JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Development/Plasticity/Repair

Anatomical diversity of the adult corticospinal tract revealed by single cell transcriptional profiling

https://doi.org/10.1523/JNEUROSCI.0811-22.2023

Cite as: J. Neurosci 2023; 10.1523/JNEUROSCI.0811-22.2023

Received: 26 April 2022
Revised: 28 July 2023
Accepted: 1 August 2023

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2023 the authors

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

1 **Title: Anatomical diversity of the adult corticospinal tract revealed by single cell**

2 **transcriptional profiling.**

3 **Abbreviated title: Transcriptional index of the adult corticospinal tract**

4

5 **Noa Golan[1,2], Daniel B Ehrlich[1,4], James Bonanno[1,2], Rory F O'Brien[2], Matias Murillo[1,2],**

6 **Sierra Kauer[2], Neal Ravindra[5,6], David van Dijk [5,6], and William BJ Cafferty[2,3]***

7 [1] Interdepartmental Neuroscience Program, [2] Department of Neurology, [3] Department of

8 Neuroscience,[4] Department of Psychiatry, [5] Department of Internal Medicine, [6] Department of

9 Computer Science, Yale University School, New Haven, CT, 06511, USA

10

11 *Correspondence: william.cafferty@yale.edu

12 Number of pages: 59

13 Number of figures: 7 main (7 extended data figures)

14 Number of words:

15 - Abstract 192/250

16 - Introduction 649/650

17 - Discussion 1494/1500

18 Conflict of interest: The authors declare no competing financial interests.

19 Acknowledgements: This work was supported by grants from the NIH.

JNeurosci Accepted Manuscript

20 **ABSTRACT**

21 The corticospinal tract (CST) forms a central part of the voluntary motor apparatus in all
22 mammals. Thus, injury, disease, and subsequent degeneration within this pathway result in
23 chronic irreversible functional deficits. Current strategies to repair the damaged CST are sub-
24 optimal in part due to underexplored molecular heterogeneity within the adult tract. Here we
25 combine spinal retrograde CST tracing with single-cell RNA sequencing in adult male and
26 female mice to index corticospinal neuron (CSN) subtypes that differentially innervate the
27 forelimb and hindlimb. We exploit publicly available datasets to confer anatomical
28 specialization among CSNs and show that CSNs segregate not only along the forelimb and
29 hindlimb axis but also by supraspinal axon collateralization. These anatomically defined
30 transcriptional data allow us to use machine learning tools to build classifiers that discriminate
31 between CSNs and cortical layer 2/3 and non-spinally terminating layer 5 neurons in M1, and
32 separately identify limb specific CSNs. Utilizing these tools, CSN subtypes can be differentially
33 identified to study postnatal patterning of the CST *in vivo*, leveraged to screen for novel limb-
34 specific axon growth survival and growth activators *in vitro*, and ultimately exploited to repair
35 the damaged CST after injury and disease.

36

37

38

39

40

JNeurosci Accepted Manuscript

41 **SIGNIFICANCE STATEMENT**

42 Therapeutic interventions designed to repair the damaged corticospinal tract (CST) after spinal
43 cord injury have remained functionally sub-optimal in part due to an incomplete understanding
44 of the molecular heterogeneity among subclasses of corticospinal tract neurons (CSNs). Here, we
45 combine spinal retrograde labeling with scRNAseq and annotate a CSN index by the termination
46 pattern of their primary axon in the cervical or lumbar spinal cord and supraspinal collateral
47 terminal fields. Using machine learning we have confirmed the veracity of our CSN gene lists to
48 train classifiers to identify CSNs among all classes of neurons in M1 to study the development,
49 patterning, homeostasis, and response to injury and disease, and ultimately target streamlined
50 repair strategies to this critical motor pathway.

51

52

53

54

55

56

57

58

59

60

JNeurosci Accepted Manuscript

61 **INTRODUCTION**

62 The corticospinal tract (CST) is the major descending motor pathway responsible for
63 driving fine coordinated movement in mammals (Lemon, 2008). Corticospinal neurons (CSNs)
64 in layer 5b (L5) of sensorimotor cortex project their axons through the internal capsule,
65 decussate in the medulla, and innervate every spinal segment along the neuroaxis. The CST is
66 wired postnatally, with pioneer axons extending to spinal enlargements followed by periods of
67 exuberant grey matter terminal arborization and pruning via activity-dependent mechanisms
68 during a protracted critical period that sculpts a mature motor pathway (Gianino et al., 1999;
69 Martin, 2005). Complex wiring of the mature CST, together with its central role in voluntary and
70 fine motor control means that damage leads to significant and lasting functional impairments.
71 Efforts to repair the damaged CST have focused on either nullifying the effects of the axon
72 growth inhibitory environment of the mature CNS, or recapitulating cell autonomous
73 developmental mechanisms (Schwab and Strittmatter, 2014; Gutilla and Steward, 2016; He and
74 Jin, 2016; Hilton and Bradke, 2017; Bradbury and Burnside, 2019; Fawcett, 2020). While strides
75 have been made in stimulating axotomized CSNs to regenerate or intact CSNs to undergo
76 plasticity after injury, recovery of significant motor function remains to be realized (Park et al.,
77 2008; Fink et al., 2017; Kauer et al., 2022). The inefficacy of current interventions could be due
78 in part to an incomplete understanding of the molecular heterogeneity among CSNs. For
79 instance, *in vitro*, and *in vivo* screening approaches designed to identify axon growth modulators
80 may overlook molecular heterogeneity within cortical neuron subclasses including CSNs,
81 consequently diluting the most potent candidates among differentially sensitive neuronal
82 subtypes (Blackmore et al., 2010; Buchser et al., 2010; Sekine et al., 2018).

JNeurosci Accepted Manuscript

83 Recent data from our laboratory supports differential sensitivity of CSN subdivisions. We
84 showed that novel pro-plasticity factors identified in intact CSNs undergoing functional
85 plasticity after unilateral pyramidotomy (PyX) stimulate growth of lesioned and intact forelimb
86 but not hindlimb CST axons (Fink et al., 2017; Kauer et al., 2022). Pro-axon growth candidates
87 *Lppr1* and *Inpp5k* were identified via retrograde labeling of sprouting CSNs in the cervical spinal
88 cord, suggesting that retrograde labeling of plastic CSNs from the lumbar cord may identify a
89 separate set of factors.

90 Single-cell RNA sequencing (scRNAseq) approaches afford the sensitivity to identify
91 transcriptional heterogeneity among CSNs. Indeed, previous studies have revealed the rich
92 phenotypic diversity among CNS cell types (Arlotta et al., 2005; Macosko et al., 2015; Saunders
93 et al., 2018; Tasic et al., 2018; Milich et al., 2020; Yao et al., 2020; Munoz-Castaneda et al.,
94 2021). Central to the utility of exploiting transcriptional indices are understanding the functional
95 or anatomical role of potential subclasses of cells. For instance, scRNAseq of corticocortical
96 projection neurons showed that while these cells form a single genetic cluster, their anatomically
97 traced subdivisions have diverse gene expression (Kim et al., 2020).

98 To identify CSN transcriptional heterogeneity, we combined retrograde AAV tracing
99 from the cervical and lumbar spinal cord with scRNAseq of adult CSNs. We used machine
100 learning to develop classifiers that go beyond cell marker characterization and provide robust
101 tools for unbiased classification of CSNs among other neurons in primary motor cortex (M1),
102 and among limb specific CSNs. We exploited taxonomic definitions of projection neuron cell
103 types in M1 to show that CSNs can be anatomically and molecularly defined based on
104 supraspinal terminals in addition to their spinal projections. Critically we present evidence
105 showing that intact adult CSNs express a constellation of genes previously shown to be unique to

JNeurosci Accepted Manuscript

106 non-spinally terminating projection neurons. Using intersectional viral tracing we confirm that
107 forelimb and hindlimb CSNs differentially innervate supraspinal structures further supporting
108 emerging data that CSNs influence motor output beyond directly modulating spinal motor
109 circuitry (Nelson et al., 2021). We believe that these sequencing data can be leveraged for *in*
110 *vitro* and *in vivo* screening, targeting, and exploitation strategies to enhance our understanding of
111 the development, patterning, housekeeping, and response to injury within the CST.

112

113

114

115

116

117

118

119

120

121

122

123

124

JNeurosci Accepted Manuscript

125 **MATERIALS AND METHODS**

126 **Mice**

127 Three mouse lines were used for experiments: *Rbp4* Cre mice (gift from Dr. David
128 Berson, Brown University), Ai14 (The Jackson Laboratory), and C57BL/6J (The Jackson
129 Laboratory). *Rbp4* Cre mice were crossed with the cre-dependent tdTomato reporter line Ai14,
130 to label deep layer cortical neurons and to confirm the veracity of our CSN classifier.

131 **Surgery**

132 All procedures and postoperative care were performed in accordance with the guidelines
133 of the Institutional Animal Use and Care Committee (IACUC) at Yale University.

134 *Intraspinal retrograde AAV injections for scRNAseq*

135 To complete retrograde labeling of cervical-projecting CSNs, adult mice (2-3 months old,
136 n= 19, 10 males, 9 females) were anesthetized with ketamine (100 mg/kg, Covetrus) and
137 xylazine (15 mg/kg, Covetrus) and placed in a stereotaxic frame (Stoelting, USA). An incision
138 was made over the cervical enlargement and the C5-C8 vertebrae revealed by blunt dissection of
139 overlying muscle. A bilateral laminectomy was performed to expose the underlying C5-C8
140 spinal cord, and a small incision was made in the dura mater. The tip of a pulled glass capillary
141 tube attached to a 5 µl Hamilton syringe loaded into a Micro4 infusion device (World Precision
142 Instruments, USA) was slowly inserted stereotaxically to a depth of 500 µm into the C6 level of
143 the spinal cord and approximately 600 µm lateral from the midline. Thirty seconds after the
144 introduction of the capillary tube, 100 nl of retro-AAV-CAG-GFP (Addgene, 37825-AAVrg)

JNeurosci Accepted Manuscript

145 was infused into the spinal cord over 2 minutes. The tip was left *in situ* for an additional 30
146 seconds prior to removal.

147 This procedure was completed 7 additional times bilaterally at the same coordinates at
148 evenly spaced injection sites between C6 and C7 resulting in a total infusion of 800 nl (100 nl
149 over 8 sites). Muscle layers were sutured with Vicryl (Ethicon, USA) and skin with
150 monofilament suture. All animals received post-surgical antibiotics (Ampicillin, 100 mg/kg
151 subcutaneously) and analgesia (Buprenorphine 0.05 mg/kg subcutaneously) for 2 days post
152 lesion. All animals recovered uneventfully.

153 To complete retrograde labeling of lumbar-projecting CSNs, adult mice (2-3 months old,
154 n=19, 10 males, 9 females) were anesthetized with ketamine (100 mg/kg) and xylazine (15
155 mg/kg) and placed in a custom-built spine stabilizer (Farrar et al., 2012). Using the last rib as a
156 landmark, an incision was made, and a bilateral laminectomy was performed to expose L4-L5
157 spinal cord. The tip of a pulled glass capillary tube attached to a 5 µl Hamilton syringe loaded
158 into a Micro4 infusion device was slowly inserted stereotaxically to a depth of 500 µm into the
159 L4 level of the spinal cord and approximately 500 µm lateral from the midline. Thirty seconds
160 after introduction of the capillary tube 100 nl of retro-AAV-CAG-GFP was infused into the
161 spinal cord over 2 minutes. The tip was left *in situ* for an additional 30 seconds prior to removal.
162 This procedure was completed 7 additional times bilaterally at the same coordinates at evenly
163 spaced injection sites between L4 and L5 resulting in a total infusion of 800 nl (100 nl over 8
164 sites). All animals received post-surgical antibiotics (Ampicillin, 100 mg/kg subcutaneously) and
165 analgesia (Buprenorphine 0.05 mg/kg subcutaneously) for 2 days post lesion. All animals
166 recovered uneventfully.

JNeurosci Accepted Manuscript

167 *Injections for single molecule fluorescent in situ hybridization validation of scRNAseq*

168 For confirmation studies, adult mice (2-3 months old, n = 3/probe) received bilateral
169 injections of retro-AAV-CAG-GFP into the cervical cord and retro-AAV-CAG-tdTomato
170 (Addgene, 59462-AAVrg) into the lumbar cord as described above for a total of 16 spinal
171 injections (100 nl over 8 sites in the cervical enlargement, and 100 nl over 8 sites in the lumbar
172 enlargement). All animals received post-surgical antibiotics (Ampicillin, 100 mg/kg
173 subcutaneously) and analgesia (Buprenorphine 0.05 mg/kg subcutaneously) for 2 days post
174 lesion. All animals recovered uneventfully.

175 *Intersectional tracing of supraspinal CST terminals for BrainJ analysis*

176 To label the forelimb (FL) and hindlimb (HL) corticospinal tract for projection analysis,
177 two groups of adult C57/blk6 mice received dual injections of pAAV-CAG-FLEX-tdTomato
178 (Addgene, 28306-AAV1) into either the caudal FL motor cortex (n=4, AP: between 0.0 mm and
179 0.5 mm rostral to bregma, ML: between 1.0 mm and 1.3 mm lateral to bregma, 4 injections 100
180 nl/site) or the HL motor cortex (n=4, AP: between -0.5 mm and -1 mm caudal to bregma, ML:
181 between 1.5 mm to 2.0 mm lateral to bregma, 4 injections 100 nl/site), and
182 pENN.AAV.hSyn.Cre.WPRE.hGH (Addgene 105553-AAVrg) into either the cervical (C6/7) or
183 lumbar spinal cord (L4/5), as described above. All animals received post-surgical antibiotics
184 (Ampicillin, 100 mg/kg subcutaneously) and analgesia (Buprenorphine 0.05 mg/kg
185 subcutaneously) for 2 days post lesion. All animals recovered uneventfully.

186 **Intact adult single-cell isolation for scRNAseq**

187 Adult intact cell isolation was done following our previously published protocol (Golan
188 and Cafferty, 2021). Briefly, mice that received retrograde injections into either the lumbar or

JNeurosci Accepted Manuscript

189 cervical cord were anesthetized with isoflurane and transcardially perfused with artificial
190 cerebrospinal fluid (aCSF) ten days after retro-AAV injection. The aCSF that consisted of
191 CaCl2 (0.5 mM), glucose (25 mM), HCl (96 mM), HEPES (20 mM), MgSO4 (10 mM),
192 NaH2PO4 (1.25 mM), myo-inositol (3 mM), N-acetylcysteine (12 mM), NMDG (96 mM), KCl
193 (2.5 mM), NaHCO3 (25 mM), sodium L-ascorbate (5 mM), sodium pyruvate (3 mM), taurine
194 (0.01 mM), thiourea (2 mM), trehalose (13.2mM) and was bubbled with carbogen gas (95% O2
195 and 5% CO2) (Tasic et al., 2018). The brain was dissected and submerged in ice-cold
196 carbogenated aCSF for three minutes before being transferred to a brain matrix (Braintree
197 Scientific, Inc.) where four 500 µm sections through M1 were sliced and sections were
198 transferred to a petri dish containing ice cold carbogenated aCSF. Regions of M1 (+1.0 → -0.5
199 mm relative to Bregma) containing labeled CSNs were dissected under a fluorescent microscope
200 and transferred to a 5 ml Eppendorf tube with dissociation buffer with papain.

201 The dissociation buffer contained sodium sulfate (82 mM), potassium sulfate (30 mM),
202 HEPES (10 mM), glucose (10 mM) and magnesium chloride (5 mM) (Saunders et al., 2018).
203 Thirty minutes prior to dissociation, the dissociation buffer was warmed to 34°C and 5 mL was
204 added to a vial of lyophilized papain (Worthington Biochemical Corporation, LK003178). The
205 solution was diluted 1:2 with additional dissociation buffer prior to sample dissociation.

206 The sample in dissociation buffer with papain was incubated at 34°C for 70 minutes on a
207 shaker at medium speed. After 70 minutes, the dissociation solution with papain was replaced
208 with a Stop buffer containing 5 ml of dissociation buffer, 5 mg Ovomucoid Protease Inhibitor
209 (Worthington Biochemical Corporation, LK003182) and 10 mg bovine serum albumin
210 (AmericanBio, AB01088) for 5 minutes on ice. The Stop solution was then replaced with 800 µl
211 of dissociation buffer and triturated with Pasteur pipettes with decreasing diameters (600 µm,

JNeurosci Accepted Manuscript

212 300 µm and 150 µm) and placed on ice. To differentiate between intact neurons and nuclei, a
213 cell-permeable dye was added at a concentration of 1:1000 for 5 minutes (Invitrogen, L34974).
214 The sample was then centrifuged at 300 g for 10 minutes at 4℃ and resuspended in 500 µl of
215 dissociation buffer.

216 **Fluorescence activated cell sorting (FACS)**

217 'In total 44 mice animals were used for the scRNAseq studies, processing of tissue was
218 completed in 7 batches on different days, 3 FL-traced cohorts (FL1: n=8 4F/4M, FL2: n=6 3F/3M, FL3:
219 n=5 2F/3M), and 3 HL-traced cohorts (HL1 n=6 3F/3M, HL2: n=7 3F/4M, HL3: n=6 3F/3M), and 1
220 naïve Rbp4 cre:Ai14 group (n=6 *rbp4* cre: Ai14 3F/3M)'. Fluorescent single cells were isolated from
221 *rbp4* cre: Ai14 mice and C57Bl/6J mice injected with retro-AAV-CAG-GFP. Sorting was
222 performed on the Sony SH800 sorter with a 130 µm nozzle, ensuring a sheath pressure and
223 sample pressure of less than 9 PSI. To ensure pyramidal neurons were sorted, size reference
224 beads were used to set a minimum diameter of 10 µm (Spherotech, PPS-5 and PPS-6) for all
225 collected cells. To exclude non-intact cells, the cells were sorted for the presence of the cell-
226 permeable dye. A sample of non-fluorescent tissue from adjacent regions of the cortex was used
227 as negative control to set the gate for the presence of fluorescence (GFP or tdTomato). The
228 sample was then sorted into Eppendorf tubes containing 4℃ dissociation buffer and immediately
229 sequenced at the Yale Center for Genome Analysis (YCGA).

230 In total 44 animals were used (4 *rbp4* cre: Ai14, 16 C57Bl/6J or Ai14 injected with retro-
231 AAV-CAG-GFP into C6/7, 24 C57Bl/6J or Ai14 injected with retro-AAV-CAG-GFP into L4/5).
232 Dissection of L5 of M1 was guided by fluorescence and consistent across all animals.

233 **10X single-cell RNAseq**

JNeurosci Accepted Manuscript

234 *Construction of 10X Genomic Single Cell/Nuclei 3' RNA-Seq libraries CITE-Seq Libraries and*
235 *sequencing.*

236 The single cell protocol enables short read sequencing to deliver a scalable microfluidic
237 platform for digital gene expression of 500 -10,000 individual cells per sample.

238 *GEM Generation and Barcoding*

239 Single cell suspension in RT Master Mix was loaded on the Single Cell A Chip and
240 partitioned with a pool of about 750,000 barcoded gel beads to form nanoliter-scale Gel Beads-
241 In-Emulsions (GEMs). Each gel bead has primers containing (i) an Illumina® R1 sequence (read
242 1 sequencing primer), (ii) a 16 nt 10x Barcode, (iii) a 12 nt Unique Molecular Identifier (UMI),
243 and (iv) a poly-dT primer sequence (30nt). Upon dissolution of the Gel Beads in a GEM, the
244 primers were released and mixed with cell lysate and Master Mix. Incubation of the GEMs then
245 produced barcoded, full-length cDNA from poly-adenylated mRNA.

246 *Post GEM-RT Cleanup, cDNA Amplification and library construction*

247 Silane magnetic beads were used to remove leftover biochemical reagents and primers
248 from the post GEM reaction mixture. Full-length, barcoded cDNA was then amplified by PCR to
249 generate sufficient mass for library construction. Enzymatic Fragmentation and Size Selection
250 were used to optimize the cDNA amplicon size prior to library construction. R1 (read 1 primer
251 sequence) was added to the molecules during GEM incubation. P5, P7, a sample index, and R2
252 (read 2 primer sequence) were added during library construction via End Repair, A-tailing,
253 Adaptor Ligation, and PCR. The final libraries contain the P5 and P7 primers were used in
254 Illumina bridge amplification.

255 *Sequencing libraries*

JNeurosci Accepted Manuscript

256 The Single Cell 3' Protocol produces Illumina-ready sequencing libraries. A Single Cell
257 3' Library comprises standard Illumina paired-end constructs which begin and end with P5 and
258 P7. The Single Cell 3' 16 bp 10x Barcode and 12 bp UMI are encoded in Read 1, while Read 2 is
259 used to sequence the cDNA fragment (91bp). Minimum sequencing depth is 20,000 read pairs
260 per cell. The RNAseq data will be deposited in the National Center for Biotechnology
261 Information (NCBI) Gene Expression Omnibus (GEO) and is accessible through GEO Series.

262 **Single-cell RNAseq analysis**

263 *Quality Control and Normalization*

264 Single-cell analyses were performed with Python 3.8.5 and the scipy (1.5.2) and numpy
265 (1.19.2) modules. After extracting the single-cell gene expression matrix, we used the scprep
266 (1.0.3) package in python to eliminate non-neurons and normalize the data (Gigante, 2021).
267 Briefly, we filtered out cells that had a library size of less than 2500 unique molecular identifiers
268 (UMIs) as well as cells that had mitochondrial gene expression in the top tenth percentile. To
269 exclude doublets, we used two methods: 1) We used the python package scrublet and removed
270 all cells that were detected as doublets (Wolock et al., 2019). 2) We removed cells that expressed
271 both the male specific genes (*Eif2s3y* or *Ddx3y*) and female specific genes (*Xist* or *Tsix*). After a
272 library size normalization, we used the python package MAGIC (3.0.0) to impute missing data
273 (van Dijk et al., 2018). Using the MAGIC output gene expression matrix, we filtered out
274 contaminating cell types (endothelial cells, microglia, astrocytes, and oligodendrocytes) as well
275 as inhibitory interneurons (via negative expression of *vglut2*), resulting in an approximate 1%
276 reduction from the starting neuronal population. We then square root transformed the data for
277 further downstream analysis.

JNeurosci Accepted Manuscript

278 *Batch Correction*

279 To batch correct collected data, we utilized the ingest function from the python package
280 scanpy (1.8.2). Once the datasets had been integrated, a graph was built in the integrated space
281 and MAGIC was used to impute gene expression based on the integrated graph.

282 To identify supraspinal connectivity of neurons sequenced in this study, we took a similar
283 approach. Briefly, the dataset collected in (Yao et al., 2021) was used as the reference dataset.
284 We then used the ingest function in scanpy and projected the dataset collected in this study into
285 the manifold calculated from Yao *et al*. Lastly, we used the label mapping output from scanpy to
286 identify intratelencephalic (IT), extratelencephalic (ET), near-project (NP), and corticothalamic
287 (CT) cells.

288 *Removing putatively mis-labeled neurons from downstream analyses*

289 To conduct accurate differential expression analyses, we first needed to identify
290 putatively mislabeled neurons. Due to inherent inefficiencies in viral labeling and FACS we
291 utilized a computational approach. We approximated the distribution for each of our cell
292 populations (L5 non-spinal, L2/3 and CSNs) using a Gaussian kernel density estimate (KDE)
293 using the python package scipy. To identify putative non-spinal non-fluorescent neurons, we
294 calculated the likelihood that each non-CSN (L5 and L2/3) came from the CSN distribution. We
295 removed putative CSNs (KDE > 50th percentile of CSN distribution) from the non-CSN samples.
296 Additionally, we used a similar approach to exclude a portion of L5 non-spinal neurons as they
297 were putative L2/3.

298 *Differential gene expression analysis*

JNeurosci Accepted Manuscript

299 Once we identified a pure non-spinally projecting layer 5 neuron population, we used the
300 python package diffxpy (Fischer, 2021) to run a Wilcoxon rank-sum test with a Benjamini-
301 Hochberg correction for multiple comparisons to identify differentially expressed genes between
302 GFP+ CSNs and non-spinal L5 cells. Using the significantly differentially expressed genes ($q <$
303 0.05), we used recursive feature elimination using sklearn.feature_selection.RFE (Pedregosa,
304 2011) to identify the top ten genes that best differentiate between CSNs and non-CSNs.
305 Differentially expressed genes were then used for QIAGEN IPA (Kramer et al., 2014) for
306 identification of commonly regulated pathways (QIAGEN Inc.,
307 https://digitalinsights.qiagen.com/IPA) and for the building of support machine vectors.

308 *Linear Support Vector Machine classification*

309 To classify cells as either CSNs or non-CSNs, we trained a Support Vector Machine
310 (SVM) classifier, with a linear kernel to predict cell origin based on the top 50 enriched genes in
311 CSNs and the top 50 enriched genes in non-CSNs, reserving half of our cells for cross validation
312 (Pedregosa, 2011). The SVM was balanced for class weight since there were an uneven number
313 of CSNs and non-CSNs. Probability outcomes were predicted by setting probability to "True"
314 during training. SVM performance was evaluated using the area under the receiver operating
315 characteristics curve (AUC ROC). Classifier performance was statistically compared using a
316 randomized permutation test.

317 The limb-specific classifier was built in a similar way. Briefly, the top 50 enriched
318 forelimb genes and the top 50 enriched hindlimb genes were used to build a linear SVM, with
319 half of the cells reserved for cross validation. SVM performance was evaluated using AUC ROC.
320 Classifier performances were statistically compared using a randomized permutation test.

321 *Classification of Rbp4+ neurons*

322 *Rbp4+* neurons were run through the same QC pipeline outlined above. They were batch
323 corrected with the sequenced cells in this study using the ingest function in scanpy. Their
324 expression was batch normalized with MAGIC based on the combined graph from the batch
325 correction. *Rbp4+* neurons with batch corrected expression values were then run through the
326 CSN-C classifier and neurons with a probability of being a CSN > 0.75 were classified as CSNs.

327 *CSN visualization*

328 For low-dimensional visualization of all scRNAseq data, we used the python package
329 PHATE (1.0.7) in three dimensions, with the default settings (Moon et al., 2019). All PHATE
330 projections are displayed in 2D, examples of 3D animated versions are included as movie files 1-
331 3, for **Figs. 3C**, **6D**, and **7A** respectively. Animated versions of all PHATE projections are
332 available upon request.

333 *MELD analysis of intermediate FL-HL CSN classification*

334 To investigate the distribution of cervical- and lumbar-labeled CSNs, and investigate the
335 possibility of an intermediate population, we used the python package MELD (1.0.0) to identify
336 the degree to which individual CSNs separated into the two groups (cervical or lumbar) defined
337 by labeling, which was expressed as a sample-associated relative likelihood (Burkhardt et al.,
338 2021). This analysis was performed on unimputed, square-root-transformed data, to avoid the
339 possibility that imputing over the whole dataset could have softened the distinction between the
340 cervical and lumbar groups.

341 **Immunohistochemistry (IHC) and single molecule fluorescent *in situ* hybridization**
342 **(smFISH) Validation**

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

343 *smFISH of adult cortical tissue*

344 Ten days after retro-AAV injections, mice were euthanized with isoflurane and
345 transcardially perfused with 0.9% NaCl with 10 units/ml heparin (Covetrus, 1000 unit/ml)
346 followed by 4% paraformaldehyde (PFA) in PBS. Brains and spinal cords were dissected and
347 post-fixed in 4% PFA overnight at 4°C. The next day, 15 µm thick sections were cut using a
348 vibratome (Leica microsystems, VT1000S), mounted on superfrost plus slides and processed for
349 smFISH according to the ACD RNAScope fluorescent protocol. Briefly, sections were either
350 immediately stored at -80°C until they were ready for processing or sections were post-fixed in
351 4% PFA and stained with a single probe: *Wnt7b* (ACDBio, Cat. #401131), *Cacng7* (ACDBio,
352 Cat. # 556171-C2), *Slc16a2* (ACDBio, Cat. # 545291). Each experiment included a positive and
353 negative control probe to ensure validity. Prior to coverslipping, immunofluorescence utilizing
354 primary antibodies directed against green fluorescent protein (GFP, 1:2000, Abcam, USA,
355 ab13970), and mCherry (1:2000, Abcam, USA, ab167453) and detected with secondary
356 antibodies Alexa Fluor-488 and -568, (1:500, 1:500 Life Technologies, Grand Island, NY,
357 A11039, A11011) was used to visualize traced CSNs.

358 *smFISH of adult cortical tissue- analysis*

359 All analyses were performed by an experimenter blinded to probe identity. For
360 quantification of smFISH results, images were taken at 40X on a Leica Sp8 confocal microscope
361 (LeicaMicrosystems). 3 images were taken per hemisphere, 3 sections were taken per animal and
362 3 animals were used for the analysis of each probe. For analysis, each image was separated into
363 individual channels (probe, forelimb and hindlimb CSN, and DAPI). The images of the traced
364 neurons were then exported to photoshop, where the quick selection tool was used to select cell
365 body regions of interest (ROI) and create masks. This was done manually as CSN axons can be

JNeurosci Accepted Manuscript

366 very thick and automatic ROI selection often cannot distinguish between CSN axons and cell
367 bodies. Images of the probe and the ROI masks are then uploaded to CellProfiler and puncta per
368 cell are automatically calculated according to the CellProfiler manual (McQuin et al., 2018).

369 *IHC protocol*

370 Ten days after retro-AAV injections, mice were euthanized with isoflurane and
371 transcardially perfused with 0.9% NaCl with 10 units/ml heparin (Covetrus, 1000 unit/ml)
372 followed by 4% paraformaldehyde in PBS. Brains and spinal cords were dissected and post-fixed
373 in 4% PFA overnight at 4°C. 40 µm thick sections of brain were then sectioned on a vibratome
374 (Leica Microsystems). The sections were then processed for mCherry with tyramide signal
375 amplification (Perkin Elmer, Waltham, MA, NEL700A001KT). Immunofluorescence utilized
376 primary antibodies directed against mCherry (1:500,000, Abcam, USA, ab167453),
377 Synaptophysin (1:10,000, EMD Millipore, MAB329), and NeuN (1:200, Cell Signaling, 24307)
378 and detected with secondary antibodies Alexa Fluor-488, -568, -647 (1:500, Life Technologies,
379 Grand Island, NY, A11039, A11011, A21235).

380 *BrainJ Processing*

381 Ten days after AAV injections, mice were euthanized with isoflurane and transcardially perfused
382 with 0.9% NaCl with 10 units/ml heparin (Covetrus, 1000 unit/ml) followed by 4%
383 paraformaldehyde (PFA) in PBS. Brains were dissected and post-fixed in 4% PFA overnight at
384 4°C. The next day, 75 µm thick sections were cut using a vibratome (Leica microsystems,
385 VT1000S). Sections underwent an immunohistochemistry protocol to enhance tdTomato
386 expression for microscopy. Briefly, sections received 3x washes of PBS-T, followed by a 10-
387 minute incubation in 0.3% H_2O_2 at room temperature, three more washes of PBS-T, and

JNeurosci Accepted Manuscript

388 blocking in 10% Normal Donkey Serum for 1 hour at room temperature. Next, sections were
389 incubated overnight in Rb-anti-mCherry (1:1,000). The next day, sections were washed 3x in
390 PBS-T and then incubated for 2 hours in 546 G-anti-Rb (1:1,000). After 3x final PBS-T washes,
391 every other section was mounted on SuperFrost Plus slides with DAPI mounting media (1
392 mg/ml, 1:10,000). Sections were imaged at 4x using a VS200 Olympus LS Slide Scanner for
393 projection analysis. BrainJ was used, as previously described (Botta et al., 2020), to perform
394 analysis of supraspinal forelimb and hindlimb ipsilateral and contralateral projections. Briefly,
395 sections were ordered and registered to the Allen Brain Atlas Common Coordinate Framework
396 adult mouse brain template using Elastix 5.0.1. Next, Ilastix 1.3.3.post3 was used to identify
397 brain-wide soma and projections in the ~80 mounted sections using pixel-based machine
398 learning. BrainJ output absolute density measurements for 1429 sub-regions of the brain, as
399 defined by the Kim Enhanced and Unified Mouse Brain Atlas (Chon et al., 2019). We then
400 classified each sub-region as ET, IT, NP, or CT using the Allen Brain Institute framework. A
401 mixed-design ANOVA was performed in R for each projection subtype to analyze differences
402 between forelimb-hindlimb and ipsilateral-contralateral projections. An alpha level of 0.05 was
403 set a priori and Bonferroni-adjusted post hoc tests were performed when main effects were
404 found.

405 **Code Accessibility**

406 All analyses were performed using in-house developed code and implemented in Python.
407 The CSN and limb-specific CSN gene lists along with example classifiers and appropriate
408 sample data and code can be found on GitHub: https://github.com/cafferty-
409 lab/CSN_classification

410

JNeurosci Accepted Manuscript

411 **RESULTS**

412 **A novel pipeline for scRNAseq of retrogradely labeled CSNs**

413 CSNs extend axons from L5b of M1 to innervate every spinal segment, synapsing
414 principally on spinal interneurons to control forelimb (FL) and hindlimb (HL) motor output.
415 Spinal retrograde tracing from C6/7 and L4/5 show that CSNs that innervate the FL are spatially
416 and anatomically distinct from those that innervate the HL, mirroring functional dexterity
417 differences and hinting that FL and HL CSNs are also molecularly heterogeneous (**Figs. 1A-D'**).
418 To explore the molecular heterogeneity among CSN populations, we completed scRNAseq on
419 retrogradely labeled FL and HL CSNs and compared gene expression between these
420 anatomically discrete populations to layer 2/3 and non-spinal projecting L5 neurons in M1 of
421 adult mice. As relatively large pyramidal neurons, CSNs have a lower nucleus to cytoplasm
422 ratio, hence a large portion of transcripts are extra-nuclear (Bakken et al., 2018). Therefore, it
423 was critical to maintain plasma membrane integrity to achieve maximal quality for whole-cell
424 sequencing. To that end, we created a pipeline for labeling, extracting, dissociating, and
425 sequencing adult neurons with long-distance axonal projections (Golan and Cafferty, 2021). To
426 independently label FL and HL CSNs, we microinfused retro-AAV-CAG-GFP into the grey
427 matter of either C6/7 or L4/5 spinal cord in adult mice (**Fig. 1E**). After ten days, we
428 macrodissected layers 2/3 and 5 of M1 (+1.0 → -0.5 mm relative to Bregma), enzymatically and
429 physically dissociated the tissue to create a single-cell suspension. Dissociated CSNs were then
430 incubated in a cytoplasmic dye to aid whole cell detection for subsequent fluorescence activated
431 cell sorting (FACS) (Golan and Cafferty, 2021). Intact CSNs were then collected for scRNAseq
432 using the 10x Chromium system. As not every CSN was fluorescently labeled due to inherent
433 inefficiencies with viral labeling, kernel density estimation (KDE) analysis was used to identify

JNeurosci Accepted Manuscript

434 putative CSNs from the non-fluorescent layer 5 samples based on gene expression. Control layer
435 5 cells with a greater than 50% probability of originating from the CSN distribution were
436 classified as putative CSNs and excluded from downstream analyses (**Fig. 1F**). To overcome
437 inexact microdissection of layer 2/3 neurons from layer 5 neurons, KDE analysis was used to
438 exclude putative layer 5 neurons from the layer 2/3 sample for downstream analysis (**G**).

439 Sequenced neurons had a mean library size of 19,432 unique molecular identifiers
440 (UMIs, median = 18,474 UMIs, **Fig. 2A**), and 5,357 genes per cell (median = 5,537). Previous
441 studies have detailed molecular differences between and among excitatory and inhibitory
442 neurons spanning the entire cortical depth (Arlotta et al., 2005; Lein et al., 2007; Morris et al.,
443 2010; Molyneaux et al., 2015; Tasic et al., 2016; Saunders et al., 2018; Tasic et al., 2018). Here
444 we focused specifically on molecular differences between layer 2/3 neurons and L5 neurons in
445 M1, as well as differences among L5 neurons, and comprehensively characterize transcriptional
446 specificity between CSNs. To achieve this, we compared gene expression between retrogradely
447 traced CSNs (n = 1776 cells), non-fluorescent non-spinal L5 neurons (n = 3000 cells), and layer
448 2/3 neurons (n = 719) in M1, sequencing a total of 5495 adult neurons.

449 Differential gene expression analysis of neurons from male and female mice revealed 32
450 differentially expressed (DE) genes in male mice and 213 DE genes in female mice (**Fig. 2B**).
451 QIAGEN IPA revealed that the synaptogenesis, protein ubiquitination and eNOS (endothelial
452 nitric oxide synthase) signaling pathways were differentially enriched between male and female
453 mice (**Fig. 2C**). To ensure there was no sex bias in this study, we assessed the expression of
454 female specific genes *Xist* and *Tsix* as well as the male-specific genes *Ddx3y* and *Eif3s3y* (**Fig.
455 2D**), showing female and male representation in all assessed populations. Next, expression of
456 *Ywhaz*, a gene whose mRNA is enriched in the cytoplasm (Yao et al., 2021), and *Neurod2*, a

457 nuclear gene, confirms that all sequenced cells contained both a nucleus and cytoplasm (**Fig.**
458 **2E**). To broadly confirm the identity of sequenced cells, we assessed their expression of
459 previously identified cell-type specific genes. All sequenced cells expressed the neuron-specific
460 gene *Snap25* and showed a dearth in expression for other cell type specific genes including the
461 endothelial-specific gene *Cldn5*, the astrocyte-specific gene *Gfap*, the microglial-specific gene
462 *Cx3cr1,* and the oligodendrocyte-specific gene *Olig2* (**Fig. 2F**). Neurotransmitter gene
463 expression confirmed that all sequenced neurons were enriched in the glutamate transporter
464 *Slc17a7* compared to the GABAergic gene *Gad1*, the cholinergic gene *Chat* or the serotonergic
465 gene *Sert* (**Fig. 2G**). Layer 2/3 neurons showed an elevated expression for the previously
466 identified layer 2/3-enriched gene *Cux2* (Tasic et al., 2016). CSNs and non-spinal L5 neurons
467 showed a moderate enrichment for established L5 genes *Ctip2, Fezf2,* and *Crym*, relative to other
468 cortical layer genes *Reln* (L1), and *Ctgf* (L6) (**Fig. 2H**), confirming L5 identity. While CSNs
469 expressed previously identified L5 genes, none were specific to any L5 population, suggesting
470 that more targeted analyses are necessary to identify unique molecular characteristics of CSNs.

471 **Identifying CSN-enriched genes in M1**

472 To identify CSN-enriched genes within M1, we performed three DE analyses: 1) traced
473 CSNs versus non-CSNs (non-spinal L5 neurons and layer 2/3 neurons), revealing 197 CSN
474 upregulated genes and 639 downregulated genes, 2) traced CSNs versus non-spinal L5 neurons,
475 revealing 215 CSN upregulated genes and 634 downregulated genes, and 3) traced CSNs versus
476 layer 2/3 neurons, revealing 467 CSN upregulated genes and 1533 downregulated genes (**Fig.**
477 **3A,** the entire DE lists have been included as an extended data table**, Fig. 3-1**). To explore the
478 biological import of the transcriptional differences among CSNs and other glutamatergic neurons
479 in M1, we completed QIAGEN IPA (Kramer et al., 2014). Consistently amongst the top enriched

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

480 pathways across all comparisons were those associated with axon guidance, signal transduction
481 and synaptic transmission (**Fig. 3B**). These pathways are consistent with previous work showing
482 that axon guidance cues including *Epha4, Efnb3, Ryk* and *Wnt3a* have previously been shown to
483 be critical in wiring the CST (Canty and Murphy, 2008).

484 To explore differences among these populations, we used potential of heat diffusion for
485 affinity-based transition embedding (PHATE) for dimensionality reduction and visualization, as
486 unlike PCA, t-SNE and uMAP, PHATE maintains both local and global structure, and was
487 designed for scRNAseq visualization (van Dijk et al., 2018; Moon et al., 2019). A PHATE
488 projection colored by cell origin (**Figs. 3C, Movie 1**) illustrates the transcriptional overlap
489 between these populations. To quantify, KDE analysis of each population was plotted, and a
490 similarity index was calculated based on the overlapping regions of the KDE (**Fig. 3D**). These
491 analyses demonstrate that CSNs are most similar to non-spinal L5 neurons (similarity index =
492 0.5), while they are least similar to layer 2/3 neurons (similarity index = 0.23). Non-spinal L5
493 neurons show an intermediate level of similarity with layer 2/3 neurons in comparison to CSNs
494 (similarity index = 0.41).

495 Based on enrichment of guidance cues in CSNs, we selected *Wnt7b* as a primary
496 candidate to validate our scRNAseq data as it is significantly enriched in all traced CSNs (**Fig.**
497 **3E)**. To confirm this finding *in vivo,* we infused retro-AAV-CAG-GFP into the grey matter of
498 C6/7 spinal cord and retro-AAV-CAG-tdTomato into the grey matter of L4/5 spinal cord in adult
499 wild type mice, differentially labeling FL and HL CSNs and completing single molecule
500 fluorescent *in situ* hybridization (smFISH, **Fig. 3F).** Photomicrographs of L5b from M1 prepared
501 2 weeks after spinal infusion show GFP+ FL CSNs and tdTomato+ HL CSNs are significantly
502 enriched in *Wnt7b* relative to DAPI+ control cells (**Figs. 3G, H**).

JNeurosci Accepted Manuscript

503 **Classifier analysis to Predict CSN identity**

504 Based on the high similarity index between CSNs and non-spinal L5 neurons, we
505 reasoned it would be unlikely that any single gene could uniquely identify CSNs from other L5
506 neurons. To that end, we first explored twelve genes that have previously been used to describe
507 L5 pyramidal neurons (**Fig. 4A**) (Arlotta et al., 2005; Fink et al., 2015; Tasic et al., 2016;
508 Economo et al., 2018; Tasic et al., 2018; Poplawski et al., 2020). Of these genes, many are
509 enriched in CSNs compared to other populations including *Ctip2***,** *Cd55*, *Cdh13*, *Crym* and *Fezf2*.
510 However, all are also expressed at moderate levels in non-traced L5 neurons and are only
511 detected in a subset of CSNs. Accordingly, PHATE projections for *Ctip2* and *Rbp4* show an
512 incomplete and nonspecific expression in CSNs and while *Wnt7b* shows greater specificity,
513 coverage remains incomplete (**Fig. 4B**). To improve upon these genes as means for CSN
514 identification, we sought to distinguish a CSN transcriptional signature. To that end, we trained a
515 linear classifier to predict cell identity based on expression of 100 genes that are most
516 differentially expressed between CSNs and non-CSNs (termed CSN classifier, CSN-C). For
517 comparison, we trained additional CSN classifiers based on the top 10 genes differentially
518 expressed, and a random 10 genes from the top 100 that were differentially expressed (**Fig. 4C**).
519 We compared the performance of these classifiers to one that was trained on sole expression of
520 all twelve previously characterized L5 genes (termed L5 classifier, L5-C), and another classifier
521 trained on a random set of 10 genes. The performance of each classifier was evaluated using the
522 area under the receiver operating characteristics curve (ROC AUC), and the classifiers were
523 compared using a permutation test. This revealed that CSN-C (AUC = 0.96), top 10 genes (AUC
524 = 0.95), random 10 CSN genes (AUC = 0.91) significantly outperform L5-C (AUC = 0.87.) (**Fig.**
525 **4C**). To visually compare the differences between these classifiers, we plotted CSN-C-classified

JNeurosci Accepted Manuscript

526 CSNs and L5-C classified CSNs in PHATE. CSN-C correctly classified 92% of CSNs while L5-
527 C classified 78% of CSNs (**Fig. 4D**).

528 **Classifier analysis to predict CSN identity of Rbp4+ neurons.**

529 To validate the use of the CSN-C as a tool for predicting CSN identity, we sequenced
530 neurons from *Rbp4* cre:Ai14 mice, a line previously used to comprehensively target and label L5
531 pyramidal neurons (Alcamo et al., 2008; Tasic et al., 2018). To quantify the proportion of Rbp4+
532 neurons that project to the spinal cord (*Rbp4+* CSNs), we delivered retro-AAV-CAG-GFP into
533 the cervical and lumbar cord of *Rbp4* cre:Ai14 mice (**Fig. 4E**). Photomicrographs show that a
534 minority of Rbp4+ neurons project to the spinal cord (**Figs. 4F, F'**). Quantification of the
535 proportion of Rbp4+ neurons that expressed GFP, and therefore have terminals in either the
536 cervical or lumbar spinal cord, revealed that on average, 7.7% of Rbp4+ neurons are also CSNs
537 (**Fig. 4G**). Classification of sequenced *Rbp4+* neurons using CSN-C predicted that 7% of Rbp4+
538 neurons were putative CSNs using a 75% probability cutoff (**Fig. 4H**). These findings confirm
539 the validity of our CSN gene list to predict CSN identity among L5 neurons in M1.

540 **Molecular identification of limb specific CSN subtypes**

541 We next wanted to explore potential transcriptional heterogeneity between anatomically
542 discrete FL and HL projecting CSNs. PHATE visualization of traced neurons colored by
543 anatomical origin illustrates that there is a transcriptional separation between FL and HL traced
544 cells (**Figs. 5A**). To quantify, a KDE of each population was plotted, and a similarity index was
545 calculated based on the overlapping regions of the KDE. FL and HL CSNs are somewhat similar
546 to each other (similarity index = 0.74). Owing to the elevated similarity index relative to the
547 more definitive separation between CSNs vs L5 controls and CSNs vs layer 2/3 (**Fig. 3D**), we

JNeurosci Accepted Manuscript

548 reasoned that a population of CSNs may have a projection to both the cervical and lumbar spinal
549 cord. Indeed, a small percentage of traced CSNs were double positive for GFP and tdTomato
550 when tracers were injected into the cervical and hindlimb respectively (**Figs. 1A-D'**). To explore
551 whether these cells represented a true population of FL/HL CSNs or whether *en passant* labeling
552 of HL axons occurred when infusing tracer into the cervical cord, we used MELD to identify
553 CSNs whose transcriptomic characteristics were intermediate between the FL and HL
554 populations as previously identified by cervical and lumbar tracing (**Figs. 5B, 5C**). We found
555 188 such CSNs, of which 60 had been previously identified as FL, and 128 as HL (**Fig. 5D**).
556 These data suggest that as much as 10% of CSNs may have a projection into both the cervical
557 and lumbar spinal cord.

558 DE analysis revealed that there were 492 FL-enriched genes and 788 HL-enriched genes
559 (**Fig. 5E,** the entire DE lists have been included as an extended data table**, Fig. 5-1**). Identifying a
560 set of enriched FL and HL CSN genes is central to understanding the potential independent
561 homeostatic biology and injury-induced changes within these discrete populations. To prioritize
562 candidate genes, recursive feature elimination was used to determine the top ten FL- and HL-
563 specific genes that most robustly distinguished these two populations. The top genes that define
564 the FL population include calcium related genes *Cacng7* and *Efcab12,* as well as transporters
565 *Slc27a1* and *Slc4a2* (**Fig. 5F**). The top genes that define the HL population include ion
566 transporters such as *Slc16a2,* transcription-related genes such as *Zfp281*, and receptors such as
567 *Ptger1* (**Fig. 5F**).

568 PHATE projections show *Cacng7* expression is significantly enriched in FL traced
569 neurons (**Fig. 5G**) and *Slc16a2* expression is significantly enriched in HL traced neurons (**Fig.**
570 **5H**). To validate these results *in vivo*, mice received injections of retro-AAV-CAG-GFP into

JNeurosci Accepted Manuscript

571 spinal grey matter at C6/7 and retro-AAV-CAG-tdTomato into spinal grey matter at L4/5.
572 Visualization of *Cacng7* using smFISH confirmed differential expression in GFP+ FL-projecting
573 CSNs compared to tdTomato+ HL-projecting CSNs (**Figs. 5I, J**). A similar analysis was then
574 repeated with the HL enriched gene *Slc16a2*. *In vivo* smFISH confirmed HL-enrichment of
575 *Slc16a2* in tdTomato+ HL-projecting CSNs relative to GFP+ FL-projecting neurons (**Figs. 5K,**
576 **L**).

577 **Classifier analysis to Predict FL and HL identity**

578 While DE analysis revealed FL and HL-enriched genes, due to the high similarity index
579 between FL and HL CSN populations, none of these genes can independently predict FL or HL
580 identity. To explore, we first examined limb-specificity in the previously identified L5-enriched
581 genes (**Fig. 5M**). *Crym,* is modestly enriched in FL-CSNs (Fink et al., 2015), and *Crim1* was
582 modestly enriched in HL-CSNs (Sahni et al., 2021a; Sahni et al., 2021b). As the majority of
583 these genes showed no limb specificity, we trained a linear classifier to predict limb specificity
584 based on expression of 100 genes that are most differentially expressed between FL and HL
585 CSNs (termed limb-specific CSN classifier, LS-CSN-C). We compared the performance of this
586 classifier to L5-C. We evaluated the performance of each classifier with ROC AUC and the two
587 classifiers were compared using a permutation test. This revealed that LS-CSN-C (AUC = 0.94)
588 significantly outperforms L5-C (AUC = 0.76.) (**Fig. 5N**).

589 **CSN subtypes defined by supraspinal connectivity**

590 Recent studies from the BICCN (Brain Initiative Cell Consensus Network) (Network,
591 2021) performing scRNAseq on M1 and anterolateral motor cortex (ALM) have classified
592 glutamatergic projection neurons into four broad subtypes: intra-telencephalic (IT) neurons that

593 have many projection targets including the striatum and cerebral cortex (Lin et al., 2018; Zhang
594 et al., 2021), extra-telencephalic (ET) neurons that project from M1 to the brainstem and medulla
595 (Economo et al., 2018; Tasic et al., 2018; Yao et al., 2021), near-projecting (NP) neurons that
596 have projections in local areas including M1, M2 and S1 (Tasic et al., 2018), and corticothalamic
597 (CT) neurons (Yao et al., 2021) (**Fig. 6A**). 10x-sequenced deep layer glutamatergic neurons
598 collected in Yao *et al*., separate into 4 discrete populations representing the IT, ET, NP, and CT
599 subtypes when examined in PHATE (**Fig. 6B**). CSNs collected in this study, when projected into
600 the same PHATE graph populate all 4 subtypes and therefore we classified them as CSN-IT,
601 CSN-ET, CSN-NP, and CSN-CT (**Figs. 6C, D, Movie 2)**. Differential gene expression analysis
602 among these 4 putative sub-divisions of CSNs revealed an enrichment for previously identified
603 projection subtype genes: *Fam84b* for CSN-ET cells*,* *Slc30a3* in CSN-IT cells, *Foxp2* in CSN-
604 CT, and *Sla2* in CSN-NP cells (Tasic et al., 2018; Yao et al., 2021) (**Fig. 6E**), and an additional
605 set of sub-category specific genes (**Figs. 6E, 6F**). We completed QIAGEN IPA on genes
606 significantly differentially expressed (SDE) among these CSN sub-categories (**Figs. 6G-J**).
607 When comparing the CSN-ET population to the other 3 categories we found that there were 972
608 DE genes showing an enrichment in pathways associated with synaptic signaling and
609 neurovascular coupling (**Fig. 6G**). Comparing CSN-IT cells to the other 3 categories we found
610 that there were 539 DE genes showing an enrichment in pathways associated with axon guidance
611 and CREB signaling (**Fig. 6H**). Comparing CSN-CT cells to the other categories we found that
612 there were 656 DE genes showing an enrichment in axon guidance and RhoGDI signaling (**Fig.**
613 **6I**). Comparing CSN-NP cells to the other categories we found that there were 1735 DE genes
614 showing an enrichment in axon guidance and cAMP signaling and GPCR signaling (**Fig. 6J**).
615 The entire DE lists have been included as an extended data table**, Fig. 6-1**.

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

616 As CSNs are represented in all four projection subtype categories, we infer that
617 transcriptional subdivisions among CSNs are therefore defined by the spinal termination of their
618 primary axon in the spinal cord and by the location of their supraspinal terminal fields (**Figs. 6K-**
619 **N**).

620 To explore CSN supraspinal connectivity *in vivo*, we injected retro-AAV-CAG-tdTomato
621 unilaterally into the grey matter of C6/7 and L4/5 spinal cord and surveyed for the presence of
622 tdTomato+ve CSN terminals in the BICCN identified supraspinal structures (**Figs. 6A, and 6K-**
623 **N**). We observed td-Tomato +ve CSN terminals in the brainstem (**Figs. 6K, K'**), the
624 caudoputamen of the striatum (**Figs. 6L, L'**), the motor thalamus (**Figs. 6M, M'**), and regions
625 near M1 including S1 and M2 (**Figs. 6N, N'**) confirming that CSNs project to both the spinal
626 cord and to these supraspinal structures. To further explore whether these collaterals represented
627 putative active synapses, we completed immunohistochemistry, showing that traced CSN
628 terminals express the presynaptic protein synaptophysin and colocalize with NeuN-positive
629 neurons.

630 **Limb-specific specialization among supraspinal CSN subtypes**

631 We found that traced CSNs maintained a supraspinal cluster separation when plotted in their
632 native PHATE space (**Figs. 7A, Movie 3**), and FL and HL CSNs were represented in each
633 supraspinal sub-category (**Figs. 7B, 5A-1**). However, the relative distribution of CSN sub-
634 category distribution was different between FL and HL. For FL CSNs 19.6% were ET, 36.5% IT,
635 40% CT, and 3.8% NP. HL CSNs were 37.6% ET, 31.5% IT, 26.8% CT, and 4% NP (**Figs. 7C,**
636 **D**). These data suggest that FL and HL CSNs may differentially innervate supraspinal structures.
637 To explore we used intersectional viral tracing to independently label all CSN terminals whose

638 axons innervate either cervical or lumbar spinal cord. We injected retroAAV-CAG-Cre into
639 either C5/6 or L4/5 spinal cord and AAV-CAG-FLEX-tdTomato into FL M1 or HL M1
640 respectively. Using the Java plugin, BrainJ (Botta et al., 2020), we analyzed the density of
641 tdTomato +ve CST projections across the entire CST. We found that the projection density was
642 higher on the ipsilateral side compared to the contralateral side (mixed model ANOVA,
643 $p<0.001$). We also found that the FL-CST had higher density of projections compared to the HL-
644 CST (mixed model ANOVA $p=0.008$), consistent with prior reports that FL-CSNs are more
645 numerous than HL-CSNs (Wang et al., 2022). Next, we assessed FL and HL-CSN supraspinal
646 projection density. We found an enrichment of FL-CSN-ET projections compared to HL-CSN-
647 ET projections (**Figs. 7E, 7F**, mixed model ANOVA $p=0.003$), and FL-CSN-NP projections
648 compared to HL-CSN-NP projections (**Figs. 7Q, 7R**, mixed model ANOVA $p<0.001$). By
649 contrast, we found no significant differences between FL and HL for IT (**Figs. 7I, 7J**, $p=0.38$)
650 and CT (**Figs. 7M, 7N**, $p=0.085$) projections. Overall, these data demonstrate the supraspinal
651 anatomical diversity of FL- and HL-CSNs and suggest that the transcriptional differences among
652 these population support unique circuit wiring.

653 **CSN Extra-telencephalic (ET) specialization**

654 These data reveal that supraspinal structures are differentially innervated by FL- and HL-
655 CSNs. To explore the transcriptional differences among these populations that may drive and/or
656 support the unique supraspinal terminal distribution within these sub-divisions we completed a
657 differential gene expression analysis among FL and HL ET, IT, CT, and NP categories.

658 There were 964 genes upregulated and 484 genes downregulated in FL-CSN-ET cells
659 compared to FL -IT, -CT, and -NP combined (**Fig. 7H**). A more granular comparison among
660 sub-categories revealed that between FL-ET and FL-IT 477 genes were upregulated and 1089

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

661 downregulated, QIAGEN IPA revealed that GPRC signaling and CREB signaling were among
662 the top pathways dysregulated. Between FL-ET and FL-CT 465 genes were upregulated and
663 1065 genes were downregulated, with GPCR signaling and RHOGDI signaling among the top
664 enriched pathways. Between FL-ET and FL-NP 630 genes were upregulated and 1808 genes
665 were downregulated, with CREB signaling Gαi signaling among the dysregulated pathways
666 (**Fig. 7H**).

667 For the HL-CSN-ET population 807 genes were upregulated, and 588 genes
668 downregulated in HL-CSN-ET cells compared to HL – IT, -CT, and -NP combined (**Fig. 7H'**).
669 Comparisons among the sub-categories revealed that between HL-ET and HL-IT 538 genes were
670 upregulated and 821 genes were downregulated, with Axon Guidance signaling and CREB
671 signaling among the top pathways dysregulated. Between HL-ET and HL-CT 735 genes were
672 upregulated and 1065 genes were downregulated, with Gai signaling and RHOGDI signaling
673 among the top enriched pathways. Between HL-ET and HL-NP 1234 genes were upregulated
674 and 2193 genes were downregulated, with spliceosomal cycle signaling and EIF2 signaling
675 among the top pathways (**Fig. 7H'**).

676 **CSN Intra-telencephalic (IT) specialization**

677 We found that 282 genes were upregulated, and 228 genes downregulated in FL-CSN-IT
678 cells compared to FL -ET, -CT, and -NP combined (**Fig. 7L**). A more specific comparison
679 among sub-categories revealed that between FL-IT and FL-ET 1089 genes were upregulated and
680 477 genes were downregulated, with Valine degradation and Isoleucine degradation among the
681 top dysregulated pathways. Between FL-IT and FL-CT 234 genes were upregulated and 359
682 genes were downregulated, with Rho Family signaling and SEMA signaling among the top
683 dysregulated pathways. Between FL-IT and FL-NP 820 genes were upregulated and 1155 genes

684 were downregulated, with Glutamate Receptor Signaling and Putrescine Degradation among
685 dysregulated pathways (**Fig. 7L**).

686 For the HL-CSN-IT population 388 genes were upregulated, and 488 genes
687 downregulated in HL-CSN-IT cells compared to HL – ET, -CT, and -NP combined (**Fig. 7L'**).
688 Comparisons among the sub-categories revealed that between HL-IT and HL-ET 821 genes were
689 upregulated and 538 genes were downregulated, with Axon Guidance signaling and Valine
690 degradation among the top dysregulated pathways. Between HL-IT and HL-CT 523 genes were
691 upregulated and 749 genes were downregulated, with SEMA Signaling and RHOGDI Signaling
692 among the top enriched pathways. Between HL-IT and HL-NP 1377 genes were upregulated and
693 1758 genes were downregulated, with Zymosterol Biosynthesis and Putrescine degradation
694 among the top pathways (**Fig. 7L'**).

695 **CSN Cortico-thalamic (CT) specialization**

696 We found that 201 genes upregulated, and 386 genes downregulated in FL-CSN-CT cells
697 compared to FL -ET, -IT, and -NP combined (**Fig. 7P**). A more specific comparison among sub-
698 categories revealed that between FL-CT and FL-ET 1065 genes were upregulated and 465 genes
699 were downregulated, with Circadian Rhythm Signaling and Synaptogenesis signaling among the
700 top pathways dysregulated. Between FL-CT and FL-IT 359 genes were upregulated and 234
701 genes were downregulated, with Axon Guidance signaling and CREB Signaling in Neurons
702 among the top dysregulated pathways. Between FL-CT and FL-NP 757 genes were upregulated
703 and 1068 genes were downregulated, with Glutamate receptor signaling and GABA receptor
704 Signaling among the dysregulated pathways (**Fig. 7P**).

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

705 For the HL-CSN-CT population 425 genes were upregulated, and 805 genes
706 downregulated in HL-CSN-IT cells compared to HL – ET, -CT, and -NP combined (**Fig. 7P'**).
707 Comparisons among the sub-categories revealed that between HL-CT and HL-ET 1370 genes
708 were upregulated and 735 genes were downregulated, with Synaptogenesis Signaling and
709 Serotonin Receptor Signaling among the top pathways dysregulated. Between HL-CT and HL-IT
710 749 genes were upregulated and 523 genes were downregulated, with Circadian Rhythm
711 Signaling and the Opioid signaling Pathway among the top enriched pathways. Between HL-CT
712 and HL-NP 1109 genes were upregulated and 1500 genes were downregulated, with Valine
713 Degradation and Glutamate Receptor Signaling among the top dysregulated pathways (**Fig. 7P'**).

714 **CSN Near Projecting (NP) specialization**

715 We found that 709 genes upregulated, and 1163 genes downregulated in FL-CSN-NP
716 cells compared to FL -ET, -IT, and -CT combined (**Fig. 7T**). A more specific comparison among
717 sub-categories revealed that between FL-NP and FL-ET 1808 genes were upregulated and 630
718 genes were downregulated, with GADD45 signaling and Senescence signaling among the top
719 pathways dysregulated. Between FL-NP and FL-IT 1155 genes were upregulated and 630 genes
720 were downregulated, with cAMP signaling and GPCR Signaling among the top dysregulated
721 pathways. Between FL-NP and FL-CT 1068 genes were upregulated and 757 genes were
722 downregulated, with cAMP signaling and GPCR Signaling among the top dysregulated
723 pathways. (**Fig. 7T**).

724 For the HL-CSN-NP, 1096 genes were upregulated, and 1706 genes were downregulated
725 in HL-CSN-NP cells compared to HL – ET, -IT, and -CT combined (**Fig. 7T'**). Comparisons
726 among the sub-categories revealed that between HL-NP and HL-ET 2193 genes were
727 upregulated and 1234 genes were downregulated, with Molecular Mechanisms in Cancer

728 Signaling and Axon Guidance Signaling among the top pathways dysregulated. Between HL-NP
729 and HL-IT 1758 genes were upregulated and 1377 genes were downregulated, with Axon
730 Guidance Signaling and the Serotonin Receptor Signaling Pathway among the top enriched
731 pathways. Between HL-NP and HL-CT 1500 genes were upregulated and 1109 genes were
732 downregulated, with GPCR Signaling and CREB Signaling in Neurons among the top
733 dysregulated pathways (**Fig. 7T'**). The entire DE lists have been included as extended data
734 tables**, Fig. 7-1A-D**.

735

736

737

738

739

740

741

742

743

744

745

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

746 **DISCUSSION**

747 In this study, we used a combination of viral retrograde labeling from the cervical and
748 lumbar spinal cord, FACS and scRNAseq to annotate a comprehensive transcriptional index of
749 adult CSNs. Using a refined extraction, dissociation, and FACS pipeline, we were able to
750 maintain cytoplasmic integrity of large diameter pyramidal neurons for whole cell sequencing
751 (**Fig.1**) (Golan and Cafferty, 2021) of intact adult CSNs (**Figs. 2, 3**). Using this approach, we
752 identified 849 genes that were significantly differentially expressed between CSNs and non-
753 spinally projecting L5 neurons. We built a linear classifier to confirm the veracity of this
754 constellation of genes to identify CSNs with improved fidelity over a classifier trained on a set of
755 established L5 genes (**Fig. 4C**). Furthermore, we found that 1280 genes were differentially
756 expressed between FL and HL CSNs. These data allowed us to train a limb-specific CSN linear
757 classifier based on the top 100 DE genes between FL and HL **(Fig. 5N)**. These gene lists can be
758 used in combination to identify FL and HL CSNs from a mixed population of neurons in any
759 single cell sequencing data set from the intact adult mouse brain. Additionally, we found that
760 adult CNSs do not fit exclusively into the BICCN-defined extra-telencephalic category (also
761 known as PT, pyramidal tract) (Network, 2021). Rather, they express a profile aligning with all
762 non-spinally projecting categories IT, CT and NP (**Figs. 6A-D**). Furthermore, these new CSN
763 sub-categories CSN-ET, CSN-IT, CSN-CT, and CSN-NP each display a unique gene expression
764 profile (**Figs. 6E-J**). Finally, we confirmed that CSNs innervate supraspinal structures consistent
765 with projection categories ET, IT, CT and NP, highlighting the capacity of the CST to influence
766 supraspinal motor processing.

767

768 **Molecular specialization among forelimb and hindlimb CSNs**

JNeurosci Accepted Manuscript

769 CSN identity has been explored during development by combining retrograde tracing
770 from the pons and upper cervical spinal cord postnatally with temporal bulk microarray (Arlotta
771 et al., 2005) and RNA sequencing (Molyneaux et al., 2015) of FACS-purified traced neurons.
772 Authors identified *Ctip2*, *Crym*, and *Crim1* as CSN-specific genes up to P14, leaving the CSN
773 transcriptome temporally and spatially incomplete. Building on these transformative studies, we
774 combined spinal retrograde labeling with scRNAseq to comprehensively profile FL and HL
775 CSNs. We were able to identify and anatomically confirm robust CSN-specific and limb-
776 specific genetic signatures that expand beyond previously identified single genes. Our findings
777 revealed a novel constellation of genes which can be used to assign CSN identity with greater
778 fidelity/confidence over previously identified CSN-markers.
779 Indeed, the power of CSN subtype identity was recently highlighted in a study that
780 sought to explore transcriptional divergence among rostral (FL) and caudal (HL) projecting
781 CSNs during postnatal patterning. Authors identified *Klhl14* and *Crim1* as crucial for segment-
782 specific axon targeting by completing bulk microarray sequencing of neurons in the rostral
783 forelimb area (RFA) and comparing them to neurons in caudal forelimb area (CFA) and
784 hindlimb M1 regions (Sahni et al., 2021a; Sahni et al., 2021b). To elaborate upon the findings,
785 we explored limb-specific differences in the adult. Neither *Klhl14* or *Crim1* are significantly
786 enriched in FL or HL CSNs in adulthood, likely reflecting downregulation of patterning
787 machinery in the adult. However, here, CSNs were harvested for scRNAseq from +1.0 $\rightarrow$ -0.5
788 mm relative to Bregma, potentially under-sampling CSNs from the RFA.
789
790 **Functional specialization among forelimb and hindlimb CSNs**

JNeurosci Accepted Manuscript

791 Functional subtypes among M1 CSNs were previously described using slice
792 electrophysiology and morphological analyses (Oswald et al., 2013). Unbiased clustering found
793 that amongst cortical neurons, FL- and HL-projecting CSNs fired action potentials at the fastest
794 rate and with the most regular pattern, and that burst firing was exclusive to FL CSNs. This is
795 consistent with FL-enriched expression of *Cacng7*, a type II transmembrane AMPA receptor
796 regulatory protein that enhances glutamate-evoked currents from glutamate receptors (Kato et
797 al., 2007).
798 There is also evidence that FL- and HL CSNs play separate roles in sensorimotor
799 function. A recent study showed that the lumbar CST functions exclusively to gate sensory input
800 via primary afferent depolarization, and CST-mediated motor commands to the lumbar spinal
801 cord are not direct, but rather use polysynaptic circuits in the upper cervical cord (Moreno-Lopez
802 et al., 2020). Our data supports this finding as mice deficient in the HL-enriched gene *Slc16a2* do
803 not display gross motor abnormalities but do exhibit shorter latency paw withdrawal during a hot
804 plate test, suggestive of hyperalgesia (Wirth et al., 2009; Moreno-Lopez et al., 2020).
805 This spinal enlargement level transcriptional resolution resolves broad FL and HL-level
806 functional differences; however, there could also be within and between spinal segment
807 specificity. Indeed, it was recently shown that different phases of FL movement are coordinated
808 by CSNs projecting to discrete spinal segments (Ueno et al., 2018).
809
810 **Plasticity of forelimb and hindlimb CSNs**
811 Data from our laboratory supports the notion that FL and HL CSNs require independent
812 interventions to stimulate functional axon growth. Previously, we completed bulk RNAseq of
813 retrogradely traced intact CSNs after contralateral PyX-induced functional sprouting into the

JNeurosci Accepted Manuscript

814 denervated side. DE analyses revealed that members of the LPAR1-interacting pathway and the
815 3-phosphoinositide degradation pathway were enriched in sprouting CSNs. Viral overexpression
816 of candidate genes within these pathways in intact CSNs after PyX resulted in significant
817 sprouting of intact CST axons into the denervated side of the spinal cord (Fink et al., 2017;
818 Kauer et al., 2022). Assessment of fine motor skills showed that the lesioned FL performed
819 better than the lesioned HL, suggesting that our treatments were having a greater impact in FL
820 CSNs. These data support limb-specific therapeutic interventions, as the pro-growth factors we
821 identified in our original *in vivo* screen were identified via retrograde tracing from the cervical
822 spinal cord. We hypothesize that tracing from the lumbar cord after contralateral PyX may reveal
823 a unique set of pro-axon growth modulators.
824

825 **CSNs segregate by supraspinal collateralization**

826 In addition to FL and HL specialization, CSNs can also be classified based on their
827 supraspinal projections. We found that FL-CSNs innervate ET and NP structures more densely
828 than HL-CSNs. Furthermore, each of these sub-divisions can also be defined transcriptionally
829 suggesting that they each influence fine motor processing in a previously under-appreciated
830 fashion. Their expression of a range of axon guidance ligands and receptors, suite of G-Protein
831 Coupled Receptors and Synaptogenesis Signaling Components suggest they may each support a
832 unique response to injury, disease, and aging. In alignment with these findings previous work
833 using scRNAseq of adult motor regions established a taxonomy of neuronal cell types (Tasic et
834 al., 2016; Tasic et al., 2018; BICCN, 2021; Yao et al., 2021). To describe the diversity observed
835 among neuronal cell types spanning the cortical depth in M1, cell-type specific genes describing
836 IT, ET, NP, and CT neurons were identified (Yao et al., 2021). The ET designation was

JNeurosci Accepted Manuscript

837 confirmed in part by retrograde tracing from the pons and medulla and scRNAseq of ALM
838 neurons (Economo et al., 2018; Tasic et al., 2018). However, these studies did not include
839 tracing from the spinal cord, thus rendering a full ET atlas incomplete. Recent studies have
840 combined scRNAseq with single-cell reconstructions from sparsely labeled L5 neurons. These
841 studies identified a subtype of L5 ET neurons with projections in the spinal cord, in addition to
842 many supraspinal regions (BICCN, 2021; Peng et al., 2021). Furthermore, axonal barcoding of
843 the CST revealed that many CSNs innervate a range of supraspinal structures in addition to their
844 terminal fields in the spinal cord (Hausmann et al., 2022). These data are consistent with our
845 findings that CSNs terminate in multiple supraspinal locations. By aligning sequenced CSNs
846 with previously published M1 sequencing datasets, we found that CSNs are not all characterized
847 solely as ET cells, but rather span all four designated subtypes.

848 There is also evidence that CSN supraspinal connectivity reflects functional differences
849 among CSNs. Studies exploring FL-CSNs with supraspinal collaterals in the striatum found that
850 even among this one CSN subclass, there was anatomical and functional heterogeneity. Using
851 intersectional polysynaptic tracing, authors found that CSNs synapse onto both D1 and D2
852 neurons in the striatum, and GAD2 and calbindin D28K interneurons in the spinal cord. By
853 combining two-photon imaging of FL-CSNs labeled via retro-AAV-GCaMP6f infusion into the
854 cervical cord with a lever pulling task, authors found that striatal CSNs display three unique
855 activity patterns during: pulling onset, pulling offset and throughout the task (Nelson et al.,
856 2021). This study provides robust evidence that supraspinal CST connectivity is critical to
857 refining motor function, further necessitating a need for a comprehensive evaluation of the
858 functional phenotypes of these molecularly divergent CSN subtypes.

859 In summary, from these data we conclude that CSNs in M1 are molecularly unique
860 among non-projection layer 2/3 and layer 5 neurons and can be further sub-divided based on the
861 location of their spinal rostrocaudal and supraspinal terminal arbors. Leveraging this molecular
862 heterogeneity provides a powerful opportunity to explore the transcriptional machinery that
863 drives development, maintains homeostasis, and actuates reactive changes after trauma and
864 disease within this critical motor pathway. Comprehensive transcriptional insight in these
865 physiological and maladaptive processes will enable refined therapeutic interventions to repair
866 the damaged CST with enhanced precision.

867

868

869

870

871

872

873

874

875

876

877

878

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

879 **BIBLIOGRAPHY**

880 Alcamo EA, Chirivella L, Dautzenberg M, Dobreva G, Farinas I, Grosschedl R, McConnell SK
881 (2008) Satb2 regulates callosal projection neuron identity in the developing cerebral
882 cortex. Neuron 57:364-377.

883 Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD (2005) Neuronal subtype-
884 specific genes that control corticospinal motor neuron development in vivo. Neuron
885 45:207-221.

886 Bakken TE et al. (2018) Single-nucleus and single-cell transcriptomes compared in matched
887 cortical cell types. PLoS One 13:e0209648.

888 BICCN (2021) A multimodal cell census and atlas of the mammalian primary motor cortex.
889 Nature 598:86-102.

890 Blackmore MG, Moore DL, Smith RP, Goldberg JL, Bixby JL, Lemmon VP (2010) High
891 content screening of cortical neurons identifies novel regulators of axon growth. Mol Cell
892 Neurosci 44:43-54.

893 Botta P, Fushiki A, Vicente AM, Hammond LA, Mosberger AC, Gerfen CR, Peterka D, Costa
894 RM (2020) An Amygdala Circuit Mediates Experience-Dependent Momentary Arrests
895 during Exploration. Cell 183:605-619 e622.

896 Bradbury EJ, Burnside ER (2019) Moving beyond the glial scar for spinal cord repair. Nat
897 Commun 10:3879.

JNeurosci Accepted Manuscript

898 Buchser WJ, Slepak TI, Gutierrez-Arenas O, Bixby JL, Lemmon VP (2010) Kinase/phosphatase
899 overexpression reveals pathways regulating hippocampal neuron morphology. Mol Syst
900 Biol 6:391.

901 Burkhardt DB, Stanley JS, 3rd, Tong A, Perdigoto AL, Gigante SA, Herold KC, Wolf G,
902 Giraldez AJ, van Dijk D, Krishnaswamy S (2021) Quantifying the effect of experimental
903 perturbations at single-cell resolution. Nat Biotechnol 39:619-629.

904 Canty AJ, Murphy M (2008) Molecular mechanisms of axon guidance in the developing
905 corticospinal tract. Prog Neurobiol 85:214-235.

906 Chon U, Vanselow DJ, Cheng KC, Kim Y (2019) Enhanced and unified anatomical labeling for
907 a common mouse brain atlas. Nat Commun 10:5067.

908 Economo MN, Viswanathan S, Tasic B, Bas E, Winnubst J, Menon V, Graybuck LT, Nguyen
909 TN, Smith KA, Yao Z, Wang L, Gerfen CR, Chandrashekar J, Zeng H, Looger LL,
910 Svoboda K (2018) Distinct descending motor cortex pathways and their roles in
911 movement. Nature 563:79-84.

912 Farrar MJ, Bernstein IM, Schlafer DH, Cleland TA, Fetcho JR, Schaffer CB (2012) Chronic in
913 vivo imaging in the mouse spinal cord using an implanted chamber. Nat Methods 9:297-
914 302.

915 Fawcett JW (2020) The Struggle to Make CNS Axons Regenerate: Why Has It Been so
916 Difficult? Neurochem Res 45:144-158.

917 Fink KL, Strittmatter SM, Cafferty WB (2015) Comprehensive Corticospinal Labeling with mu-
918 crystallin Transgene Reveals Axon Regeneration after Spinal Cord Trauma in ngr1-/-
919 Mice. J Neurosci 35:15403-15418.

920 Fink KL, Lopez-Giraldez F, Kim IJ, Strittmatter SM, Cafferty WBJ (2017) Identification of
921 Intrinsic Axon Growth Modulators for Intact CNS Neurons after Injury. Cell Rep
922 18:2687-2701.

923 Fischer D, Hölzlwimmer, F. R. (2021) Fast and scalable differential expression analysis on
924 single-cell RNA-seq data. GitHub Repository.

925 Gianino S, Stein SA, Li H, Lu X, Biesiada E, Ulas J, Xu XM (1999) Postnatal growth of
926 corticospinal axons in the spinal cord of developing mice. Brain Res Dev Brain Res
927 112:189-204.

928 Gigante S, Krishnaswamy, S. (2021) scprep. GitHub Repository.

929 Golan N, Cafferty WB (2021) Dissociation of intact adult mouse cortical projection neurons for
930 single-cell RNA-seq. STAR Protoc 2:100941.

931 Gutilla EA, Steward O (2016) Selective neuronal PTEN deletion: can we take the brakes off of
932 growth without losing control? Neural Regen Res 11:1201-1203.

933 Hausmann FS, Barrett JM, Martin ME, Zhan H, Shepherd GMG (2022) Axonal Barcode
934 Analysis of Pyramidal Tract Projections from Mouse Forelimb M1 and M2. J Neurosci
935 42:7733-7743.

936 He Z, Jin Y (2016) Intrinsic Control of Axon Regeneration. Neuron 90:437-451.

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

937 Hilton BJ, Bradke F (2017) Can injured adult CNS axons regenerate by recapitulating
938 development? Development 144:3417-3429.

939 Kato AS, Zhou W, Milstein AD, Knierman MD, Siuda ER, Dotzlaf JE, Yu H, Hale JE,
940 Nisenbaum ES, Nicoll RA, Bredt DS (2007) New transmembrane AMPA receptor
941 regulatory protein isoform, gamma-7, differentially regulates AMPA receptors. J
942 Neurosci 27:4969-4977.

943 Kauer SD, Fink KL, Li EHF, Evans BP, Golan N, Cafferty WB (2022) Inositol Polyphosphate-5-
944 phosphatase K (Inpp5k) enhances sprouting of corticospinal tract axons after
945 CNS trauma. The Journal of Neuroscience:JN-RM-0897-0821.

946 Kramer A, Green J, Pollard J, Jr., Tugendreich S (2014) Causal analysis approaches in Ingenuity
947 Pathway Analysis. Bioinformatics 30:523-530.

948 Lein ES et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature
949 445:168-176.

950 Lemon RN (2008) Descending pathways in motor control. Annu Rev Neurosci 31:195-218.

951 Lin HM, Kuang JX, Sun P, Li N, Lv X, Zhang YH (2018) Reconstruction of Intratelencephalic
952 Neurons in the Mouse Secondary Motor Cortex Reveals the Diverse Projection Patterns
953 of Single Neurons. Front Neuroanat 12:86.

954 Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR,
955 Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A,

956 McCarroll SA (2015) Highly Parallel Genome-wide Expression Profiling of Individual
957 Cells Using Nanoliter Droplets. Cell 161:1202-1214.

958 Martin JH (2005) The corticospinal system: from development to motor control. Neuroscientist
959 11:161-173.

960 McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, Doan M, Ding
961 L, Rafelski SM, Thirstrup D, Wiegraebe W, Singh S, Becker T, Caicedo JC, Carpenter
962 AE (2018) CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol
963 16:e2005970.

964 Milich LM, Choi J, Ryan C, Yahn SL, Tsoulfas P, Lee JK (2020) Single cell analysis of the
965 cellular heterogeneity and interactions in the injured mouse spinal cord.
966 bioRxiv:2020.2005.2013.094854.

967 Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Hrvatin S, Rinn JL, Arlotta P (2015) DeCoN:
968 genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate
969 selection in neocortex. Neuron 85:275-288.

970 Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, Yim K, Elzen AVD, Hirn
971 MJ, Coifman RR, Ivanova NB, Wolf G, Krishnaswamy S (2019) Visualizing structure
972 and transitions in high-dimensional biological data. Nat Biotechnol 37:1482-1492.

973 Moreno-Lopez Y, Bichara C, Isope P, Cordero-Erausquin M (2020) Sensory gating precedes
974 motor command in evolution of the corticospinal tract. bioRxiv:2020.2001.2006.895912.

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

975 Morris JA et al. (2010) Divergent and nonuniform gene expression patterns in mouse brain. Proc
976 Natl Acad Sci U S A 107:19049-19054.

977 Munoz-Castaneda R et al. (2021) Cellular anatomy of the mouse primary motor cortex. Nature
978 598:159-166.

979 Nelson A, Abdelmesih B, Costa RM (2021) Corticospinal populations broadcast complex motor
980 signals to coordinated spinal and striatal circuits. Nat Neurosci 24:1721-1732.

981 Network BICC (2021) A multimodal cell census and atlas of the mammalian primary motor
982 cortex. Nature 598:86-102.

983 Oswald MJ, Tantirigama ML, Sonntag I, Hughes SM, Empson RM (2013) Diversity of layer 5
984 projection neurons in the mouse motor cortex. Front Cell Neurosci 7:174.

985 Park KK, Liu K, Hu Y, Smith PD, Wang C, Cai B, Xu B, Connolly L, Kramvis I, Sahin M, He Z
986 (2008) Promoting axon regeneration in the adult CNS by modulation of the PTEN/mTOR
987 pathway. Science 322:963-966.

988 Pedregosa F, Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
989 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,
990 Brucher, M., Perrot, M., Duchesnay, E (2011) Scikit-learn: Machine Learning in Python.
991 Journal of Machine Learning Research 12:2825--2830.

992 Peng H et al. (2021) Morphological diversity of single neurons in molecularly defined cell types.
993 Nature 598:174-181.

994 Poplawski GHD, Kawaguchi R, Van Niekerk E, Lu P, Mehta N, Canete P, Lie R, Dragatsis I,
995 Meves JM, Zheng B, Coppola G, Tuszynski MH (2020) Injured adult neurons regress to
996 an embryonic transcriptional growth state. Nature 581:77-82.

997 Sahni V, Itoh Y, Shnider SJ, Macklis JD (2021a) Crim1 and Kelch-like 14 exert complementary
998 dual-directional developmental control over segmentally specific corticospinal axon
999 projection targeting. Cell Rep 37:109842.

1000 Sahni V, Shnider SJ, Jabaudon D, Song JHT, Itoh Y, Greig LC, Macklis JD (2021b)
1001 Corticospinal neuron subpopulation-specific developmental genes prospectively indicate
1002 mature segmentally specific axon projection targeting. Cell Rep 37:109843.

1003 Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, Bien E, Baum M,
1004 Bortolin L, Wang S, Goeva A, Nemesh J, Kamitaki N, Brumbaugh S, Kulp D, McCarroll
1005 SA (2018) Molecular Diversity and Specializations among the Cells of the Adult Mouse
1006 Brain. Cell 174:1015-1030 e1016.

1007 Schwab ME, Strittmatter SM (2014) Nogo limits neural plasticity and recovery from injury. Curr
1008 Opin Neurobiol 27:53-60.

1009 Sekine Y, Lin-Moore A, Chenette DM, Wang XX, Jiang ZX, Cafferty WB, Hammarlund M,
1010 Strittmatter SM (2018) Functional Genome-wide Screen Identifies Pathways Restricting
1011 Central Nervous System Axonal Regeneration (vol 23, pg 415, 2018). Cell Reports
1012 24:269-269.

1013 Tasic B et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics.
1014 Nat Neurosci 19:335-346.

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

1015 Tasic B et al. (2018) Shared and distinct transcriptomic cell types across neocortical areas.
1016 Nature 563:72-78.

1017 Ueno M, Nakamura Y, Li J, Gu Z, Niehaus J, Maezawa M, Crone SA, Goulding M, Baccei ML,
1018 Yoshida Y (2018) Corticospinal Circuits from the Sensory and Motor Cortices
1019 Differentially Regulate Skilled Movements through Distinct Spinal Interneurons. Cell
1020 Rep 23:1286-1300 e1287.

1021 van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL,
1022 Pattabiraman D, Bierie B, Mazutis L, Wolf G, Krishnaswamy S, Pe'er D (2018)
1023 Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell 174:716-
1024 729 e727.

1025 Wirth EK, Roth S, Blechschmidt C, Holter SM, Becker L, Racz I, Zimmer A, Klopstock T,
1026 Gailus-Durner V, Fuchs H, Wurst W, Naumann T, Brauer A, de Angelis MH, Kohrle J,
1027 Gruters A, Schweizer U (2009) Neuronal 3',3,5-triiodothyronine (T3) uptake and
1028 behavioral phenotype of mice deficient in Mct8, the neuronal T3 transporter mutated in
1029 Allan-Herndon-Dudley syndrome. J Neurosci 29:9439-9449.

1030 Wolock SL, Lopez R, Klein AM (2019) Scrublet: Computational Identification of Cell Doublets
1031 in Single-Cell Transcriptomic Data. Cell Syst 8:281-291 e289.

1032 Yao Z et al. (2020) An integrated transcriptomic and epigenomic atlas of mouse primary motor
1033 cortex cell types. bioRxiv:2020.2002.2029.970558.

1034 Yao Z et al. (2021) A transcriptomic and epigenomic cell atlas of the mouse primary motor
1035 cortex. Nature 598:103-110.

1036 Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, Dong H, Zhuang X (2021) Spatially
1037 resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature 598:137-
1038 143.

1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052

JNeurosci Accepted Manuscript

1053 **Figure Legends**

1054 **Figure 1: Single cell RNAseq pipeline for adult CSNs.**

1055 Schematic of spinal tracing via injecting retro-AAV-CAG-GFP at spinal level C6/7 and retro-
1056 AAV-CAG-tdTomato at spinal level L4/5 (**A**). Photomicrographs (**B-D**) from M1 (~ 0.0 mm A-
1057 P Bregma, medial cortex) showing FL-CSNs (**B**), HL-CSNs (**C**), and an overlay with both single
1058 and dual-projecting CSNs (**D, D'**) 14 days post spinal infusion. Schematic overview of the
1059 dissection-sequencing procedure. Mice received injections of retro-AAV-CAG-GFP into either
1060 C6/7 or L4/5 (**E**). After a 2-week incubation, three 500 µm sections through M1 were collected
1061 (**E**i), layer 2/5 and 5were macrodissected (**E**ii), dissociated, and traced CSNs (green), L2/3 cells
1062 (yellow) and L5 nonfluorescent cells (blue) were then sorted using FACS (**E**iii) and sequenced
1063 using 10x Chromium (**E**iv). To exclude putative CSNs from the unlabeled layer 5 population of
1064 sequenced neurons, we used Kernel density estimation (KDE) analysis to quantify the
1065 transcriptional overlap between populations in low dimensional space. Non-traced layer 5
1066 neurons (blue) with >50% probability (blue, white outline) of originating from the CSN
1067 distribution (green) were excluded (**F**). To exclude putative layer 5 neurons from the layer 2/3
1068 population of sequenced neurons, we used KDE to quantify the transcriptional overlap between
1069 these populations in low dimensional space. Layer 2/3 neurons (orange) with >50% probability
1070 (orange, white outline) of originating from the layer 5 distribution (blue) were excluded (**G**).
1071 Scale bars = 500 µm (**B-D**) and 100 µm (**D'**).

1072

1073 **Figure 2: Gene expression confirms sequenced cell identity.**

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

1074 Sequenced CSNs, L2/3 neurons and non-traced L5 cells have overlapping distributions of library
1075 size, mean library size of 19,432 UMIs, median = 18,474 UMIs (**A**). Differential gene expression
1076 analysis revealed 32 genes were enriched in neurons from male mice and 213 genes enriched in
1077 neurons from female mice (**B,** Wilcoxon rank sum test with a Benjamini-Hochberg correction, p-
1078 adj < 0.05). IPA was used to gain biological insight between the sexes, only three pathways
1079 were significantly enriched, the synaptogenesis, protein ubiquitination and eNOS signaling (**C**,
1080 right-tailed Fisher Exact test with a Benjamini-Hochberg correction, p < 0.05). Heatmaps
1081 representing normalized imputed expression of sex-specific genes (**D**), intact cell (IC) genes (**E**),
1082 cell-type genes (**F**), neurotransmitter (NT) genes (**G**), and cortical layer genes (**H**), confirming
1083 that sequenced cells were sex balanced, intact, neurons, glutamatergic and from expected layers,
1084 respectively.

1085

1086 **Figure 3: CSNs are unique amongst neurons in M1 and are enriched in *Wnt7b* expression.**

1087 Differential gene expression analysis between traced CSNs and a combination of layer 2/3 and
1088 L5 nonfluorescent cells shows that197 genes were upregulated and 639 downregulated in CSNs.
1089 217 genes were upregulated and 634 downregulated in CSNs compared to layer 5 non-traced
1090 neurons, and 467 genes upregulated and 1533 downregulated in CSNs compared to layer 2/3
1091 neurons (**A,** Wilcoxon rank sum test with a Benjamini-Hochberg correction, p-adj < 0.05).
1092 QIAGEN IPA was used to identify enriched pathways in each of these analyses, with pathway
1093 names abbreviated for clarity (**B**, right-tailed Fisher Exact test with a Benjamini-Hochberg
1094 correction, p-adj < 0.05). PHATE projection (**C, Movie 1**) and kernel density estimate plots with
1095 hierarchical clustering and a similarity index (**D**) colored by cell identity illustrate differences
1096 between CSNs, L5 non-traced, and L2/3 neurons. *Wnt7b* is significantly enriched in CSNs

JNeurosci Accepted Manuscript

1097 compared to layer 5 non-traced and L2/3 neurons (**E**, Wilcoxon rank sum test with a Benjamini-
1098 Hochberg correction, ****p-adj < 0.001). Schematic showing labeling of cervical CSNs with
1099 retro-AAV-CAG-GFP (rAAV-GFP, cyan) and lumbar CSNs with retro-AAV-CAG-tdTomato
1100 (rAAV-tdT, magenta) (**F**). smFISH of retrogradely labeled CSNs shows an enrichment of *Wnt7b*
1101 in FL- and HL-CSNs compared to nonfluorescent, DAPI+ cells in layer 5b (**G-H**, data shown are
1102 average number of puncta per cell ± SEM, one-way ANOVA with a Bonferroni correction,
1103 ****p < 0.0001). Scale bars = 10 µm. Data tables from the scRNAseq containing the results of
1104 the DE analysis comparing CSNs to non-traced L5 and L2/3 neurons are provided in **Fig. 3-1**.

1105

1106 **Movie 1: 3D PHATE visualization of CSNs, L2/3, and non-traced L5 neurons in M1**

1107 3D PHATE projection of 5495 intact adult neurons from M1 colored by cell identity (green =
1108 traced CSNs, blue non-traced L5, yellow = L2/3 neurons).

1109

1110 **Figure 3-1: DE genes for CSNs versus non-CSNs in M1**

1111 Data tables from the scRNAseq containing the results of the DE analysis comparing CSNs to
1112 non-traced L5 and L2/3 neurons (Wilcoxon rank sum test with a Benjamini-Hochberg correction,
1113 p-adj < 0.05).

1114

1115 **Figure 4: Utilizing linear SVM classifiers to identify CSNs among cells in M1.**

1116 A dotplot of previously characterized L5-enriched genes illustrates enrichment and expression of
1117 each gene in CSNs, non-traced L5, and L2/3 neurons. The size of the dot denotes the proportion

JNeurosci Accepted Manuscript

1118 of each population that expresses a given gene, and the color denotes the normalized mean
1119 expression within that subtype (**A**). PHATE projections colored by normalized gene expression
1120 of previously characterized L5 genes including *Ctip2*, and *Rbp4,* along with newly-identified
1121 *wnt7b,* illustrate expression in CSNs, non-traced L5 neurons, and L2/3 neurons. Boxed PHATE
1122 projection illustrates neuronal cell types in their transcriptional space (CSNs = green, non-traced
1123 L5 neurons = blue, L2/3 neurons = orange, **B**). A receiver operating characteristic (ROC) curve
1124 showing the performance of a linear support vector machine (SVM) trained to classify CSNs
1125 based on previously characterized L5 genes (L5-C); the top 100, top 10, or random 10 novel
1126 genes enriched in CSNs (CSN-C); or 10 random genes. AUC is calculated for each classifier,
1127 showing that CSN-C outperforms L5-C, and top 10 and random 10 CSN genes outperforms the
1128 L5-C classifier (randomized permutation test, $p < 0.001$, **C**). PHATE projections showing CSN
1129 classification performance with either the CSN-C (top, forest green) or L5-C (bottom, purple),
1130 CSNs missed by L5-C (black), and a quantification of the proportion of CSNs that were
1131 classified correctly. CSN-C correctly classified 92% of CSNs. L5-C correctly classified 76% of
1132 CSNs (**D**). *Rbp4* cre:Ai14 mice received infusions of retro-AAV-CAG-GFP into either the
1133 cervical, C6/7, or the lumbar, L4/5, spinal cord (**E**). Low (F) and high (F') power
1134 photomicrographs show GFP (cyan) expression in a section through M1 from an *Rbp4* cre:Ai14
1135 (magenta) mouse. Cyan cells represent traced CSNs. Magenta cells are *Rbp4*+ neurons that do
1136 not project to the spinal cord, white cells are GFP+/*Rbp4*+ neurons (**F'**). Quantification of the
1137 proportion of *Rbp4*+ neurons that express GFP in coronal sections through M1 (1.7→ -1.2,
1138 relative to Bregma) (**G,** Mean = 7.7%). CSN-C prediction of CSN identity of sequenced *Rbp4*+
1139 neurons. CSN-C predicts 7% of *Rbp4* neurons are CSNs with a 75% probability cutoff (**H**). Scale
1140 bars = 500 µm and 100 µm.

1141

1142 **Figure 5: Forelimb and hindlimb CSN specialization.**

1143 PHATE projection colored by cell identity of retrogradely labeled GFP+ cervical CSNs (cyan)
1144 and tdTomato+ lumbar CSNs (magenta, **A**). Histogram of MELD prediction confidence of
1145 lumbar identity in all CSNs, showing predicted intermediate cells in grey (**B**). PHATE projection
1146 showing MELD prediction confidence of lumbar identity (**C**). Comparison of counts for tracing-
1147 identified lumbar and cervical CSN populations (top) with MELD-predicted counts for lumbar,
1148 cervical, and intermediate CSN populations (bottom) (**D**). DE analysis identified 492 upregulated
1149 FL genes and 788 upregulated HL genes (**E,** Wilcoxon rank sum test, Benjamini-Hochberg
1150 correction, p-adj < 0.05). Violin plots of the top 10 most differentially expressed genes in FL
1151 CSNs (cyan) and HL CSNs (**F**, magenta, Wilcoxon rank sum test, Benjamini-Hochberg
1152 correction, p-adj < 0.05). PHATE projection colored by *Cacng7* expression illustrates FL
1153 enrichment (**G**). PHATE projection colored by *Slc16a2* expression illustrates HL enrichment
1154 (**H**). smFISH of retrogradely labeled GFP+ FL-CSNs and tdTomato+ HL-CSNs shows an
1155 enrichment of *Cacng7* in FL-CSNs (**I**, **J**, data shown are average number of puncta per traced
1156 CSN ± SEM, unpaired t-test, ***p < 0.0001). smFISH of retrogradely labeled GFP+ FL-CSNs
1157 and tdTomato+ HL-CSNs shows an enrichment of *Slc16a2* in HL-CSNs (**K, L**, data shown are
1158 average number of puncta per traced CSN ± SEM, unpaired t-test, ***p < 0.001). A dotplot of
1159 previously characterized L5-enriched genes illustrates enrichment and expression of each gene in
1160 FL and HL CSNs. The size of the dot denotes the proportion of each population that expresses a
1161 given gene, and the color denotes the normalized mean expression within that subtype. These L5
1162 genes are not uniquely nor differentially enriched in FL or HL CSNs (**M**). A ROC curve showing
1163 the performance of a linear SVM trained to classify limb-specific CSNs based on previously

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

1164 characterized L5 genes (L5-C), or our novel limb-specific gene list (LS-CSN-C). AUC is
1165 calculated for each classifier, showing that LS-CSN-C outperforms C-L5 (**N**). Scale bars = 30
1166 µm and 15 µm. Data tables from the scRNAseq containing the results of the DE analysis
1167 comparing FL versus HL CSNs are provided in **Fig. 5-1**.

1168

1169 **Figure 5-1: DE genes for FL versus HL CSNs**

1170 Data tables from the scRNAseq containing the results of the DE analysis comparing FL versus
1171 HL CSNs (Wilcoxon rank sum test with a Benjamini-Hochberg correction, p-adj < 0.05).

1172

1173 **Figure 6: Transcriptionally distinct CSN subtypes can be defined by supraspinal terminal**
1174 **fields.**

1175 Schematic based on the BICCN M1 taxonomy showing that projection neurons can be classified
1176 as into 4 categories based on their primary terminal field, including ET (areas shaded in orange
1177 in the schematic), IT (areas shaded green), CT (areas shaded purple), and NP (areas shaded red,
1178 **A**). PHATE projection showing 10x V3-sequenced M1 neurons from Yao, *et al.,* 2021, colored
1179 by ET (orange), IT (green), CT (purple) and NP (red) projection subtype (**B**). PHATE projection
1180 of CSNs collected in this study projected onto the same PHATE space as in (**B**) from Yao, *et al,*
1181 colored by projection subtype show that CSNs are represented in all 4 projection classes (**C**).
1182 PHATE projections of CSNs and M1 projection neurons from Yao, *et al*., (**D, Movie 2**). Violin
1183 plot showing gene expression in each projection subtype (**E**). PHATE projections (**F**) showing
1184 enrichment of established ET projection gene *Fam84b* and novel ET enriched gene *Gprc5*,
1185 enrichment of established IT gene *Slc30a3*, and novel IT enriched gene *Frzb*, enrichment of

JNeurosci Accepted Manuscript

1186 established CT gene *Foxp2*, and novel CT enriched gene *Syt6*, enrichment in established NP
1187 gene *Sla2*, and novel NP enriched gene *Lypd1*. DE analysis identified 791 upregulated and 181
1188 down regulated genes in CSN-ET cells vs CSN-IT, -CT, and -NP cells, listed with the top 6
1189 enriched pathways from QIAGEN IPA (**G,** Wilcoxon rank sum test, Benjamini-Hochberg
1190 correction, p-adj < 0.05). DE analysis identified 275 upregulated and 263 down regulated genes
1191 in CSN-IT cells vs CSN-ET, -CT, and -NP cells, listed with the top 6 enriched pathways from
1192 QIAGEN IPA (**H,** Wilcoxon rank sum test, Benjamini-Hochberg correction, p-adj < 0.05). DE
1193 analysis identified 215 upregulated and 441 down regulated genes in CSN-CT cells vs CSN-ET,
1194 -IT, and -NP cells, listed with the top 6 enriched pathways from QIAGEN IPA (**I,** Wilcoxon rank
1195 sum test, Benjamini-Hochberg correction, p-adj < 0.05). DE analysis identified 675 upregulated
1196 and 1060 down regulated genes in CSN-NT cells vs CSN-ET, -IT, and -CT cells, listed with the
1197 top 6 enriched pathways from QIAGEN IPA (**J,** Wilcoxon rank sum test, Benjamini-Hochberg
1198 correction, p-adj < 0.05). To validate CSN innervation of ET, IT, CT, and NP structures, retro-
1199 AAV-CAG-tdTomato was injected in the cervical and lumbar cord. Terminals in supraspinal
1200 brain regions were examined including the brainstem (ET, orange, **K**), the striatum (IT, green,
1201 **L**), motor thalamus (CT, purple, **M**), and primary motor cortex (NP, red, **N**) and (**F**).
1202 Photomicrographs of CSN terminals in the brainstem (ET, **K'**), striatum (IT, **L'**), motor thalamus
1203 (CT, **M'**), and primary motor cortex (NP, **N'**) stained with antibodies against synaptophysin
1204 (green) and NeuN (white) show presumptive synaptic connections within these supraspinal
1205 terminal fields. IT: Intratelencephalic, ET: Extratelencephalic, NP: Near-projecting, CT:
1206 Corticothalamic. Scale bars = 30 μm. Data tables from the scRNAseq containing the results of
1207 the DE analysis comparing each CSN supraspinal subtype versus all the other subtypes are
1208 provided in **Fig. 6-1**.

JNeurosci Accepted Manuscript

1209

1210 **Movie 2: 3D PHATE visualization of CSNs and deep layer glutamatergic projection**
1211 **neurons sequenced from Yao et al., 2021 in the Yao et al., 2021 manifold.**

1212 PHATE projection of CSNs and M1 projection neurons from Yao, et al., colored by BICCN
1213 taxonomy, ET (orange), IT (green), CT (purple) and NP (red) projection subtype.

1214

1215 **Figure 6-1: DE genes among CSN supraspinal subtypes**

1216 Data tables from the scRNAseq containing the results of the DE analysis for comparing each
1217 CSN supraspinal subtype versus all the other subtypes (Wilcoxon rank sum test with a
1218 Benjamini-Hochberg correction, $p\text{-adj} < 0.05$).

1219

1220 **Figure 7: Forelimb and hindlimb-specific supraspinal CSN innervation and specialization.**

1221 CSNs in native PHATE space colored by supraspinal categories (**A, Movie 3**), and FL and HL
1222 identity (**B**). Divergent numbers of FL and HL CSNs transcriptionally defined by BICCN
1223 supraspinal structures (**C, D**). Photomicrographs (**E**) show representative images of FL (cyan
1224 outline) and HL (magenta outline) td-tomato +ve CST projections after intersectional tracing in
1225 the brainstem (ET population). Projection density of FL-CSN-ET was significantly higher than
1226 in HL-CSN-ET (n=4 FL traced mice, n=3 HL traced mice, mixed-design ANOVA, ** $p=0.003$,
1227 **F**). Schematic illustrates FL axons (thick cyan lines) more densely innervating ET structures
1228 compared to HL axons (thin magenta lines, **G**). Number of DE genes comparing FL-CSN-ET to
1229 FL-CSN-IT, -CT, and NP combined (orange to black), and FL-CSN-ET vs IT (orange

JNeurosci Accepted Manuscript

1230 [upregulated] to green downregulated]), FL-CSN-ET vs CT (orange to purple), and FL-CSN-ET
1231 vs -NP (orange to red, **H**). These comparisons are repeated for the HL-CSN-ET population (**H'**).

1232 Photomicrographs (**I**) show FL and HL td-tomato +ve CST projections after intersectional
1233 tracing in the striatum (IT population). There was no significant difference in the projection
1234 density between ipsilateral and contralateral FL- and HL-CSN-IT projections (mixed-design
1235 ANOVA, p=0.38, **J**). Schematic illustrates FL axons and HL axons innervating IT structures
1236 (green, **K**). Number of DE genes comparing FL-CSN-IT to FL-CSN-ET, -CT, and -NP
1237 combined (green to black), and FL-CSN-IT vs ET (orange to green), FL-CSN-IT vs CT (purple
1238 to green), and FL-CSN-IT vs NP (green to red, **L**). These comparisons are repeated for the HL-
1239 CSN-IT population (**L'**).

1240 Photomicrographs (**M**) show images of FL and HL td-tomato +ve CST projections after
1241 intersectional tracing in the thalamus (CT population, VPL = ventral posterolateral nucleus of
1242 thalamus). There was no significant difference in the projection density between FL- and HL-
1243 CSN-CT projections (mixed-design ANOVA, p=0.085, **N**). Schematic illustrates FL axons and
1244 HL axons innervating the thalamus (**O**). Number of DE genes comparing FL-CSN-CT to FL-
1245 CSN-ET, -IT, and NP combined (purple to black), and FL-CSN-CT vs ET (purple to orange),
1246 FL-CSN-CT vs -IT (purple to green), and FL-CSN-CT vs -NP (purple to red, **P**). These
1247 comparisons are repeated for the HL-CSN-ET population (**P'**).

1248 Photomicrographs (**Q**) show images of FL and HL td-tomato +ve CST projections after
1249 intersectional tracing in M2 (NP population). Projection density of FL-CSN-NP was significantly
1250 higher than in HL-CSN-ET (n=4 FL traced mice, n=3 HL traced mice, mixed-design ANOVA,
1251 *** p<0.001, **R**). Schematic illustrates FL axons more densely innervating NP structures
1252 compared to HL axons (**S**). Number of DE genes comparing FL-CSN-NP to FL-CSN-ET, -IT,

JNeurosci Accepted Manuscript

1253 and -CT combined (red to black), and FL-CSN-NP vs TT (red to orange), FL-CSN-IT vs NP
1254 (green to red), and FL-CSN-NP vs CT (purple to red, **T**). These comparisons are repeated for the
1255 HL-CSN-ET population (**T'**). Scale bars: E = 500 µm, I, M, Q = 100 µm. Data tables from the
1256 scRNAseq containing the results of the DE analysis comparing between FL CSN supraspinal
1257 subtypes, and HL within each CSN supraspinal subtypes are provided in **Fig. 7-1**.
1258

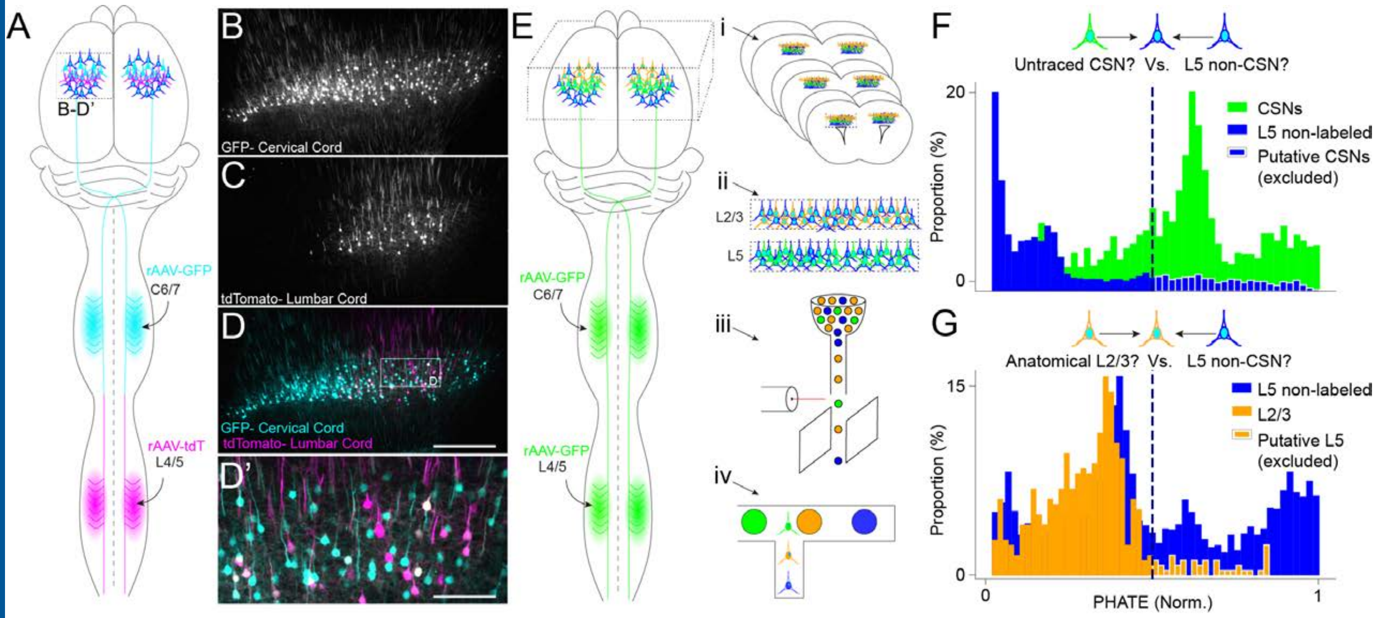
1259 **Movie 3: 3D PHATE visualization of CSNs traced from the cervical and lumbar**
1260 **enlargements colored by BICCN taxonomy in their native manifold.**

1261 3D PHATE visualization of CSNs in their native manifold colored by BICCN taxonomy, CSN-
1262 ET (orange), CSN-IT (green), CSN-CT (purple) and CSN-NP (red) projection subtype.
1263

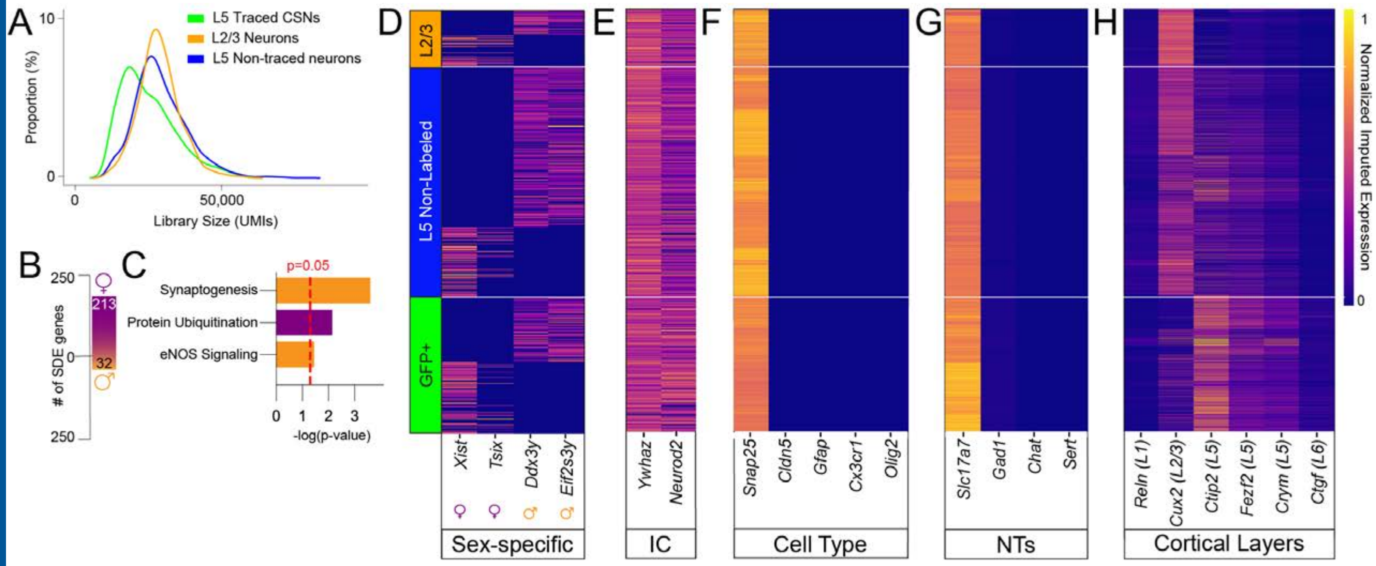
1264 **Figure 7-1: DE genes for FL and HL supraspinal CSN subtypes**

1265 Data tables from the scRNAseq containing the results of the DE analysis comparing between FL
1266 CSN supraspinal subtypes, and HL within each CSN supraspinal subtypes (Wilcoxon rank sum
1267 test with a Benjamini-Hochberg correction, p-adj < 0.05).
1268

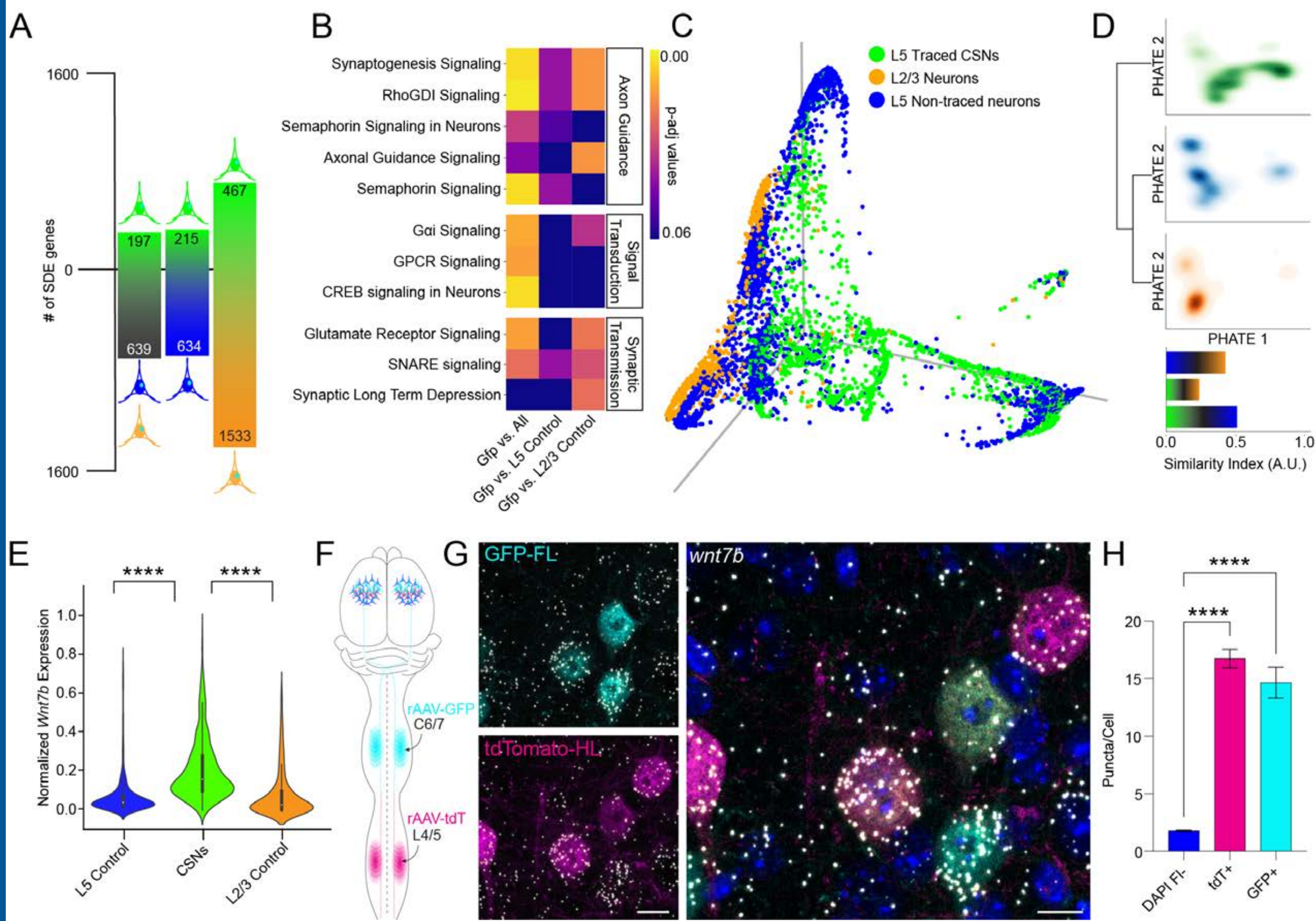
JNeurosci Accepted Manuscript



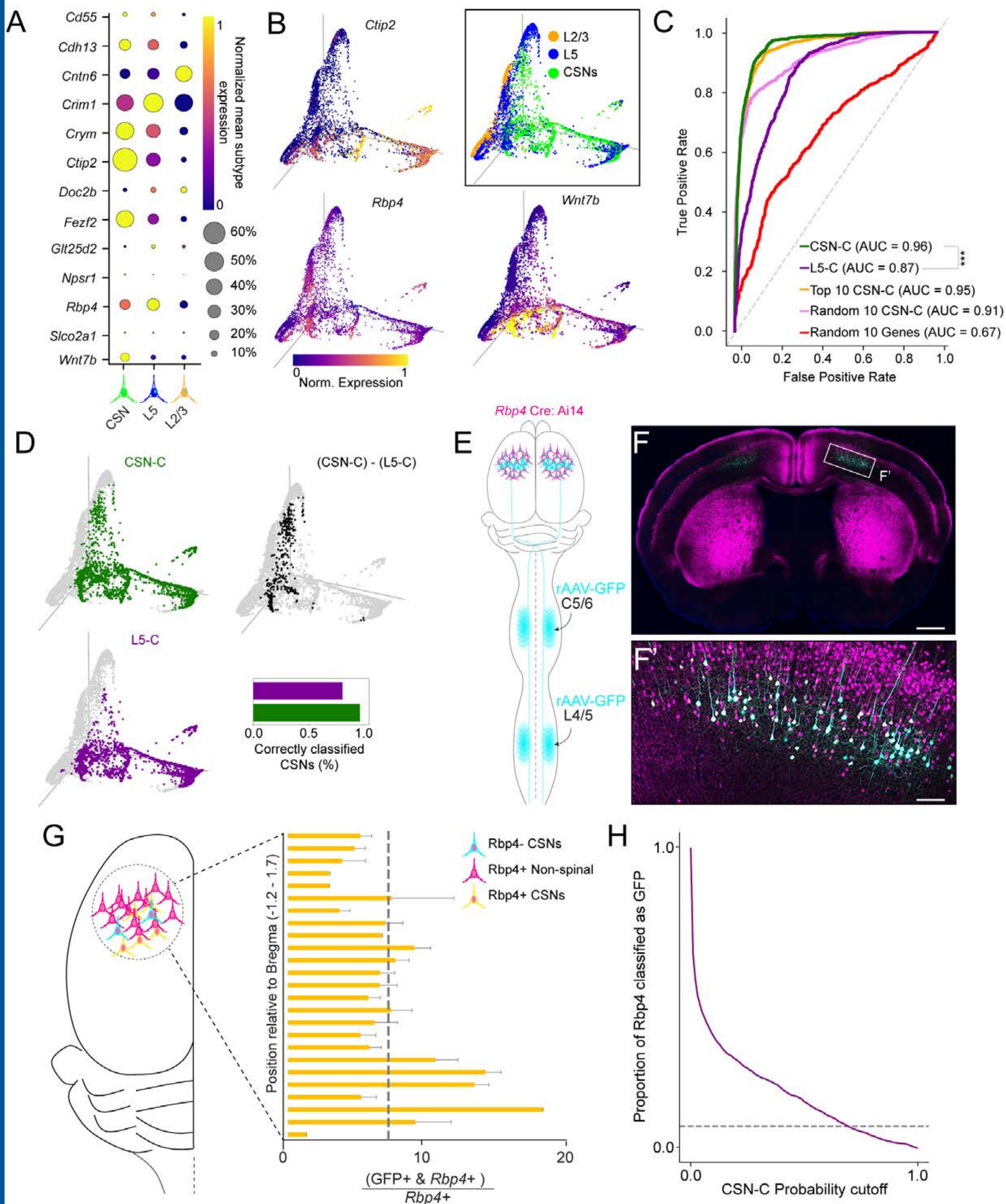
JNeurosci Accepted Manuscript



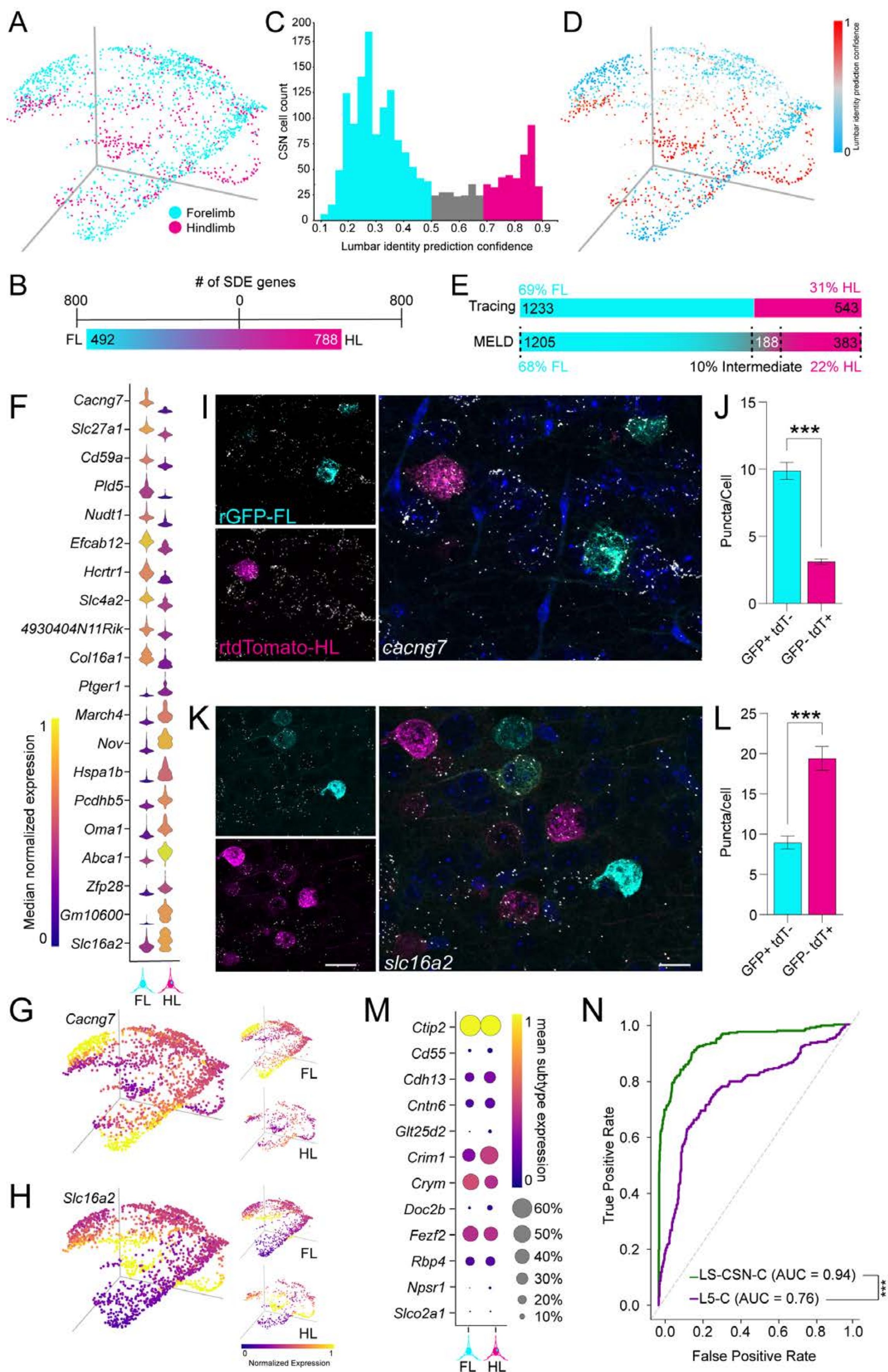
JNeurosci Accepted Manuscript



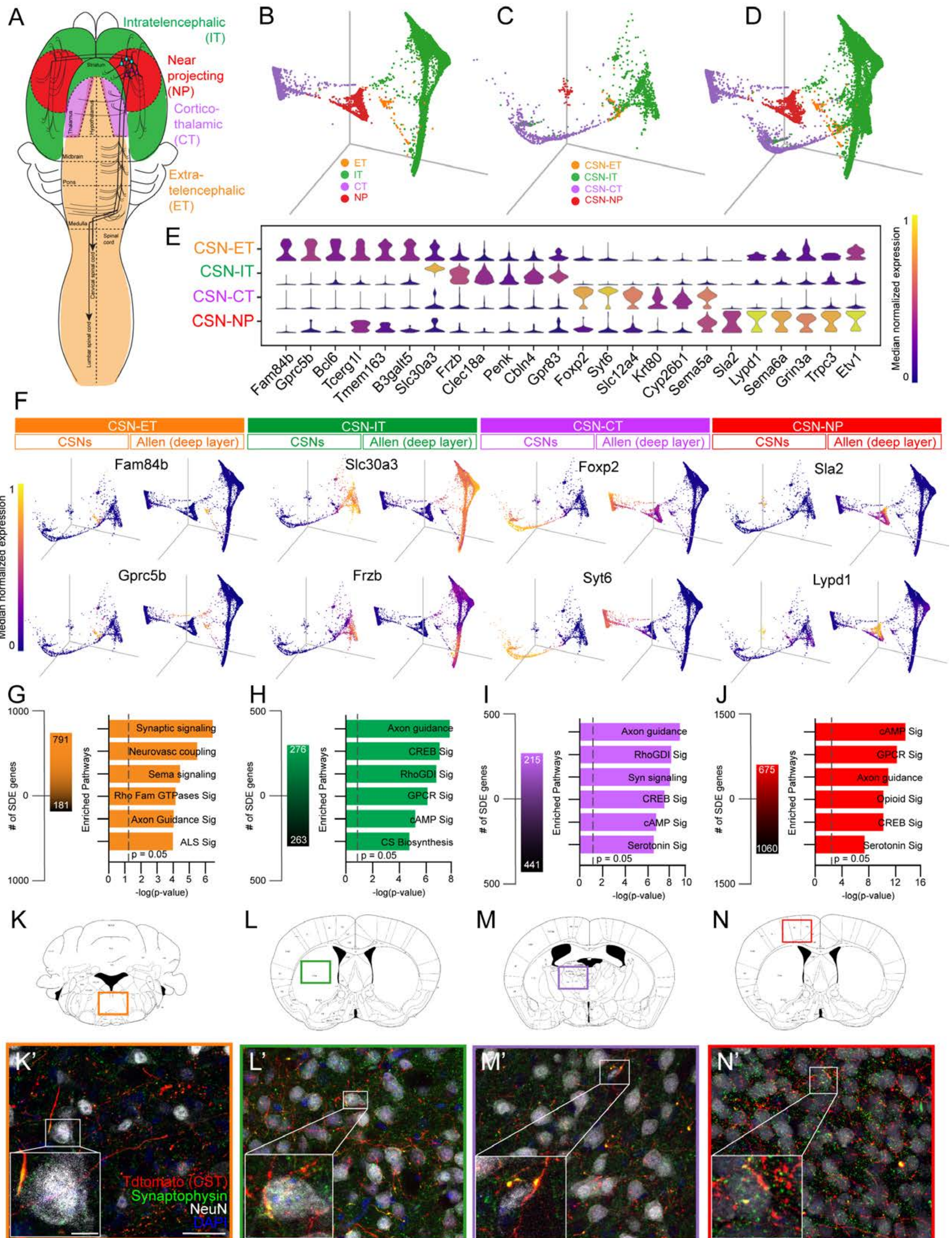
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript