydia Ng[5], Pavel Osten[6], Ramesh Palaniswamy[6],
1746 Hanchuan Peng[69], Rodrigo Muñoz-Castañeda[6], Quanxin Wang[5], Yimin Wang[74], Yun Wang[5],
1747 Peng Xie[69], Feng Xiong[69], Yang Yu[5], Hongkui Zeng[5]

1748

1749 **Spatially resolved single-cell transcriptomics (MERFISH)**
1750 Hong-Wei Dong[2,†], Stephen W. Eichhorn[14], Zizhen Yao[5], Hongkui Zeng[5], Meng Zhang[14],
1751 Xiaowei Zhuang[14], Brian Zingg[2,†]

1752

1753 **Multimodal profiling (Patch-seq)**
1754 Philipp Berens[35,36,37,38], Jim Berg[5], Matteo Bernabucci[11,12], Yves Bernaerts[35], Cathryn René
1755 Cadwell[75], Jesus Ramon Castro[11,12], Rachel Dalley[5], Leonard Hartmanis[76], Gregory D.
1756 Horwitz[39,77], Xiaolong Jiang[11,12,78], Brian E. Kalmbach[5,39], C. Dirk Keene[48], Andrew L. Ko[79,80],
1757 Dmitry Kobak[35], Sophie Laturnus[35], Ed S. Lein[5], Elanine Miranda[11,12], Shalaka Mulherkar[11,12],
1758 Philip R. Nicovich[5,†], Scott F. Owen[5,†], Rickard Sandberg [76], Federico Scala[11,12], Kimberly
1759 Smith[5], Staci A. Sorensen[5], Zheng Huan Tan[11,12], Jonathan T. Ting[5,39], Andreas Savas Tolias[11,12],
1760 Hongkui Zeng[5]

1761

1762 **Transgenic tools**

A multimodal cell census and atlas of the mammalian primary motor cortex

Shona Allen[7], Xu An[6], Helen S. Bateup[7,8,64], Rebecca K. Chance[7], Tanya L. Daigle[5], William Galbavy[6,66], Joshua T. Hatfield[6], Dirk Hockemeyer[7,64,81], Z. Josh Huang[6,†], Gukhan Kim[6], Rodrigo Muñoz-Castañeda[7], Angus Y. Lee[82], Katherine S. Matho[6], John Ngai[7,8,†], David A. Stafford[7], Bosiljka Tasic[5], Matthew B. Veldman[32,33], X. William Yang[32,33], Zizhen Yao[5], Hongkui Zeng[5]

NeMO archive and analytics
Ricky S. Adkins[13], Seth A. Ament[13,83], Héctor Corrada Bravo[84], Robert Carter[13], Apaala Chatterjee[13], Carlo Colantuoni[13,85], Jonathan Crabtree[13], Heather Creasy[13], Victor Felix[13], Michelle Giglio[13], Brian R. Herb[13], Ronna Hertzano[13,25], Jayaram Kancherla[84], Anup Mahurkar[13], Carrie McCracken[13], Lance Nickel[13], Dustin Olley[13], Joshua Orvis[13], Michael Schor[13], Owen White[13]

Brain Image Library (BIL) archive
Greg Hood[30], Alexander J Ropelewski[30]

DANDI archive
Benjamin Dichter[86], Satrajit S. Ghosh[23], Michael Grauer[87], Yaroslav O. Halchenko[24], Brian Helba[87]

Brain Cell Data Center (BCDC)
Anita Bandrowski[27], Nikolaos Barkas[88], Benjamin Carlin[88], Florence D. D'Orazi[5], Kylee Degatano[88], James C. Gee[22], Thomas H Gillespie[27], Mike Hawrylycz[5], Farzaneh Khajouei[31], Kishori Konwar[31], Maryann E. Martone[27,28], Lydia Ng[5], Carol Thompson[5], Timothy L. Tickle[31]

Project management
Philipp Berens[35,36,37,38], Florence D. D'Orazi[5], Hui Gong[65,67], Houri Hintiryan[2,†], Kathleen Kelly[6], Dmitry Kobak[35], Blue B. Lake[34], Katherine S. Matho[6], Stephanie Mok[5], Michael Nunn[3], Federico Scala[11,12], Susan Sunkin[5], Carol Thompson[5]

Affiliations
1) Systems Neurobiology Laboratories, The Salk Institute for Biological Studies, La Jolla, CA 92037
2) Center for Integrative Connectomics, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Department of Neurology, Zilkha Neurogenetic Institute, Keck School of Medicine at USC, University of Southern California, Los Angeles, CA 90033
3) Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037
4) Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037
5) Allen Institute for Brain Science, Seattle, WA, 98109

A multimodal cell census and atlas of the mammalian primary motor cortex

1803   6) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

1804   7) Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

1805   8) Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720

1806   9) Center for Epigenomics, Department of Cellular and Molecular Medicine, University of
1807   California, San Diego School of Medicine, La Jolla, CA, USA

1808   10) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA

1809   11) Department of Neuroscience, Baylor College of Medicine, Houston, TX, 77030, USA

1810   12) Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston,
1811   TX, 77030, USA

1812   13) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD

1813   14) Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology,
1814   Department of Physics, Harvard University, Cambridge, MA 02138, USA.

1815   15) Center for Neural Informatics, Krasnow Institute for Advanced Study, George Mason
1816   University, Fairfax, VA, USA

1817   16) Bioengineering Department, George Mason University, Fairfax, VA, USA.

1818   17) Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla,
1819   CA 92037

1820   18) Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA

1821   19) McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge,
1822   Massachusetts 02139

1823   20) Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology,
1824   Cambridge, Massachusetts 02139

1825   21) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge,
1826   Massachusetts 02142

1827   22) Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104

1828   23) Massachusetts Institute of Technology, Cambridge, MA

1829   24) Dartmouth College, Hannover, NH

1830   25) Department of Otorhinolaryngology, Anatomy and Neurobiology, University of Maryland
1831   School of Medicine, Baltimore, MD

1832   26) Division of Biological Science, Neurobiology Section, University of California, San Diego,
1833   La Jolla, CA, USA

1834   27) Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093

1835   28) SciCrunch, Inc.

1836   29) California Institute of Technology, Pasadena, CA 91125

1837   30) Pittsburgh Supercomputing Center, Carnegie Mellon University, 300 South Craig Street,
1838   Pittsburgh PA 15213 USA

1839   31) Klarman Cell Observatory and Data Sciences Platform, Broad Institute of MIT and Harvard,
1840   Cambridge, MA 02142

1841   32) Semel Institute & Department of Psychiatry and Biobehavioral Science, UCLA, Los
1842   Angeles, CA 90095

A multimodal cell census and atlas of the mammalian primary motor cortex

1843   33) David Geffen School of Medicine, UCLA, Los Angeles, CA 90095
1844   34) Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093
1845   35) Institute for Ophthalmic Research, University of Tübingen, 72076, Germany
1846   36) Center for Integrative Neuroscience, University of Tübingen, 72076, Germany
1847   37) Institute for Bioinformatics and Medical Informatics, University of Tübingen, 72076,
1848   Germany
1849   38) Bernstein Center for Computational Neuroscience, University of Tübingen, 72076 Germany
1850   39) Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195
1851   40) Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093
1852   41) Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92037
1853   42) School of Pharmaceutical Sciences, Tsinghua University, Beijing, China, 100084
1854   43) Department of Physics, University of California, San Diego, La Jolla, CA 92037
1855   44) Bioinformatics and Systems Biology Graduate Program, University of California San Diego,
1856   La Jolla, CA 92093
1857   45) Broad Institute of MIT and Harvard, Cambridge, MA 02142
1858   46) Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur Boston MA
1859   02115
1860   47) Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics,
1861   Karolinska Institute, S-17177 Stockholm, Sweden
1862   48) Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA,
1863   USA
1864   49) Biomedical Sciences Program, School of Medicine, University of California, San Diego, La
1865   Jolla, CA, USA
1866   50) Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge MA 02140
1867   51) Flow Cytometry Core Facility, The Salk Institute for Biological Studies, La Jolla, CA 92037
1868   52) Peptide Biology Laboratories, The Salk Institute for Biological Studies, La Jolla, CA 92037
1869   53) Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA
1870   92037
1871   54) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring
1872   Harbor, NY 11724
1873   55) Department of Statistics and Division of Biostatistics, University of California, Berkeley,
1874   Berkeley, CA 94720-3860
1875   56) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
1876   57) University of California San Francisco, San Francisco, California, 94143,USA
1877   58) Department of Statistics, University of California, Berkeley, Berkeley, CA, USA
1878   59) Department of Statistical Sciences, University of Padova, Italy
1879   60) Division of Biostatistics, School of Public Health, University of California, Berkeley, CA,
1880   USA
1881   61) Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, 02215

A multimodal cell census and atlas of the mammalian primary motor cortex

62) Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

63) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

64) Chan Zuckerberg Biohub, San Francisco, CA 94158

65) Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, MoE Key Laboratory for Biomedical Photonics, Huazhong University of Science and Technology, Wuhan 430074, China

66) Program in Neuroscience, Department of Neurobiology and Behavior, Stony Brook University, Stony Brook, NY, 11794, USA

67) HUST-Suzhou Institute for Brainsmatics, Suzhou 215123, China

68) School of Biomedical Engineering, Hainan University, Haikou 570228, China

69) SEU-ALLEN Joint Center, Institute for Brain and Intelligence, Southeast University, Nanjing, Jiangsu, China

70) School of Optometry and Ophthalmology, Wenzhou Medical University, Wenzhou, Zhejiang, China

71) Key Laboratory of Intelligent Computation & Signal Processing, Ministry of Education, Anhui University, Hefei, Anhui, China

72) Anhui University, Hefei, Anhui, China

73) Shanghai University, Shanghai, China

74) School of Computer Engineering and Science, Shanghai University, Shanghai, China

75) Department of Pathology, University of California San Francisco, San Francisco, California, 94143,USA

76) Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, 17177, Sweden

77) Washington National Primate Research Center, University of Washington, Seattle, WA 98195

78) Jan and Dan Ducan Neurological research Institute, Houston,77030, TX, USA

79) Department of Neurological Surgery, University of Washington School of Medicine, Seattle, WA, USA

80) Regional Epilepsy Center at Harborview Medical Center, Seattle, WA, USA

81) Innovative Genomics Institute, University of California, Berkeley, CA, 94720

82) Cancer Research Laboratory, University of California, Berkeley

83) Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD

84) Center for Bioinformatics and Computational Biology, University of Maryland, College Park, College Park, MD, 20742.

85) Department Neurology & Department Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD

86) CatalystNeuro, LLC

87) Kitware Inc

A multimodal cell census and atlas of the mammalian primary motor cortex

1922 88) Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142
1923
1924 †Current address for I.B., H.W.D., N.N.F., L.G., H.H, B.Z.: Department of Neurobiology, David
1925 Geffen School of Medicine at UCLA, Los Angeles, CA 90095
1926 Current address for J.A.H., K.E.H., P.R.N.: Cajal Neuroscience, 1616 Eastlake Ave E, Suite 208,
1927 Seattle, WA 98102
1928 Current address for Z.J.H.: Department of Neurobiology, Duke University School of Medicine,
1929 Durham, NC 27705
1930 Current address for C.L.: Department of Human Genetics, University of California, Los Angeles
1931 Current address for J.N.: National Institute of Neurological Disorders and Stroke, National
1932 Institutes of Health, Bethesda, MD USA
1933 Current address for S.F.O.: Department of Neurosurgery, Stanford University School of
1934 Medicine, Stanford, CA 94305
1935 Current address for A.R.: Genentech, 1 DNA Way, South San Francisco, CA
1936 Current address for X.W.: McDonnell Genome Institute, Washington University School of
1937 Medicine, St. Louis, MO, USA
1938
1939 **Author contributions**
1940
1941 See the consortium author list for full details of author contributions. BICCN Contributing PIs:
1942 G.A.A., M.M.B., E.M.C., J.C., H.D., J.R.E., G.F., J.C.G., S.S.G., Y.O.H., M.H., R.H., Z.J.H.,
1943 E.S.L., B.L., M.E.M., L.N., J.N., P.O., L.P., A.J.R., T.L.T., A.S.T., O.W., X.W.Y., H.Z., K.Z.,
1944 X.Z. Principal manuscript editors: Z.J.H., E.S.L., H.Z. Manuscript writing and figure generation:
1945 G.A.A., T.E.B., P.B., E.M.C., T.L.D., H.D., J.R.E., J.A.H., M.H., Z.J.H., N.L.J., B.E.K., D.K.,
1946 E.S.L., Y.E.L., H.L., K.S.M., E.A.M., M.N., J.N., P.O., B.R., F.S., P.T., J.T.T., A.S.T., F.X.,
1947 H.Z., M.Z., Z.Z., J.Z., X.Z., B.Z. Analysis coordination: T.E.B., E.M.C., H.D., J.R.E., J.A.H.,
1948 M.H., Z.J.H., E.S.L., E.A.M., J.N., P.O., B.R., A.S.T., H.Z., X.Z. Integrated data analysis: E.A.,
1949 T.E.B., P.B., H.D., J.R.E., J.A.H., Z.J.H., N.L.J., B.E.K., D.K., E.S.L., Y.E.L., H.L., E.A.M.,
1950 P.O., B.R., F.S., P.T., A.S.T., F.X., Z.Y., H.Z., M.Z., Z.Z., J.Z., X.Z. Sc/snRNA-seq data
1951 generation and processing: D.B., T.C., J.C., K.C., N.D., D.D., S.D., W.D., E.L.D., G.F., O.F.,
1952 M.G., J.G., R.D.H., L.H., C.D.K., F.M.K., M.K., B.B.L., K.L., E.S.L., S.L., C.S.L., E.Z.M.,
1953 S.A.M., D.M., N.M.N., T.N.N., C.R.P., T.P., N.P., N.M.R., A.R., C.R., W.J.R., S.S., K.S., K.S.,
1954 J.S., B.T., M.T., A.T., H.T., C.T.V.V., C.R.V., A.M.Y., H.Z., K.Z. ATAC-seq data generation
1955 and processing: M.M.B., J.C., D.D., W.D., R.F., X.H., B.B.L., Y.E.L., C.S.L., J.D.L., J.K.O.,
1956 C.R.P., A.P., N.P., O.P., S.P., B.R., W.J.R., X.W., K.Z. Methylcytosine data production and
1957 analysis: A.I.A., A.B., M.M.B., L.B., C.O., R.G.C., H.C., J.R.E., C.F., H.L., J.D.L., C.L., J.R.N.,
1958 M.N., J.K.O., A.P., A.C.R., W.T., J.Z. Epi-Retro-Seq data generation and processing: A.B.,
1959 M.M.B., L.B., E.M.C., C.O., R.G.C., B.D., J.R.E., C.F., T.I., M.J., X.J., C.L., K.L., P.A.M.,
1960 E.A.M., J.R.N., M.N., Y.P., A.P., M.R., A.C.R., J.B.S., P.T., M.V., E.W., Z.Z., J.Z. Omics data
1961 analysis: E.A., T.E.B., T.B., A.S.B., M.C., D.D., S.D., J.R.E., R.F., S.F., O.F., J.G., J.G., Q.H.,

A multimodal cell census and atlas of the mammalian primary motor cortex

N.L.J., P.V.K., F.M.K., B.B.L., E.S.L., Y.E.L., S.L., H.L., E.Z.M., E.A.M., J.N., S.N., V.N., L.P., O.P., E.P., A.R., D.R., H.R.D.B., K.S., K.S., S.S., K.S., V.S., B.T., W.T., E.D.V., K.V.D.B., C.T.V.V., J.D.W., F.X., Z.Y., H.Z., J.Z. Tracing and connectivity data generation: X.A., H.S.B., I.B., R.K.C., H.D., N.N.F., W.G., H.G., L.G., J.A.H., J.T.H., H.H., K.E.H., Z.J.H., G.K., D.J.K., A.L., X.L., B.L., Q.L., K.S.M., L.N., J.N., R.M., D.A.S., H.Z., B.Z. Morphology data generation and reconstruction: T.L.D., H.D., Z.F., H.G., J.A.H., K.E.H., Z.J.H., X.J., S.J., T.J., X.K., R.L., P.L., X.L., Y.L., Y.L., L.L., Q.L., H.P., L.Q., M.R., Z.R., E.S., Y.S., W.W., P.W., Y.W., Y.W., L.Y., J.Y., H.Z., S.Z., X.Z. OLST/STPT and other data generation: X.A., W.G., J.T.H., Z.J.H., G.K., K.S.M., A.N., P.O., R.P., R.M. Morphology, connectivity and imaging analysis: X.A., G.A.A., S.B., L.D., H.D., Z.F., N.N.F., W.G., H.G., J.A.H., J.T.H., Z.J.H., B.H., X.J., G.K., H.K., S.L., A.L., X.L., K.S.M., P.P.M., J.M., M.N., A.N., L.N., P.O., R.P., H.P., R.M., Q.W., Y.W., Y.W., P.X., F.X., Y.Y., H.Z. Spatially resolved single-cell transcriptomics (MERFISH): H.D., S.W.E., Z.Y., H.Z., M.Z., X.Z., B.Z. Multimodal profiling (PATCH-seq): P.B., J.B., M.B., Y.B., C.R.C., J.R.C., R.D., L.H., G.D.H., X.J., B.E.K., C.D.K., A.L.K., D.K., S.L., E.S.L., E.M., S.M., P.R.N., S.F.O., R.S., F.S., K.S., S.A.S., Z.H.T., J.T.T., A.S.T., H.Z. Transgenic tools: S.A., X.A., H.S.B., R.K.C., T.L.D., W.G., J.T.H., D.H., Z.J.H., G.K., D.J.K., A.Y.L., K.S.M., J.N., D.A.S., B.T., M.B.V., X.W.Y., Z.Y., H.Z. NeMO archive and analytics: R.S.A., S.A.A., H.C.B., R.C., A.C., C.C., J.C., H.C., V.F., M.G., B.R.H., R.H., J.K., A.M., C.M., L.N., D.O., J.O., M.S., O.W. Brain Image Library (BIL) archive: G.H., A.J.R. DANDI archive: B.D., S.S.G., M.G., Y.O.H., B.H. Brain Cell Data Center (BCDC): A.B., N.B., B.C., F.D.D., K.D., J.C.G., T.H.G., M.H., F.K., K.K., M.E.M., L.N., C.T., T.L.T. Project management: P.B., F.D.D., H.G., H.H., K.K., D.K., B.B.L., K.S.M., S.M., M.N., F.S., S.S., C.T.

Competing interests

A.B. is a cofounder of SciCrunch, a company devoted to improving scientific communication. J.R.E. is a member of Zymo Research SAB. J.A.H. is currently employed by Cajal Neuroscience. K.E.H. is currently employed by Cajal Neuroscience. P.V.K. serves on the Scientific Advisory Board to Celsius Therapeutics Inc. M.E.M. is a founder and CSO of SciCrunch Inc., a UCSD tech start up that produces tools in support of reproducibility including RRIDs. P.R.N. is currently employed by Cajal Neuroscience. A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and until August 31, 2020 was an SAB member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and ThermoFisher Scientific. From August 1, 2020, A.R. is an employee of Genentech. B.R. is shareholder of Arima Genomics, Inc. and Epigenome Technologies, Inc. K.Z is a co-founder, equity holder and serves on the Scientific Advisor Board of Singlera Genomics. X.Z. is a co-founder and consultant of Vizgen.

REFERENCES

A multimodal cell census and atlas of the mammalian primary motor cortex

1. Waldeyer, W. Ueber einige neuere Forschungen im Gebiete der Anatomie des Centralnervensystems1. *DMW-Deutsche Medizinische Wochenschrift* **17**, 1213–1218 (1891).

2. Ramón y Cajal, S. *Histologie Du Système Nerveux De L'homme & Des Vertébrés*. (Maloine, 1909).

3. Somogyi, P. & Klausberger, T. Defined types of cortical interneurone structure space and spike timing in the hippocampus. *J. Physiol.* **562**, 9–26 (2005).

4. Sanes, J. R. & Masland, R. H. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu. Rev. Neurosci.* **38**, 221–246 (2015).

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

6. Huang, Z. J. & Paul, A. The diversity of GABAergic neurons and neural communication elements. *Nat. Rev. Neurosci.* **20**, 563–572 (2019).

7. Mukamel, E. A. & Ngai, J. Perspectives on defining cell types in the brain. *Curr. Opin. Neurobiol.* **56**, 61–68 (2019).

8. Petilla Interneuron Nomenclature Group *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).

9. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).

10. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).

11. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).

12. Svensson, V., da Veiga Beltrame, E. & Pachter, L. A curated database reveals trends in single-cell transcriptomics. *bioRxiv* 742304 (2019).

13. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).

14. Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015–1030.e16 (2018).

15. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

16. Harris, K. D. *et al.* Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLoS Biol.* **16**, e2006387 (2018).

17. Romanov, R. A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat. Neurosci.* **20**, 176–188 (2017).

18. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).

19. Kim, D.-W. *et al.* Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. *Cell* **179**, 713–728.e17 (2019).

2042   20.   Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in
2043   mammalian cortex. *Science* **357**, 600–604 (2017).
2044   21.   Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse
2045   forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439
2046   (2018).
2047   22.   Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in
2048   the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
2049   23.   Sinnamon, J. R. *et al.* The accessible chromatin landscape of the murine hippocampus at
2050   single-cell resolution. *Genome Res.* **29**, 857–869 (2019).
2051   24.   Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell
2052   chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
2053   25.   Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin
2054   Accessibility. *Cell* **174**, 1309–1324.e18 (2018).
2055   26.   Winnubst, J. *et al.* Reconstruction of 1,000 Projection Neurons Reveals New Cell Types
2056   and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* **179**, 268–281.e13
2057   (2019).
2058   27.   Gong, H. *et al.* Continuously tracing brain-wide long-distance axonal projections in mice at
2059   a one-micron voxel resolution. *Neuroimage* **74**, 87–98 (2013).
2060   28.   Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging.
2061   Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090
2062   (2015).
2063   29.   Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the
2064   hypothalamic preoptic region. *Science* **362**, eaau 5324 (2018).
2065   30.   Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of
2066   single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).
2067   31.   Fuzik, J. *et al.* Integration of electrophysiological recordings with single-cell RNA-seq data
2068   identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).
2069   32.   Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for
2070   neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017).
2071   33.   Huang, Z. J. & Zeng, H. Genetic approaches to neural circuits in the mouse. *Annu. Rev.*
2072   *Neurosci.* **36**, 183–215 (2013).
2073   34.   Graybuck, L. T. *et al.* Enhancer viruses and a transgenic platform for combinatorial cell
2074   subclass-specific labeling. *bioRxiv* 525014 (2020).
2075   35.   Mich, J. K. *et al.* Functional enhancer elements drive subclass-selective expression from
2076   mouse to primate neocortex. *bioRxiv* 555318 (2020).
2077   36.   Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced
2078   Brain-Cell-Type Targeting and Functionality. *Cell* **174**, 465–480.e22 (2018).
2079   37.   He, M. *et al.* Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in
2080   Mouse Cerebral Cortex. *Neuron* **91**, 1228–1243 (2016).
2081   38.   Dimidschstein, J. *et al.* A viral strategy for targeting and manipulating interneurons across

2082   vertebrate species. *Nat. Neurosci.* **19**, 1743–1749 (2016).

2083   39. Vormstein-Schneider, D. *et al.* Viral manipulation of functionally distinct interneurons in
2084   mice, non-human primates and humans. *Nat. Neurosci.* (2020) doi:10.1038/s41593-020-
2085   0692-9.

2086   40. Ecker, J. R. *et al.* The BRAIN Initiative Cell Census Consortium: Lessons Learned toward
2087   Generating a Comprehensive Brain Cell Atlas. *Neuron* **96**, 542–557 (2017).

2088   41. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference
2089   Atlas. *Cell* **181**, 936–953.e20 (2020).

2090   42. Lemon, R. N. Descending pathways in motor control. *Annu. Rev. Neurosci.* **31**, 195–218
2091   (2008).

2092   43. Svoboda, K. & Li, N. Neural mechanisms of movement planning: motor cortex and beyond.
2093   *Curr. Opin. Neurobiol.* **49**, 33–41 (2018).

2094   44. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal
2095   formation. *bioRxiv* 2020.03.30.015214 (2020).

2096   45. Yao, Z. *et al.* An integrated transcriptomic and epigenomic atlas of mouse primary motor
2097   cortex cell types. *bioRxiv* 2020.02.29.970558 (2020).

2098   46. Kozareva, V. *et al.* A transcriptomic atlas of the mouse cerebellum reveals regional
2099   specializations and novel cell types. *bioRxiv* 2020.03.04.976407 (2020).

2100   47. Krienen, F. M. *et al.* Innovations in Primate Interneuron Repertoire. *bioRxiv* 709501 (2019).

2101   48. Bakken, T. E. *et al.* Evolution of cellular diversity in primary motor cortex of human,
2102   marmoset monkey, and mouse. *bioRxiv* 2020.03.31.016972 (2020).

2103   49. Luo, C. *et al.* Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.*
2104   **9**, 3824 (2018).

2105   50. Liu, H. *et al.* DNA Methylation Atlas of the Mouse Brain at Single-Cell Resolution. *bioRxiv*
2106   2020.04.30.069377 (2020).

2107   51. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by
2108   combinatorial cellular indexing. *Science* **348**, 910–914 (2015).

2109   52. Li, Y. E. *et al.* An Atlas of Gene Regulatory Elements in Adult Mouse Cerebrum. *bioRxiv*
2110   2020.05.10.087585 (2020).

2111   53. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and
2112   chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).

2113   54. Zhang, M. *et al.* Molecular, spatial and projection diversity of neurons in primary motor
2114   cortex revealed by in situ single-cell transcriptomics. *bioRxiv* 2020.06.04.105700 (2020).

2115   55. Fang, R. *et al.* SnapATAC: A Comprehensive Analysis Package for Single Cell ATAC-seq.
2116   *bioRxiv* 615179 (2020).

2117   56. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-
2118   connected communities. *Sci. Rep.* **9**, 5233 (2019).

2119   57. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of
2120   Brain Cell Identity. *Cell* **177**, 1873–1887.e17 (2019).

2121   58. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell

A multimodal cell census and atlas of the mammalian primary motor cortex

2122   RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.*
2123   **36**, 421–427 (2018).

2124 59. Luo, C. *et al.* Single nucleus multi-omics links human cortical cell regulatory genome
2125   diversity to disease risk variants. *bioRxiv* 2019.12.11.873398 (2019).

2126 60. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21
2127   (2019).

2128 61. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of
2129   cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.*
2130   **9**, 884 (2018).

2131 62. Cadwell, C. R. *et al.* Multimodal profiling of single-cell morphology, electrophysiology,
2132   and gene expression using Patch-seq. *Nat. Protoc.* **12**, 2531–2553 (2017).

2133 63. Gouwens, N. W. *et al.* Toward an integrated classification of neuronal cell types:
2134   morphoelectric and transcriptomic characterization of individual GABAergic cortical
2135   neurons. *bioRxiv* 2020.02.03.932244 (2020).

2136 64. Scala, F. *et al.* Phenotypic variation within and across transcriptomic cell types in mouse
2137   motor cortex. *bioRxiv* 2020.02.03.929158 (2020).

2138 65. Berg, J. *et al.* Human cortical expansion involves diversification and specialization of
2139   supragranular intratelencephalic-projecting neurons. *bioRxiv* 2020.03.31.018820 (2020).

2140 66. Gong, H. *et al.* High-throughput dual-colour precision imaging for brain-wide connectome
2141   with cytoarchitectonic landmarks at the cellular level. *Nat. Commun.* **7**, 12142 (2016).

2142 67. Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In
2143   Situ Sequencing. *Cell* **179**, 772–786.e19 (2019).

2144 68. Peng, H. *et al.* Brain-wide single neuron reconstruction reveals morphological diversity in
2145   molecularly defined striatal, thalamic, cortical and claustral neuron types. *bioRxiv* 675280
2146   (2020).

2147 69. Muñoz-Castaneda, R. *et al.* Cellular Anatomy of the Mouse Primary Motor Cortex. *bioRxiv*
2148   2020.10.02.323154 (2020).

2149 70. Zingg, B. *et al.* Neural networks of the mouse neocortex. *Cell* **156**, 1096–1111 (2014).

2150 71. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature*
2151   **575**, 195–202 (2019).

2152 72. Zingg, B. *et al.* AAV-Mediated Anterograde Transsynaptic Tagging: Mapping
2153   Corticocollicular Input-Defined Neural Pathways for Defense Behaviors. *Neuron* **93**, 33–47
2154   (2017).

2155 73. Hintiryan, H. *et al.* The mouse cortico-striatal projectome. *Nat. Neurosci.* **19**, 1100–1114
2156   (2016).

2157 74. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

2158 75. Matho, K. S. *et al.* Genetic dissection of glutamatergic neuron subpopulations and
2159   developmental trajectories in the cerebral cortex. *bioRxiv* 2020.04.22.054064 (2020).

2160 76. Reardon, T. R. *et al.* Rabies Virus CVS-N2c(ΔG) Strain Enhances Retrograde Synaptic
2161   Transfer and Neuronal Viability. *Neuron* **89**, 711–724 (2016).

2162   77.   Tervo, D. G. R. *et al.* A Designer AAV Variant Permits Efficient Retrograde Access to
2163   Projection Neurons. *Neuron* **92**, 372–382 (2016).

2164   78.   Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single,
2165   genetically targeted neurons. *Neuron* **53**, 639–647 (2007).

2166   79.   Zhang, Z. *et al.* Epigenomic Diversity of Cortical Projection Neurons in the Mouse Brain.
2167   *bioRxiv* 2020.04.01.019612 (2020).

2168   80.   Veldman, M. B. *et al.* Brainwide Genetic Sparse Cell Labeling to Illuminate the
2169   Morphology of Neurons and Glia with Cre-Dependent MORF Mice. *Neuron* (2020)
2170   doi:10.1016/j.neuron.2020.07.019.

2171   81.   Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: themes and variations. *Nat.*
2172   *Neurosci.* **18**, 170–181 (2015).

2173   82.   Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype
2174   specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 427–437 (2007).

2175   83.   Scheibel, M. E., Davies, T. L., Lindsay, R. D. & Scheibel, A. B. Basilar dendrite bundles of
2176   giant pyramidal cells. *Exp. Neurol.* **42**, 307–319 (1974).

2177   84.   Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat.*
2178   *Commun.* **10**, 5416 (2019).

2179   85.   Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain.
2180   *Neuron* **86**, 1369–1384 (2015).

2181   86.   Economo, M. N. *et al.* Distinct descending motor cortex pathways and their roles in
2182   movement. *Nature* **563**, 79–84 (2018).

2183   87.   Battiste, J. *et al.* Ascl1 defines sequentially generated lineage-restricted neuronal and
2184   oligodendrocyte precursor cells in the spinal cord. *Development* **134**, 285–293 (2007).

2185   88.   Memic, F. *et al.* Ascl1 Is Required for the Development of Specific Neuronal Subtypes in
2186   the Enteric Nervous System. *J. Neurosci.* **36**, 4339–4350 (2016).

2187   89.   Bouyain, S. & Watkins, D. J. The protein tyrosine phosphatases PTPRZ and PTPRG bind to
2188   distinct members of the contactin family of neural recognition molecules. *Proc. Natl. Acad.*
2189   *Sci. U. S. A.* **107**, 2443–2448 (2010).

2190   90.   Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to
2191   study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**,
2192   1368–1383 (2013).

2193   91.   Harris, J. A. *et al.* Anatomical characterization of Cre driver mice for neural circuit mapping
2194   and manipulation. *Front. Neural Circuits* **8**, 76 (2014).

2195   92.   Hahn, J. D. *et al.* An open access mouse brain flatmap and upgraded rat and human brain
2196   flatmaps based on current reference atlases. *J. Comp. Neurol.* (2020)
2197   doi:10.1002/cne.24966.

2198   93.   Claudi, F., Tyson, A. L. & Branco, T. Brainrender. A python based software for
2199   visualisation of neuroanatomical and morphological data. *bioRxiv* 2020.02.23.961748
2200   (2020).

2201   94.   Dong, H. W. *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male*

A multimodal cell census and atlas of the mammalian primary motor cortex

2202   *mouse.* (John Wiley & Sons Inc, 2008).

2203   95. Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D. & Shepherd, G. M. G. A genuine
2204   layer 4 in motor cortex with prototypical synaptic circuit connectivity. *Elife* **3**, e05422
2205   (2014).

2206   96. Narayanan, R. T., Udvary, D. & Oberlaender, M. Cell Type-Specific Structural
2207   Organization of the Six Layers in Rat Barrel Cortex. *Front. Neuroanat.* **11**, 91 (2017).

2208   97. Yin, L. *et al.* Epigenetic regulation of neuronal cell specification inferred with single cell
2209   "Omics" data. *Comput. Struct. Biotechnol. J.* **18**, 942–952 (2020).

2210   98. Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and
2211   behaviors relevant to neurodevelopmental disorders. *Elife* **5**, (2016).

2212   99. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).

2213   100. Yuste, R. *et al.* A community-based transcriptomics classification and nomenclature of
2214   neocortical cell types. *Nat. Neurosci.* (2020) doi:10.1038/s41593-020-0685-8.

2215   101. Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature*
2216   **555**, 457–462 (2018).

2217   102. Mi, D. *et al.* Early emergence of cortical interneuron diversity in the mouse embryo.
2218   *Science* **360**, 81–85 (2018).

2219   103. Ginhoux, F. & Prinz, M. Origin of microglia: current concepts and past controversies. *Cold
2220   Spring Harb. Perspect. Biol.* **7**, a020537 (2015).

2221   104. Shepherd, G. M. *et al.* Neuron Names: A Gene- and Property-Based Name Format, With
2222   Special Reference to Cortical Neurons. *Front. Neuroanat.* **13**, 25 (2019).

2223   105. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757
2224   (2016).

2225   106. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression
2226   data analysis. *Genome Biol.* **19**, 15 (2018).

2227   107. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell
2228   transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).

2229   108. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

2230   109. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers.
2231   *Science* **362**, (2018).

2232   110. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell
2233   Chromatin Accessibility Data. *Mol. Cell* **71**, 858–871.e8 (2018).

2234   111. Delignette-Muller, M. & Dutang, C. fitdistrplus: An R Package for Fitting Distributions.
2235   *Journal of Statistical Software, Articles* **64**, 1–34 (2015).

2236   112. Hoyer, P. O. Non-negative Matrix Factorization with Sparseness Constraints. *J. Mach.
2237   Learn. Res.* **5**, 1457–1469 (2004).

2238   113. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor
2239   binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).

2240   114. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an
2241   evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).

2242   115. Saiki, A. *et al.* In Vivo Spiking Dynamics of Intra- and Extratelencephalic Projection
2243   Neurons in Rat Motor Cortex. *Cereb. Cortex* **28**, 1024–1038 (2018).
2244   116. Baker, A. *et al.* Specialized Subpopulations of Deep-Layer Pyramidal Neurons in the
2245   Neocortex: Bridging Cellular Properties to Functional Consequences. *J. Neurosci.* **38**,
2246   5441–5455 (2018).

2247

2248