Projection-TAGs enable multiplex projection tracing and multi-modal profiling of projection neurons

Lite Yang[1,2], Fang Liu[1], Hannah Hahm[1], Takao Okuda[1], Xiaoyue Li[1], Yufen Zhang[1], Vani Kalyanaraman[1], Monique R. Heitmeier[1], Vijay K. Samineni[1*]

1 Department of Anesthesiology, Washington University Pain Center, Washington University School of Medicine, St. Louis, MO, United States

2 Neuroscience Graduate Program, Division of Biology & Biomedical Sciences, Washington University School of Medicine, St. Louis, MO, United States

*Contact information: vksamineni@wustl.edu

Abstract

Single-cell multiomic techniques have sparked immense interest in developing a comprehensive multi-modal map of diverse neuronal cell types and their brain-wide projections. However, investigating the complex wiring diagram, spatial organization, transcriptional, and epigenetic landscapes of brain-wide projection neurons is hampered by the lack of efficient and easily adoptable tools. Here we introduce Projection-TAGs, a retrograde AAV platform that allows multiplex tagging of projection neurons using RNA barcodes. By using Projection-TAGs, we performed multiplex projection tracing of the cortex and high-throughput single-cell profiling of the transcriptional and epigenetic landscapes of the cortical projection neurons in female mice. Projection-TAGs can be leveraged to obtain a snapshot of activity-dependent recruitment of distinct projection neurons and their molecular features in the context of a specific stimulus. Given its flexibility, usability, and compatibility, we envision that Projection-TAGs can be readily applied to build a comprehensive multi-modal map of brain neuronal cell types and their projections.

Main

Understanding brain function requires the elucidation of the complex wiring diagram and constituent cell types across brain regions. Revolutionary work from Golgi and Cajal laid the foundation for understanding the diversity of neurons and their anatomical connections based on morphological features[2,3]. Over the past few decades, advancements have led to the further classification of neuronal cell types incorporating additional modalities, including molecular, electrophysiological, morphological, and anatomical features[4-20]. Recent breakthroughs in high-throughput single-cell techniques and spatially-resolved molecular assays have sparked immense interest in developing a comprehensive multi-modal map of diverse neuronal cell types and their brain-wide projections[21-28]. Despite the rapid integration of novel multiomic technologies in studying brain-wide connectivity, the investigation of spatially mingled neuronal projections continues to be hampered by the lack of broadly available tools to simultaneously trace multiple neuronal projections and/or profile projection neurons for multi-modal investigations.

Traditional neuroanatomical tracing methods, performed often with tracers or with viral vectors use fluorescence as the projection identifier, have been invaluable in mapping distinct neuronal projections[29-37]. However, these methods are limited by the spectral properties of fluorophores that can be detected in a single experiment, thus limiting the number of neuronal projections that can be examined simultaneously. While the recent advance in fluorescence microscopy enables simultaneous imaging of up to 10 fluorophores[38], common microscopy equipment in neuroscience labs can reliably detect only 3-4 fluorophores, limiting the number of projections to be examined simultaneously[27,39,40]. Furthermore, such approaches are not directly suited for high-throughput sequencing-based assays such as single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq), as the detection of exogenous transcripts is usually low by short-read sequencing. Alternatively, one can still build a multi-modal atlas for projection neurons by tracing one or two projections at a time using fluorophores and incorporating this tracing paradigm with fluorescently-activated cell sorting (FACS) for single-cell profiling[41,42]. Methods, such as retro-seq and epi-retro-seq, characterized the molecular features of projection neurons[22,26,28,43-45], but they are often inefficient, costly, and not suitable for studying the complex wiring diagrams of projections to multiple targets.

RNA barcode-based tracing tools provide a promising avenue for high-throughput neuroanatomical studies, as diverse RNA barcodes can be parallelly detected via sequencing or imaging, thus offering a very powerful and scalable strategy to trace neuronal projections. Employing an anterograde tracing scheme, MAPSeq and BARseq utilize a sindbis viral library to encode a diverse collection of short random RNA barcodes, which are strategically designed to label individual cells and engineered to be anterogradely transported to axonal terminals for projection tracing[46-50]. Despite the ability to quantitatively measure the projection strength to any target regions of investigation, the requirement of highly customized instruments and pipelines by those methods has restricted their adoption beyond several expert labs. The cellular toxicity associated with the sindbis virus also poses a challenge to integrating their use with chronic experimental paradigms[47,51]. More recently, employing a retrograde tracing scheme using adeno-associated viruses (AAVs), both Projection-seq and MERGE-seq used RNA barcodes as the projection identifier for multiplex projection tracing[52,53]. The low cellular toxicity of AAVs allowed them to correlate gene expression to projections using scRNA-seq, but their compatibility for additional modalities, such as epigenetic profiling, spatial analysis, and activity-dependent circuit mapping is lacking or not sufficiently demonstrated.

Here we introduce Projection-TAGs, a retrograde AAV platform that allows multiplex tracing of projection neurons using RNA barcodes, which we call Projection-TAGs. Similar to Projection-seq and MERGE-seq, the key component of Projection-TAGs is a set of engineered retrograde AAVs each expressing a unique RNA barcode TAG, which acts as the projection identifier (Extended Data Fig. 1a). With this scheme, neurons projecting to a target region are uniquely labeled by a retrograde AAV-mediated TAG and multiplex projection tracing can be achieved by injecting a unique Projection-TAG AAV into each of the target regions predefined for investigation. We developed a toolbox for Projection-TAG detection using various commercial assays, enabling multiplex spatial neuroanatomical studies, high-throughput multi-modal profiling of projection neurons, and investigation of activity-dependent populations in response to a stimulus of interest. In this study, we utilized Projection-TAGs to study the complex wiring diagram, spatial organization, transcriptional, and epigenetic landscapes of intracortical-, subcortical-, and corticospinal-projecting neurons in the cortex and identify activity-dependent recruitment of distinct cortical projections in a mouse model of visceral pain. Projection-TAGs are designed for incorporation into existing experimental pipelines with minimal effort and easy application in studying the nervous system.

Results

Design of Projection-TAGs

To democratize and facilitate the use of Projection-TAGs in neuroscience labs without any specialized equipment, we incorporated the following features into the Projection-TAG plasmid design: First, we employed a chicken beta-actin (CAG) promoter to enable ubiquitous Projection-TAG expression in various tissue and cell types (Fig. 1a). Second, we reasoned that the inclusion of a fluorescent marker could enable flexibility in downstream processes and confirming viral expression. We screened several fusion fluorescent proteins for the ability to label both intact cells and nuclei in suspension (Extended Data Fig. 1b, c), and included in our AAV plasmids a fusion of Sad1 And UNC84 Domain Containing 1 (Sun1) to Green Fluorescent Protein (GFP), allowing enrichment of target cell/nucleus populations using FACS (Extended Data Fig. 1e). Sun1-GFP labels the nuclear membrane without interfering with the chromosome structure and has been employed in numerous epigenetic studies (Extended Data Fig. 1d)[44,54]. To further enhance Projection-TAG functionality, we developed a protocol for photobleaching Sun1-GFP fluorescence (Extended Data Fig. 2f,h) and created a version of Projection-TAGs expressing a red fluorophore oScarlet (Fig. 1a, Methods). Third, to enable projection tracing using various commercial assays, we cataloged 50 unique 100-nt RNA barcodes[55] (Extended Data Fig. 2a, Supplementary Table 1, Methods). By insertion into the 3' untranslated region (UTR) of the *Sun1-GFP* transcript, we generated 12 plasmids each expressing a unique Projection-TAG (TAGs 1-12, Fig. 1a). This design aimed to optimize its detection by 3' end RNA-seq assays and enable the development of fluorescent *in situ* hybridization (FISH) probes for spatial analysis.

As a proof of principle to validate Projection-TAG expression and detection, we first performed RNA-seq and multiplexed FISH on human embryonic kidney (HEK) cells transfected with individual Projection-TAG plasmids (Extended Data Fig. 2b,d). We faithfully detected Projection-TAGs with high specificity, as the detection of Projection-TAGs expressed in each HEK sample is at least 422.4 times (RNA-seq) and 6.9 times (multiplexed FISH) greater than that of Projection-TAGs not expressed or the background (Extended Data Fig. 2c,e). To use Projection-TAGs for multiplex tracing of neuronal projections in vivo, we packaged these validated plasmids into recombinant AAV serotype 2 retro (rAAV2-retro)[30] and generated a set of 12 Projection-TAG AAVs each expressing a unique Projection-TAG (Fig. 1a).

To evaluate the ability to use Projection-TAGs in high-throughput multi-modal profiling of projection neurons, we applied it to performing multiplex projection tracing in the adult mouse primary motor (MOp) and primary somatosensory cortex (SSp). Neurons in the MOp and SSp are anatomically organized by cortical layers and exhibit distinct layer-specific connectivity with other brain regions and have been extensively known to coordinate a wide range of innate and learned behaviors[43,56-58]. Tremendous progress has been made to characterize the diverse MOp and SSp cell types using single-cell multiomic approaches and create a comprehensive molecular taxonomy of projection neurons and their brain-wide projections[27,45,50,59-61], allowing us to validate the efficacy of Projection-TAGs by comparing the data generated in this study with the ground truth data. To test Projection-TAGs in vivo,

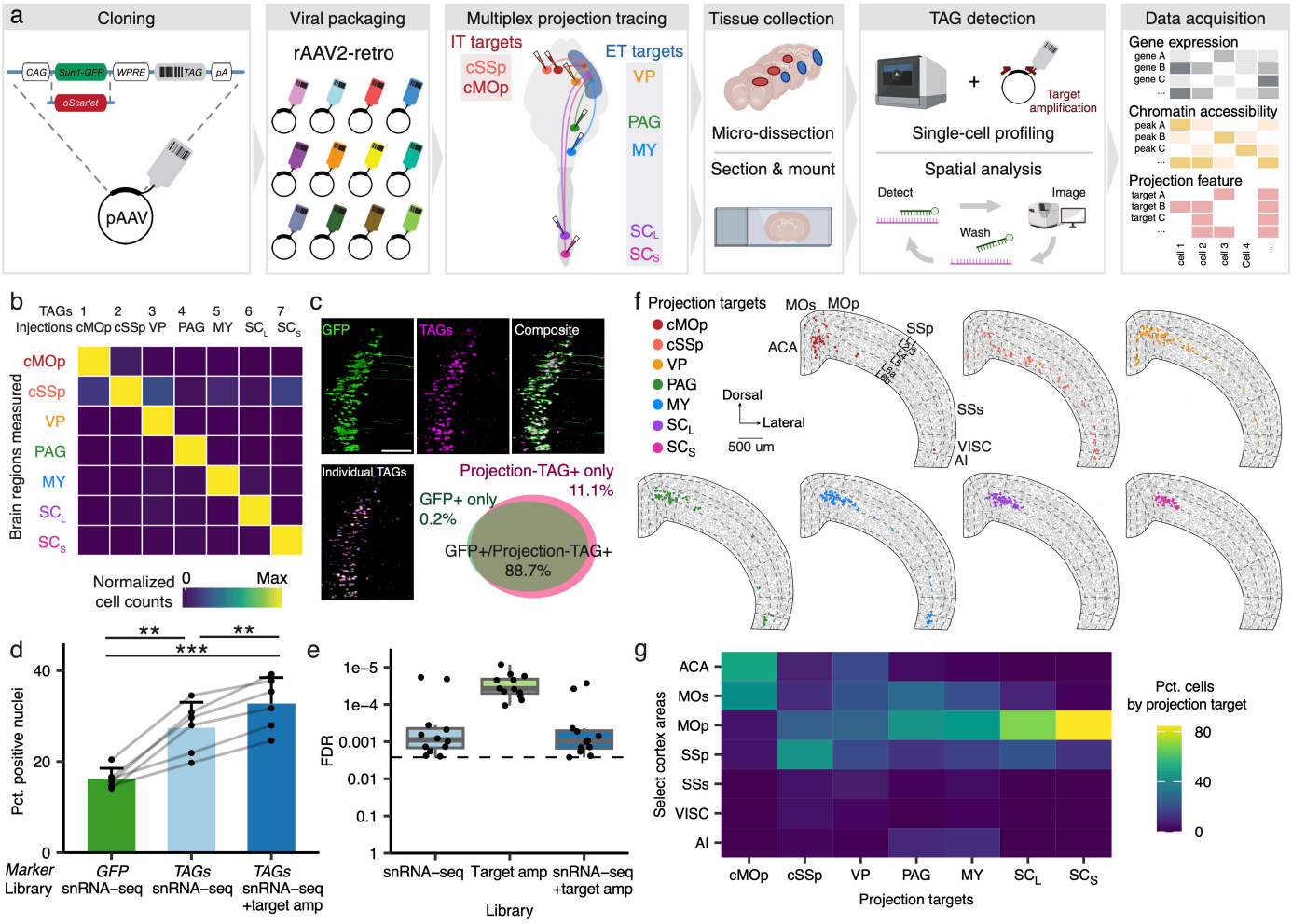


Figure 1. Multiplex projection tracing using Projection-TAGs
a, Experimental and analytical workflow of tagging neuronal projections of the MOp and SSp using Projection-TAGs. b, Heatmap showing the enrichment of cells labeled with individual Projection-TAGs by multiplexed FISH in brain regions of injections. Cell counts are normalized by the max values in each brain region. Individual TAGs and their injection sites are shown as on the column. c, Representative MOp images showing cells labeled with the fluorescent Sun1-GFP signal and fluorescent FISH signals from Projection-TAGs. Quantification shows the overlap between cells that are GFP+ and/or Projection-TAG+. FISH probes uniquely targeting Projection-TAGs (BCs 1-7) were visualized in separate imaging channels, shown as individual TAGs, and the signals were aggregated pos-hoc, shown as TAGs. Scale bar = 100 µm. d, Bar plot showing the percentage of nuclei expressing select marker genes (>0 UMIs for snRNA-seq, > cutoffs for target amplification, see Methods) in individual snRNA-seq libraries prepared by FACS and target amplification. Bar and error bar indicate the average and standard deviation across libraries. (p = 5.3e-06, $F(2,10)$ = 51.77, one-way ANOVA; **p = 0.006 between Projection-TAGs and *GFP* in snRNA-seq, **p = 0.003 for Projection-TAGs between snRNA-seq only and snRNA-seq with target amplification, ***p = 0.001 between *GFP* in snRNA-seq and Projection-TAGs in snRNA-seq with target amplification, post-hoc t-tests with Bonferroni correction). e, Box plot showing the false detection rate of Projection-TAGs due to ambient RNA contamination in snRNA-seq libraries. Dotted line shows the highest FDR due to ambient RNA contamination at 0.0026 among all libraries. f, Spatial map of Projection-TAG+ cells in select cortex areas from a brain section (-0.65 mm relative to bregma). Cells are downsampled to 200 per Projection-TAG and colored by the injection site of the corresponding TAG. Grey dots are cells segmented based on DAPI signal. g, Heatmap showing the distribution of projection neurons in select cortex areas from Fig. 1f. ACA: anterior cingulate area, MOs: secondary motor area, MOp: primary motor area, SSp: primary somatosensory area, SSs: supplemental somatosensory area, VISC: visceral area, AI: agranular insular area.

122 we identified seven downstream projection targets of MOp and SSp, including two intratelencephalic (IT)
123 targets (contralateral MOp [cMOp] and contralateral SSp [cSSp]) and five extratelencephalic (ET) targets,
124 which can be further classified into three subcortical targets (ipsilateral ventral posterior nucleus of the
125 thalamus [VP], ipsilateral periaqueductal grey [PAG], ipsilateral medulla [MY]) and two corticospinal
126 targets (lumbar spinal cord [SC_L], and sacral spinal cord [SC_S], Fig. 1a, Extended Data Fig. 5a).

127 **Multiplex projection tracing using Projection-TAG rAAV2-retro**

128 While previous studies using rAAV2-retro have shown that two weeks are typically sufficient for
129 fluorescently labeling cortical projections[21,22,43,44,62], a multiplex tracing experiment would benefit from
130 understanding the temporal kinetics and stability of cargo gene expression across different cortical
131 projections, which remains largely elusive. We thus investigated the Projection-TAG expression in two
132 cortical projections with notably long axons that may represent the upper limit of waiting time: MY, one of
133 the longest cortical projections in the brain, and SC_S, one of the longest cortical projections in the central
134 nervous system. We injected a Projection-TAG AAV into either MY or SC_S and quantified the Projection-
135 TAG expression overtime in the cortex using qPCR (Extended Data Fig. 3a). Projection-TAG expression
136 is detectable as soon as one week after injection and increased at week two, consistent with previous
137 reports. However, the expression in mice receiving MY injection increased rapidly and reached its peak
138 at week three, whereas the expression in mice receiving SC_S injection increased more steadily and
139 peaked at around week five. In both cases, expression plateaued until week ten, the endpoint of our
140 study. While the initial peak timing may vary across projections, our results suggest that the stable
141 Projection-TAG expression after peaking provides a flexible time window for synchronizing peak
142 expression in different projections and coupling Projection-TAGs with other experimental paradigms.

143 To make sure Projection-TAGs can be unbiasedly used for multiplex projection tracing, we next
144 investigated if significant viral competition exists among Projection-TAG AAVs that may reduce the tracing
145 efficiency. Prior experiments showed that, in the medial part of the posterior parietal association area of
146 the cortex (PTLp), more than 50% of neurons projecting to PAG also project to VP. We hypothesized that
147 the competition between Projection-TAG AAVs would reduce tracing efficiency in mice receiving both VP
148 and PAG injections compared to mice receiving only one of the injections. However, neither VP- nor PAG-
149 TAG+ cell counts are significantly different from mice receiving either injection than mice receiving both
150 injections, regardless of whether the injections were made simultaneously or separately (Extended Data
151 Fig. 3d-f). Additionally, VP- and PAG-TAG UMI counts are not significantly different in most snRNA-seq
152 libraries from nuclei expressing one of the TAGs than nuclei co-expressing both TAGs (Extended Data
153 Fig. 3g,h). These results suggest that the Projection-TAG expression is largely unaffected by viral
154 competition and thus can be safely used for multiplex tracing.

155 Moreover, we tested if Projection-TAG AAVs alter the gene expression and induce immune
156 response in infected cells. In snRNA-seq, Projection-TAG+ nuclei (nuclei with >0 UMI for any Projection-
157 TAGs) co-cluster well with Projection-TAG- nuclei on the UMAP, indicating little gene expression alteration
158 due to AAV infection (Extended Data Fig. 3b). Gene ontology analysis revealed no significant immune
159 response to viral infection in Projection-TAG+ nuclei compared to Projection-TAG- nuclei (Extended Data
160 Fig. 3c). Therefore, rAAV2-retro stably express Projection-TAGs with minimal immune response.

161 **Detection of Projection-TAGs with high specificity and efficiency**

162 Multiplex projection tracing relies on the demultiplexing of individual Projection-TAGs with various
163 detection assays. In this section, we examined the specificity and efficiency of Projection-TAG detection

164 in multiplexed FISH and snRNA-seq and highlighted some considerations that may potentially affect the
165 interpretation of Projection-TAG experimental results.

166 To examine the specificity of Projection-TAG detection in vivo, we performed multiplexed FISH
167 on brain sections from mice receiving Projection-TAG AAV injections into the projection targets of the
168 cortex (Fig 1a, Extended Data Fig. 5a). We observed a high correspondence in Projection-TAG detection
169 at the injection sites with 19.5 – 393 times greater cells labeled with Projection-TAGs injected into a given
170 region than those injected into other regions (Fig. 1b, Extended Data Fig. 5b,c). We next compared the
171 efficiency of Projection-TAG labeling to fluorescent labeling by examining MOp cells retrogradely labeled
172 with Projection-TAGs in multiplexed FISH and those labeled with the Sun1-GFP fluorescence, expressed
173 by the same Projection-TAG plasmids. We observed a high degree of overlap between Projection-TAG+
174 cells and Sun1-GFP+ cells (Fig. 1c, Extended Data Fig. 3i,j), despite Sun1-GFP and Projection-TAGs
175 exhibiting distinct subcellular compartmentalization (Fig. 1c). Therefore, Projection-TAG labeling
176 demonstrated high specificity and efficiency in multiplexed FISH, and that GFP fluorescence can serve
177 as a surrogate to estimate Projection-TAG+ cells in vivo.

178 As different 3' scRNA-seq technologies use varying strategies to capture RNA transcripts and
179 construct sequencing libraries, which may affect the 3' gene feature detection in sequencing, we next
180 compared the Projection-TAG detection by splitting one nuclear resuspension sample into two scRNA-
181 seq assays (Supplementary Table 2). The 10X Genomics assay yielded higher Projection-TAG detection
182 and identified proportionally more Projection-TAG+ nuclei compared to that from Parse Biosciences,
183 despite lower UMI recovery (Extended Data Fig. 4a). Therefore, we used the 10X Genomics assay for
184 the sequencing experiments described in this study. Besides sequencing assays, sequencing setups also
185 affect Projection-TAG detection. We targeted at least 100k reads/nucleus or 80% saturation rate for
186 library sequencing, which recovered 85.2% of the Projection-TAG UMIs and 90.5% of Projection-TAG+
187 nuclei compared to sequencing at 500k reads/nucleus (Extended Data Fig. 4b). Sequencing read 2 length
188 exceeding 75-nt had little effect on Projection-TAG detection (Extended Data Fig. 4c). To further improve
189 detection and reduce overall sequencing cost, we devised a PCR-based protocol that target amplifies
190 Projection-TAG UMIs from the cDNA library (Methods). Target amplification increased the discovery of
191 Projection-TAG+ nuclei by 1.2 ± 0.1 fold and detection of positive nuclei for individual Projection-TAGs
192 by 1.4 ± 0.6 fold per library (Extended Data Fig. 4d). Notably, target amplification recovered proportionally
193 more Projection-TAG+ nuclei when the library is shallowly sequenced (Extended Data Fig. 4e), while
194 maintaining high Projection-TAG detection.

195 To examine the efficiency of Projection-TAG detection, we performed snRNA-seq on FACS sorted
196 nuclei, in which >95% of the nuclei are GFP+, from mice receiving Projection-TAG AAV injections into the
197 projection targets of the cortex (Fig. 1a). We identified significantly more Projection-TAG+ nuclei (32.8 ±
198 5.7% with target amplification, 27.4 ± 5.6% in snRNA-seq library alone) than *Sun1-GFP*+ nuclei (Fig. 1d).
199 To investigate the specificity of Projection-TAG detection, we assessed the Projection-TAG mismatching
200 in snRNA-seq. Projection-TAGs used in experiments are highly expressed (75.9 ± 6.6 percentile by
201 expression ranks among all genes) while Projection-TAGs not used yielded zero counts in both snRNA-
202 seq library and target amplification. We also assessed the false positive rate of Projection-TAG detection
203 due to ambient RNA contamination and clustering and annotation errors. Projection-TAGs were detected
204 in only 0.1 ± 0.08% of the empty droplets per library (Fig. 1e) and 0.32% of the non-neuronal cells. Target
205 amplification marginally increased the average Projection-TAG detection in empty droplets and non-
206 neuronal cells to 0.104% and 0.34%, respectively. These data suggest that in snRNA-seq, Projection-
207 TAG detection is highly specific and more efficient compared to the detection of GFP expressed by the

same plasmid. To control for false discovery due to technical artifacts, we applied the FDR at 0.34% in the downstream Projection-TAG analyses (Methods).

However, one should be mindful of several limitations when interpreting the Projection-TAG results. First, while the Projection-TAG detection in multiplexed FISH assay is highly efficient (only 0.2% GFP+ cells do not express Projection-TAGs), the false negative rate of Projection-TAG detection is nontrivial in snRNA-seq (67.2% ± 5.7% nuclei from snRNA-seq libraries prepared by FACS do not express Projection-TAGs, discussion). Consequently, a lack of Projection-TAG expression in a snRNA-seq cell should not be simply interpreted as lack of projection. In addition, technical and biological variables may introduce bias in Projection-TAG tracing efficiency. For example, the distribution of cells labeled for each projection differs across animals (Extended Data Fig. 4f), likely due to variations of stereotaxic injections, and the Projection-TAG expression is different across projection pathways (Extended Data Fig. 4g). However, the exact TAGs used to trace each projection pathway did not significantly affect their expression in most cases (Extended Data Fig. 4h), suggesting Projection-TAGs can be used interchangeably in tracing experiments.

Spatial organization of projection neurons across the cortex

As it has been well established that neurons projecting to IT and ET targets exhibit distinct spatial distribution across the cortex[26,63], we next sought to validate the spatial distribution of Projection-TAG labeled neurons following AAV injection into the projection targets mentioned above. We quantified the distribution of neurons projecting to each target in several neocortical areas (Extended Data Fig. 6c). We observed that neurons projecting to different targets are enriched in distinct areas across the cortex (Fig. 1f, g). For example, cMOp-projecting neurons are most enriched in the ACA, MOs, and the medial part of MOp, whereas cSSp-projecting neurons are dominantly found in MOp and SSp. While VP-, PAG-, and MY-projecting neurons are distributed across multiple cortex areas, SC_L- and SC_S-projecting neurons are highly enriched around MOp and SSp. The spatial distribution of projection neurons is largely consistent with previous reports[23,64-66]. It has been reported that thalamus- and MY-projecting neurons in the anterior lateral motor cortex exhibit distinct spatial distribution in layer 5 (L5)[26,67]. Similarly, we observed that VP-projecting neurons, while also found in L6, were more enriched in L5a than L5b (p = 0.005), whereas MY-projecting neurons were more abundant in L5b than L5a (p = 0.002) in the anterior part of the MOp (Extended Data Fig. 6d). Interestingly, their distinct sub-layer distribution appeared to attenuate towards the posterior MOp, accompanied by an increased proportion of neurons co-projecting to both targets. While numerous cortex areas contain neurons projecting to the seven targets examined in this study, our analysis highlights the MOp and the SSp as particularly rich in projection neurons among all cortex areas analyzed and are highly enriched with neurons projecting to each of the seven targets (Extended Data Fig. 6e).

A multi-modal single-cell atlas of mouse cortex

We next assessed whether we could profile simultaneously the gene expression, chromatin accessibility, and projection feature of the same cells using Projection-TAGs in multiomic analysis of snRNA-seq and single-nucleus ATAC-seq (snATAC-seq, Extended Data Fig. 7a). We first performed snRNA-seq on four MOp samples and six SSp samples from mice receiving injections of Projection-TAG AAVs into seven downstream projection targets mentioned above, generating 14 libraries containing a total of 69,657 nuclei with an average of 3,621 genes per nucleus (Extended Data Fig. 7b, Supplementary Table 2). After quality control and removal of low-quality nuclei and doublets, 61,387 nuclei were retained in the snRNA-

seq dataset. Notably, nuclei from individual libraries co-clustered together, indicating a low batch effect across libraries and largely shared gene expression profiles between nuclei from MOp and SSp (Extended Data Fig. 7c,d). Clustering analysis revealed 35 distinct transcriptional clusters, which were further classified into three major classes (Glutamatergic neurons, GABAergic neurons, and non-neuronal cells, Fig. 2a, Methods). We assigned each cluster to a known cell type, based on the

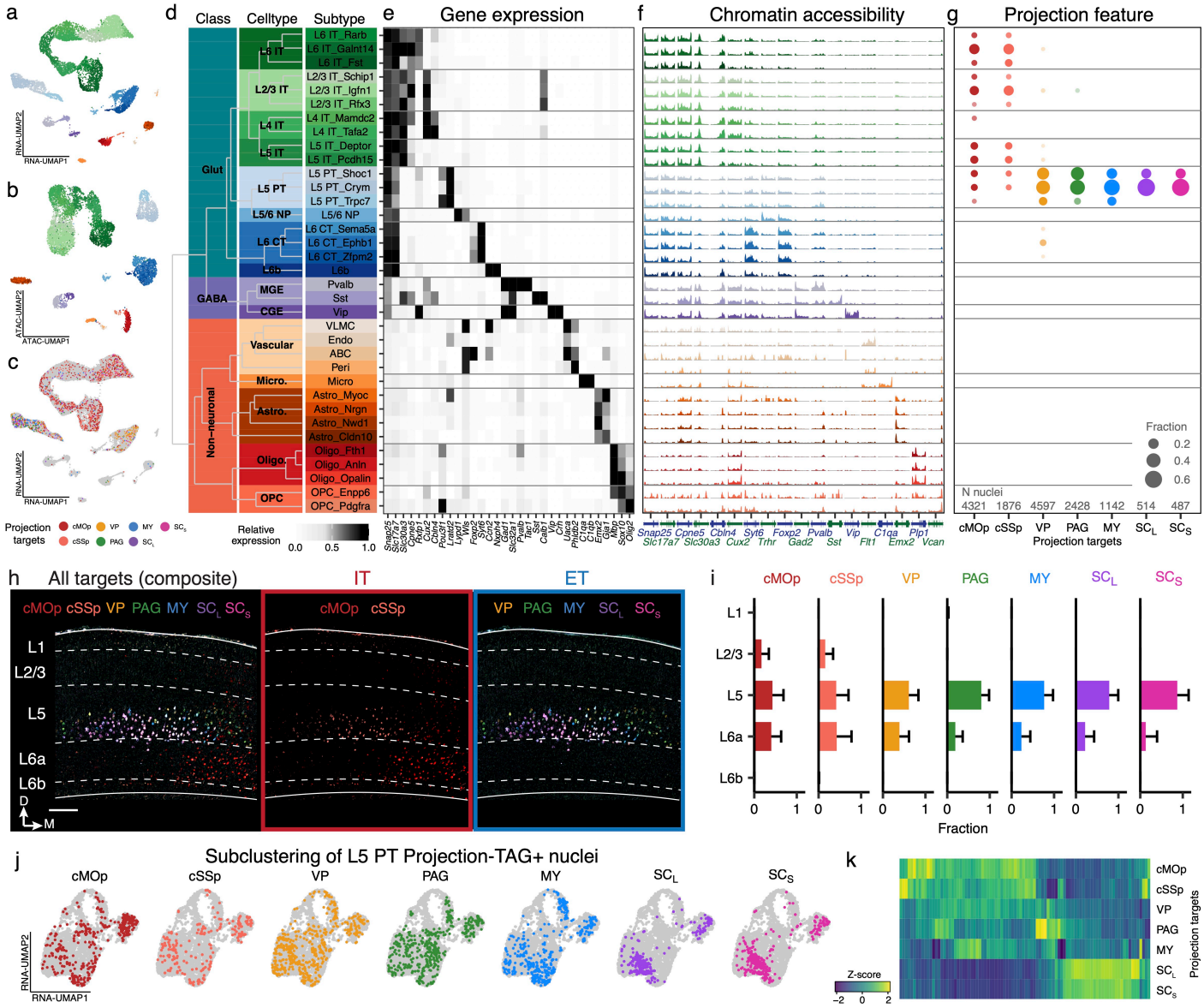


Figure 2. A multi-modal single-cell atlas of mouse cortex

a, Uniform manifold approximation and projection (UMAP)[1] visualization of snRNA-seq data showing 10,000 downsampled nuclei, colored by transcriptional subtypes. b, UMAP visualization of snATAC-seq data showing 10,000 downsampled nuclei, colored by transcriptional subtypes. c, UMAP visualization of snRNA-seq nuclei, colored by their projection targets. 10,000 downsampled Projection-TAG- nuclei are colored grey and displayed as the background. 500 nuclei (or all nuclei if total nuclei count of one projection is less than 500) expressing only the corresponding Projection-TAG are downsampled for each projection target and displayed. d, Hierarchical clustering of 35 snRNA-seq clusters based on the average expression of top 100 marker genes (by FDR) per cluster, and their annotations of class, celltype, and subtype. e, Heatmap showing the average expression of select marker genes in individual snRNA-seq clusters. f, Coverage plot of average accessibility of canonical celltype-specific marker genes in snATAC-seq nuclei, grouped by transcriptional clusters. The chromatin accessibility is displayed as the average frequency of sequenced DNA fragments per nucleus for each cluster, grouped by 50 bins per displayed genomic region. g, Dot plot showing fraction of snRNA-seq nuclei, positive for Projection-TAG of each projection, from each transcriptional cluster. h, A representative FISH image showing the layer distribution of projection neurons in MOp from a brain section (-0.91 mm relative to bregma). Colors denote the projection targets based on the expression of Projection-TAGs. Scale bar = 150 μm. Left: composite image of all FISH channels, middle and right: images of a subset of Projection-TAGs injected into IT targets and ET targets, respectively. i, FISH quantification showing the fraction of MOp cells retrogradely labeled by each Projection-TAG in each layer. Bar and error bar indicate the average and standard deviation across six slices from two mice. j, RNA-UMAP showing subclustering of L5 PT neurons labeled with each Projection-TAG. k, Heatmap showing the z-score of average expression of genes (column) differentially expressed in L5-PT neurons positive for each Projection-TAG (row). Data shown on the heatmap can be found in Supplementary Table 4.

expression of canonical marker genes (Fig. 2d,e, Methods). Previous studies revealed that different glutamatergic neuronal types, characterized with different anatomical properties such as layer distributions and projection targets, exhibited distinct gene expression profiles[27,43,60]. Following this naming convention, we assigned glutamatergic celltypes to 18 glutamatergic neuronal clusters based on their layer enrichment and projection patterns previously described: IT neurons from layers 2/3 (L2/3 IT), 4 (L4 IT), 5 (L5 IT), and 6 (L6 IT), pyramidal tract neurons from layer 5 (L5 PT), near-projecting neurons from layers 5/6 (L5/6 NP), corticothalamic-projecting neurons from layer 6 (L6 CT), and neurons from layer 6b (L6b). We defined three GABAergic neuronal celltypes based on developmental origins: medial ganglionic eminence (MGE) neurons and caudal ganglionic eminence (CGE) neurons. We also identified five non-neuronal celltypes: vascular cells, microglial cells (Micro.), astrocytes (Astro.), oligodendrocytes (Oligo.), and oligodendrocyte progenitor cells (OPC). We assigned subtypes to each cluster based on the marker gene distinctly expressed in that cluster compared to other clusters of the same celltype (Extended Data Fig. 7e,f). While we observed some variability of celltype distribution across libraries, likely due to different nuclear preparation approaches (Extended Data Fig. 7d), the gene expression profiles of individual subtypes are largely consistent with previous reports[15,26,61] (Extended Data Fig. 7g, Supplementary Table 3).

To enable simultaneous investigation of the gene expression and chromatin accessibility profiles in the same cells, we performed combinatorial snATAC-seq on eight snRNA-seq libraries reported above, generating a dataset of 40,188 high-quality nuclei with an average sequencing depth of 25,412 transposase-sensitive fragments per nucleus (Extended Data Fig. 8a)[68]. These snATAC-seq fragments captured the open chromatin regions of the genome as they followed the expected nucleosomal size distribution (Extended Data Fig. 8b). Clustering analysis of the snATAC-seq data revealed 28 distinct clusters. The chromatin accessibility and gene expression profiles of the sequenced nuclei are highly correlated, with 87.7 ± 18.0% nuclei in each snATAC-seq cluster assigned to the same transcriptional clusters as those made when independently analyzed for snRNA-seq data (Fig. 2b, Extended Data Fig. 7c). In addition, nuclei in individual snATAC-seq clusters display distinct chromatin accessibility around the genomic loci of canonical marker genes for the corresponding transcriptional subtypes (Fig. 2f and Supplementary Table 3). Consequently, we grouped snATAC-seq nuclei by transcriptionally defined celltypes and subtypes in subsequent analysis.

Detection and demultiplexing of Projection-TAGs by snRNA-seq enabled us to investigate the projection feature of individual sequenced cells, thus allowing multi-modal profiling of projection neurons at single-cell resolution (Fig. 2c). We determined whether a snRNA-seq nucleus project to each of the seven targets based on its expression of the corresponding TAG (Methods). To correlate the transcriptional identity with projection targets, we first examined the subtype composition of nuclei that are positive for each TAG (Fig 2f, extended Data Fig. 8d). In this analysis, a nucleus positive for the corresponding Projection-TAG will be included for the analysis of a projection target, regardless of the expression of other Projection-TAGs in the same nucleus (Methods). Hierarchical clustering based on the subtype composition for each project target divided the seven targets into three clusters (Extended Data Fig. 8e). The first cluster contains two IT targets, cMOp and cSSp. They both consist of transcriptionally-defined L2/3 IT, L4 IT, L5 IT, L6 IT, and L5 PT neurons and their subtype composition are not significantly different from each other (23.2% and 27% L2/3 IT, 4.9% and 4% L4 IT, 20.6% and 18.4% L5 IT, 29.9% and 35.8% L6 IT, 19% and 12.5% L5 PT, MOp and SSp, respectively). The second cluster contains three ET/subcortical targets: VP, PAG, and MY. While they are all enriched of L5 PT neurons (78.1% VP, 85.4% PAG, and 86.6% MY), MY-projecting neurons have proportionally fewer nuclei from L5 PT_Shoc1 cluster (20.7%, compared to 28.3% VP and 33.7% PAG). In addition, 10.4% of the VP-

projecting neurons are from transcriptionally defined L6 CT subtypes. The third cluster contains two ET/corticospinal targets with SC_L and SC_S projecting neurons are enriched of L5 PT subtypes but are underrepresented in L5 PT_Trpc7 subtype (3.9% each) compared to other ET projections (subcortical: VP, PAG, and MY). Spatial analysis of multiplexed FISH further confirmed their layer distribution in MOp (Fig. 2h,i). The cell types of projection neurons are consistent with the explicit correspondence between neurons projecting to IT and ET targets previously described[26,63].

As transcriptionally defined L5 PT neurons are enriched of nuclei positive for Projection-TAGs of all seven projection targets, we next asked whether L5 PT neurons projecting to different targets are transcriptionally distinct. Subclustering of 4,168 L5 PT Projection-TAG+ nuclei present in both snRNA-seq and snATAC-seq data revealed distinct transcriptional and epigenetic profiles of L5 PT neurons projecting to each target (Fig 2j, k, Supplementary Table 4). *Hpgd* and *Slco2a1* have been reported as marker genes for thalamus- and MY-projecting neurons, respectively[26]. Interestingly, we observed an expression gradient of those genes across L5 PT neurons projecting to different targets (Extended Data Fig. 8f), which appears to correlate to the distance of the projections, raising the possibility that the projection targets of neurons may be dictated/maintained by a shared, fine-tuned transcriptional program.

Projection-TAGs revealed axonal collaterals and complex wiring diagram

Brain regions are interconnected with complex wiring diagrams. If a brain region projects to N downstream targets, the possible projection pattern of a single neuron will have up to 2^N combinations. As Projection-TAGs enable multiplex projection tracing in single animals, it opens new possibilities for studying neuronal collaterals to multiple targets. A neuron with its axonal collaterals terminated in multiple targets may be labeled by multiple Projection-TAGs. Indeed, 30.6% of Projection-TAG+ snRNA-seq nuclei express multiple Projection-TAGs (Fig. 3a). To investigate the overall pattern of axonal collaterals, we first calculated the overlap of neurons projecting to any two targets (Fig. 3b). Two IT targets (cMOp and cSSp) exhibited highly significant overlap (Extended Data Fig. 9a) with each other with 16.8% MOp-TAG+ nuclei or 36.8% SSp-TAG+ nuclei are co-labeled by both IT-TAGs. ET targets also showed highly significant overlap with each other, with the most notable overlapping to VP. 56.3 – 66.7% nuclei labeled with other ET-TAGs (PAG, MY, SC_L, and SC_S) are also co-labeled with VP-TAG. Additionally, PAG and MY also highly significantly overlap, as well as SC_L and SC_S. To elucidate the complex projection feature of single neurons, we next investigated any possible combinations of projection to the seven targets. Out of all 128 (2^7) possible projection patterns, we identified 16 neuronal populations with distinct projection patterns that passed the FDR cutoff: seven populations with single projections (positive for only one Projection-TAG) (Fig. 3c) and nine populations with multiple projections to two or three of the targets (positive for the Projection-TAGs of the corresponding projection targets and negative for all other Projection-TAGs) (Fig. 3d). The axonal collaterals were further confirmed by multiplexed FISH (Extended Data Fig 9b). These observations are highly correlated with the anatomical hierarchy of axonal projections of the cortex previously demonstrated by anterograde bulk tracing techniques (Extended Data Fig. 9c,d). Moreover, the hierarchical organization of projections is also reflected in the gene expression and chromatin accessibility profiles of neurons projecting to individual targets (Extended Data Fig. 9e,f). This complementary analysis provides a multi-modal and high-resolution view of the spatial and molecular intricacies of neuronal projections in the mouse brain.

Among 5,358 snRNA-seq nuclei positive for only ET-TAGs, 43.6% and 26.1% express only VP-TAG and only other ET-TAGs (PAG, MY, SC_L, and SC_S), respectively, and 30.3% express both VP- and other ET-TAGs (Fig. 3e,f). Transcriptionally defined CT neurons restrict their projection to only VP, while

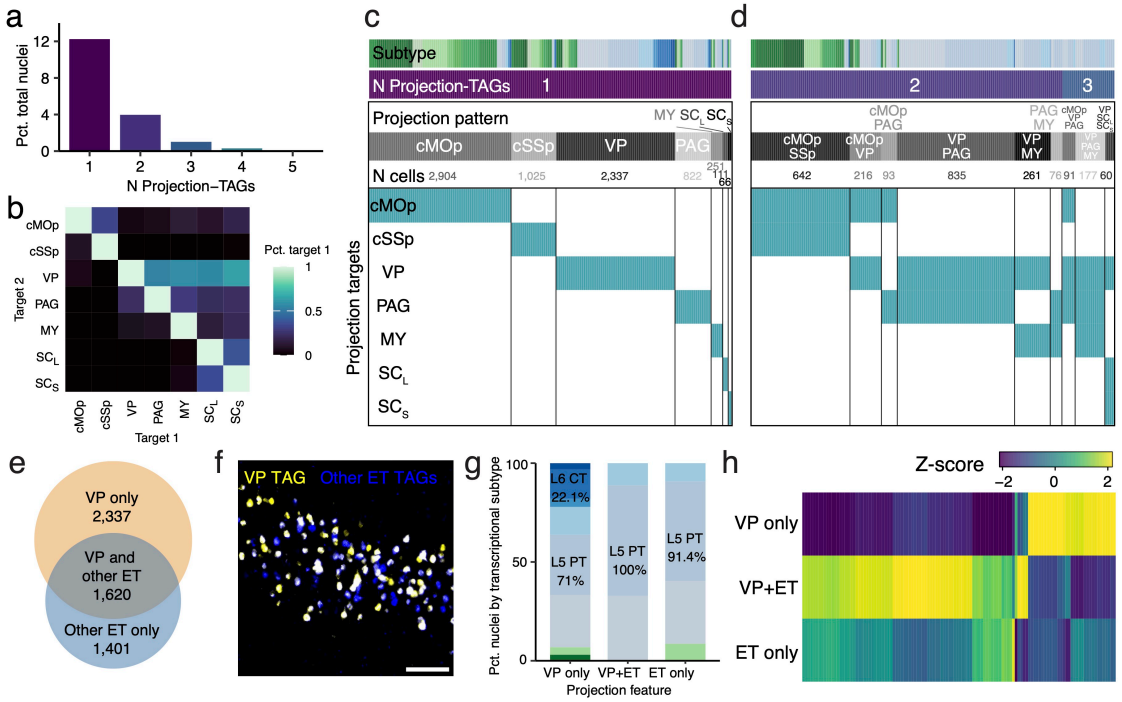


Figure 3. Projection-TAGs revealed axonal collaterals and complex wiring diagram
a, Bar plot showing the percentage of snRNA-seq nuclei, grouped by the number of unique Projection-TAGs detected in each nucleus. **b**, Heatmap demonstrating the pairwise overlapping of snRNA-seq nuclei positive for any two Projection-TAGs, shown as the percentage of nuclei positive for target 1 TAG (column) that are also positive for target 2 TAG (row). **c** and **d**, Heatmaps showing the Projection-TAG (row) detected in individual snRNA-seq nuclei (column) expressing only one Projection-TAG (c) and multiple Projection-TAGs (d). Nuclei are ordered based on the projection patterns, and only groups that passed the FDR cutoff (at least 60 nuclei) were shown on the heatmaps. **e**, Venn diagram showing the overlap between VP-TAG-expressing and other ET-TAG-expressing snRNA-seq nuclei. Other ET-TAGs include Projection-TAGs injected into PAG, MY, SC_L, and SC_S. **f**, Representative FISH image showing MOp cells labeled with VP- and other ET-TAGs. Scale bar = 100 µm. **g**, Distribution of transcriptional subtypes in each projection group. Transcriptional subtypes with < 60 nuclei from each projection feature group were excluded and celltypes that make up at least 10% nuclei in the projection group are annotated on the plot. **h**, Heatmap showing the z-score of average expression of 118 genes (row) differentially expressed in transcriptional L5 PT nuclei in each projection group (column). Data shown on the heatmap can be found in Supplementary Table 5.

transcriptionally defined PT neurons broadly project to various ET targets (Fig. 3g). Our observations of axonal collaterals corroborate with the intricate single-neuron projection patterns reported in the MouseLight dataset (Extended Data Fig. 10 a,b)[69] and further demonstrate the complexity of single-neuron projections, which has been overlooked and warrants further investigation. To uncover transcriptional programs that fine-tune cortical projection patterns, we identified 1,971 differentially expressed (DE) genes and 2,737 differentially accessible (DA) peaks in neurons with single projections compared to those with multiple projections (Fig. 3h, Supplementary Table 5).

Characterization of genomic cis-regulatory elements and regulated genes using Projection-TAGs

While the gene expression program in the diverse MOp and SSp cell types has been largely elucidated by scRNA-seq studies, the regulatory networks that govern the distinct gene expression pattern in individual cell types and projections are under investigated[22,44,60,72]. The genomic regulatory elements (GREs), mostly identified within the non-coding region of the genome and act in a cell-type-specific and tissue-specific manner, fine tune the expression level of the regulated genes[73]. Among different types of GREs, enhancers initiate the recruitment of transcription complexes and drive the transcription of regulated genes while the silencers prevent the expression of regulated genes[74,75]. Identification of cell-type-specific GREs, specifically enhancers, has gained increased attention in the neuroscience field as they can be used as a valuable tool to mediate transgene (e.g eYFP, ChR2, DREADDS, etc) expression in the target cell populations for basic science research and potentially the development of novel therapeutics[76-81]. Though recent progress has been made in characterizing the landscape of GREs for the projection neurons, the current experimental workflow is tedious and pain-

staking as individual projection neurons are labeled via injection of retro Cre in floxed nuclear reporter mice, and regions of interest are pooled for analysis of potential GREs[22,44]. We tested whether Projection-TAGs could potentially be used to perform high throughput analysis of identifying projection-specific GREs.

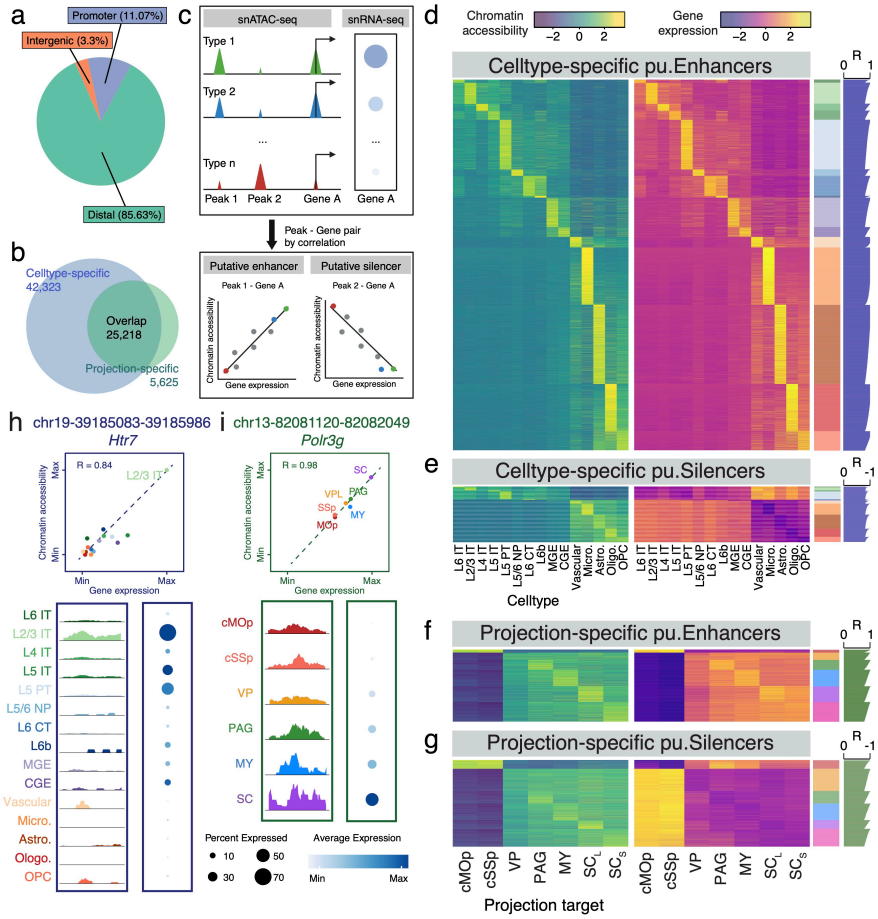


Figure 4. Genomic regulatory networks of cortical cell types using Projection-TAGs
a, Pie chart categorizing snATAC-seq peaks based on their genomic locations. Peaks are mapped to one of the three categories: promoter regions (-1,000 bp to +100 bp of transcription start site [TSS]), distal regions (<200 kb upstream or downstream of TSS or within gene body, excluding promoter), and intergenic regions (>200kb upstream or downstream of TSS, excluding gene body). b, Overlapping of celltype-specific peaks and projection-specific peaks in snATAC-seq data. c, Analytical workflow of identifying putative enhancers and silencers and their regulated genes. d and e, Heatmaps showing 18,088 pairs of celltype-specific pu.Enhancers and regulated genes (d) and 2,739 pairs of celltype-specific pu.Silencers and regulated genes (e). Left heatmaps show the average accessibility of pu.Enhancer peaks (row) in individual celltypes (column) in snATAC-seq data and right heatmaps show the average expression of regulated genes (row) in individual celltypes (column) in snRNA-seq data. On the right side next to the heatmaps, the color bar indicates the celltypes in which the peak and gene have highest accessibility and expression, and the bar plot shows the Pearson correlation coefficient (R) between the accessibility of the peak and the expression of the gene in individual celltypes. Data shown on the plots can be found in Supplementary Table 6. f and g, Heatmaps showing 3,545 pairs of projection-specific puEnhancers and regulated genes (f) and 4,200 pairs of projection-specific pu.Silencers and regulated genes (g). Left heatmaps show the average accessibility of pu.Enhancer peaks (row) in nuclei positive for each Projection-TAG (column) in snATAC-seq data and right heatmaps show the average expression of regulated genes (row) in nuclei positive for each Projection-TAG (column) in snRNA-seq data. On the right side next to the heatmaps, the color bar indicates the projection targets in which the peak and gene have highest accessibility and expression, and the bar plot shows the Pearson correlation coefficient (R) between the accessibility of the peak and the expression of the gene in nuclei positive for each Projection-TAG. Data shown on the plots can be found in Supplementary Table 6. h, Top: Scatter plot showing the correlation between the average accessibility of peak chr19-39,185,083-39,185,986 in snATAC-seq and the average expression of gene *Htr7* in snRNA-seq in individual celltypes (chromatin accessibility and gene expression are normalized to their max values). The dotted line represents the line of best fit. Bottom left: chromatin accessibility at the genomic locus of chr19-39,185,083-39,185,986, displayed as the average fraction of transposase-sensitive fragments per nucleus in each celltype (grouped by 50 bins per displayed genomic region). Accessibility at each locus (y axis) is scaled to the max value across all cell types. Bottom right: expression of *Htr7* in each celltype. i, Top: Scatter plot showing the correlation between the average accessibility of peak chr13-82,081,120-82,082,049 and the average expression of gene *Polr3g* in nuclei positive for each Projection-TAG (chromatin accessibility and gene expression are normalized to their max values). Bottom left: chromatin accessibility at the genomic locus of chr13-82,081,120-82,082,049, displayed as the average fraction of transposase-sensitive fragments per nucleus in nuclei positive ach Projection-TAG (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Polr3g* in nuclei positive for ach Projection-TAG. Colors indicate the injection site of the corresponding TAG. Nuclei positive for SC_L- and SC_S-TAGs are combined for visualization due to low cell number.

To this end, we integrated the snRNA-seq and snATAC-seq data from MOp and SSp to produce a unified, multi-modal cell census. We first identified 166,540 peaks when aggregated across all snATAC-seq libraries. We believe many of those peaks are likely to contain functionally relevant GREs, as 85.6% of the peaks are located at the distal regions in the genome (Fig. 4a), and 43.9% of them are differentially accessible in individual transcriptional celltypes and/or neurons projecting to different targets (Fig. 4b). To identify putative GREs and their correspondence to genes they regulate, we first identified 67,541 D peaks (Log2FC > 1, FDR < 0.05) in transcriptional celltypes and 30,843 DA peaks in neurons projecting to each target, we then calculated the average expression of genes from snRNA-seq data and the average accessibility of DA peaks from snATAC-seq data in each celltypes or projections. We next calculated the Pearson's correlation between the accessibility of DA peaks with the expression of genes for any peak-gene pairs that are within a 5M bp window of the same chromosome (Fig. 4c). A peak-gene pair with a strong positive correlation (Pearson's r > 0.75) is identified as a putative enhancer (pu.Enhancer) and its putative regulated gene, whereas a pair with strong negative correlation (Pearson's r < -0.75) is categorized as a putative silencer (pu.Silencer) and its regulated gene. We identified 18,088 pu.Enhancer-gene pairs and 2,739 pu.Silencer-gene pairs that are celltype-specific (Fig. 4d,e, Supplementary Table 6), and 3,545 pu.Enhancer-gene pairs and 4,200 pu.Silencer-gene pairs that are projection-specific (Fig. 4f,g, Supplementary Table 6). Several pieces of evidence support the authenticity of identified putative GREs. First, they exhibited a greater overlap with the DNase hypersensitive sites (96.4%) detected in the adult mouse brain compared to randomly selected genomic regions (42.7%) of similar sizes and GC contents[82]. Second, these putative GREs show significant alignment with GREs previously annotated by the ENCODE consortium, particularly those identified in the mouse brain as opposed to other organs (Extended Data Fig. 11a)[83]. Furthermore, our analysis reveals four experimentally validated functional enhancers that coincide with peaks identified in our snATAC-seq data[77]. Notably, two of these validated enhancers overlap with the celltype-specific pu.Enhancers identified in our analysis, and the predicted transcriptional celltypes aligned with the cell types demonstrating the highest enhancer activity as verified experimentally (Extended Data Fig. 11b).

Among identified putative GREs, the peak 39,185,083 - 39,185,986 on chromosome 19 is ~3.2Mbp upstream of the TSS of *Htr7* gene (Fig. 4h). This peak, differentially accessible in L2/3 IT neurons, is positively correlated (Pearson's r = 0.84) to *Htr7*, which is differentially expressed in the same celltype, suggesting that this peak may contain a pu.Enhancer that might drive the expression of *Htr7* specifically in L2/3 IT neurons. The accessibility of peak 82,081,120 - 82,082,049 on chromosome 13 is positively correlated to the expression of *Polr3g* in neurons of different projections (Fig. 4i), Both the peak and gene are highly accessed/expressed in corticospinal neurons, suggesting this peak may contain a pu.Enhancer that might drive the expression of *Polr3g* specifically in the corticospinal projection neurons. Additionally, peaks chr19-10,784,771-10,785,580 and chr18-39,362,615-39,363,500 may contain celltype-specific and projection-specific pu.Silencers, respectively. While accessible in the genome, they may reduce the expression of *Fth1* in the IT neurons and *Kctd16* in ET-projecting neurons, respectively (Extended Data Fig. 11c,d). Thus, Projection-TAGs offer a powerful, high-throughput platform to perform systemic multiomic analyses to gain insight into the gene expression and chromatin accessibility profiles of diverse projection neurons.

Projection-TAGs enable the detection of projection neurons tuned to a behavioral stimulus

Projection-TAGs allow us to elucidate the spatial, gene expression, and chromatin accessibility profiles of diverse projection neurons at single-cell resolution. As Projection-TAG AAVs mediate stable BC expression with minimal immune response, we sought to leverage it to obtain a snapshot of activity-dependent recruitment of distinct projection neurons and their molecular features in the context of a

specific stimulus. The MOp and the SSp circuitry have been extensively implicated in the modulation of distinct pain modalities[84-89], we wanted to gain insight into the gene expression changes in the cell types and their projections neurons in response to visceral pain stimulus. To study the cell populations acutely activated by visceral pain, we first labeled the seven projection targets of the MOp and SSp as described earlier (Fig. 1a). On the day of experiment, we induced acute inflammatory visceral pain with cyclophosphamide (CYP) or inject saline as the control, followed by combinatorial snRNA-seq and snATAC-seq of the MOp and SSp 30 minutes after the stimulus, at which we observed significant spontaneous behaviors (Extended Data Fig. 12a).

In both snRNA-seq and snATAC-seq data, nuclei from CYP- and saline-treated mice co-cluster together, suggesting visceral pain did not significantly alter the gene expression or chromatin accessibility of MOp and SSp cell types at the acute time point (Extended Data Fig. 12b). To pinpoint the cell populations that are acutely activated by visceral pain, we next applied Act-seq[90-92], an approach that links the transcriptional activity following stimulus to individual neuronal cell types and projections based on the expression of immediate early genes (IEG). To aggregate expression across IEGs, we generated an IEG score for each snRNA-seq nucleus (Methods). While the IEG scores did not significantly differ in

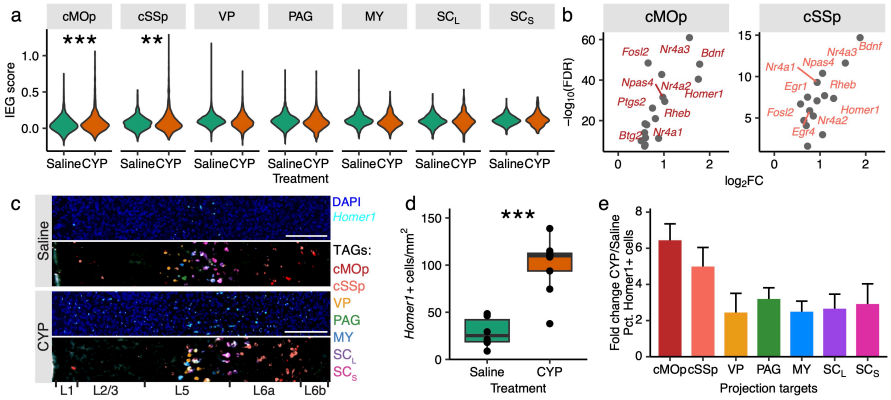


Figure 5. Projection-TAGs enable the detection of projection neurons activated by visceral pain
a, Violin plots showing IEG scores in snRNA-seq nuclei for each projection between treatments (***p = 4.6e-14 for cMOp, **p = 0.004 for cSSp, p = 1 for VP, p = 0.99 for PAG, p = 0.99 for MY, p = 0.4 for SC_L, p = 0.2 for SC_S, one-side t-test with Bonferroni correction). **b**, Volcano plots showing significantly induced IEGs (log2FC > 0.5) in activated nuclei from CYP-treated mice compared to the same number of randomly sampled nuclei (with matched transcriptional subtypes) from saline-treated mice for the two IT projections. **c**, Representative FISH images showing the expression of *Homer1* and Projection-TAGs in MOp (from brain section approximately -0.2 mm relative to bregma) from saline- and CYP-treated mice. Scale bar: 100 µm. **d**, Quantification of *Homer1*+ cells in MOp and SSp areas from CYP- and saline-treated mice (0 to -0.92 mm relative to bregma, nine slices from three mice per group, ***p = 1.9e-5, one-side t-test). **e**, Percentage of *Homer1*+ MOp and SSp cells in each projection. Fold change is calculated as the percentage in individual CYP slices, divided by the average percentage across saline slices. Bar and error bar indicate the mean and standard deviations across nine slices per group.

all nuclei between treatments (Extended Data Fig. 12c), CYP significantly activated 12 transcriptional subtypes, including seven IT subtypes and two GABAergic subtypes (Extended Data Fig. 12d), suggesting their preferential recruitments following the acute visceral pain stimulus. Further analysis of the projections suggests that CYP significantly activated only IT projections (both cMOp and cSSp), while none of the ET-projecting populations exhibited significant activation (Fig. 5a). Our results pinpointed the cell types and projections that visceral pain selectively recruits in the acute phase of the pain induction state.

We next expanded our analysis to determine the IEG profiles of the activated IT-projecting neurons, as it has been recently appreciated that distinct stimuli exhibit varying IEG activation profiles[89,93-96]. DE analysis comparing all nuclei of the same projection between treatments failed to identify any CYP-induced IEGs. To maximize "signal-to-noise" ratio, we first identified transcriptionally "activated" nuclei based on their IEG scores (Methods), followed by DE analysis comparing activated nuclei from CYP-treated mice to inactivated nuclei from Saline-treated mice. We found that CYP induced 17 IEGs in activated neurons projecting to IT targets (Log2FC > 0.5 in either cMOp or cSSp, Fig. 5b). Among them, *Nr4a3*, *Bdnf*, *Rheb*, and *Homer1* showed higher fold change (Log2FC > 1) in both cMOp- and cSSp-

projecting neurons. Interestingly, CYP did not significantly induce *Fos* expression (Fig. 5b, Extended Data Fig. 12e). *Homer1* expression is directly linked to neuronal activity and synaptic plasticity[97-100]. To validate these findings, we performed multiplexed FISH screening for *Homer1* as an activity-dependent marker for the CYP-induced pain state. We observed that CYP increased *Homer1* expression in MOp and SSp compared to saline (Fig. 5c,d). Next, to investigate if *Homer1* is selectively enriched in the IT-projecting neurons following visceral pain, we examined the colocalization of *Homer1* with each Projection-TAG. While CYP induced *Homer1* expression in all projection pathways analyzed (Extended Data Fig. 12f), IT-projection neurons exhibited a higher magnitude of *Homer1* induction, as the percentage of *Homer1*+ cells increased by 6.4 and 5-fold in cMOp and cSSp-projecting neurons, compared to on average 2.7-fold in the ET-projecting neurons (Fig. 5e). In summary, our data shows that Projection-TAGs could be readily implemented with transcriptional analysis such Act-seq to identify the molecular markers and recruitment of distinct specific brain-wide projections and cell types following the stimulus of interest.

Discussion

Here we have developed Projection-TAGs, a retrograde AAV platform for multiplex neuroanatomical studies and high-throughput multi-modal profiling of projection neurons. Projection-TAG AAVs retrogradely label distinct projections with unique RNA barcodes with high specificity and efficiency. It can be easily adaptable to existing neuroscience workflow and is optimized for commercial assays, such as multiplexed FISH, which allows simultaneous spatial integration of multiple projections in the same animals, and single-cell profiling assays that enable multiomic profiling of projection neurons. In this study, we applied Projection-TAGs to examine the transcriptional and epigenetic landscapes of the cortex using combinatorial snRNA-seq and snATAC-seq. Projection-TAGs could facilitate the development of novel tools for projection-specific targeting/manipulation by opening new avenues for studying multiple projections in the same animals and for identifying key gene expression features and genomic regulatory elements in distinct cortical cell types and diverse projection neurons. Lastly, we demonstrated that Projection-TAGs can be incorporated with additional experimental paradigms, providing users with flexibility for studying activity-dependent recruitment of distinct cell populations and projections in the stimulus of interest.

Available high-throughput neuroanatomy tools

RNA barcode-based high-throughput neuroanatomical tools have greatly expanded the multiplexing capacity of neuronal tracing. Available high-throughput neuroanatomical techniques can be broadly classified based on their retrograde and barcoding schemes. Anterograde tracing techniques, namely MAP-seq and its derivatives, such as BRICseq, BARseq, and BARseq2[46,49,50,101]. MAPseq utilizes a sindbis viral library to encode a diverse collection of short random RNA barcodes, which act as the cell barcode/identifier. The sindbis viral library is injected into the source region, and a unique barcode is expressed in individual neurons and anterogradely transported to the axonal terminals. Multiplexed projection tracing is achieved by assaying the RNA barcodes present in each of the target regions. MAPseq can be coupled with scRNA-seq for multi-modal profiling of projection neurons[48] and with in-situ sequencing for spatial analysis (BARseq) and spatially-resolved transcriptional assays (BARseq2). Due to the anterograde tracing scheme, BARseq and BARseq2 can achieve high spatial resolution in both source and target regions, but the multiplexing capacity and tracing accuracy of MAPseq may be limited by tissue dissections. With the simple surgery procedure (one injection into the source region), MAPseq and derivatives are ideal for quantitatively measuring the projection strength. However, the requirement of highly customized instruments and pipelines by those methods has restricted their adoption beyond

485 several expert labs. The high replication rate of sindbis virus results in cellular toxicity, posing a challenge
486 to integrating their use with chronic experimental paradigms.

487 Retrograde tracing techniques, including Projection-seq, MERGE-seq, and Projection-TAGs[53,102].
488 These techniques utilize a limited number of RNA barcodes with known sequences, which act as the
489 projection barcode/identifier. A retrograde AAV expressing a unique BC is injected into each of the target
490 regions, which retrogradely label projection neurons in the source region. Multiplexed projection tracing
491 is achieved by assaying the barcodes present in the source region, which can be simply read out by
492 scRNA-seq and spatial assays. Due to the retrograde tracing scheme, Projection-TAGs and similar
493 techniques can achieve high spatial resolution in the source region, but not necessarily the target regions.
494 The surgery procedures are relatively complicated (one injection into each of the target regions), which
495 may limit the multiplexing capacity, and the accuracy of targeting each target region may introduce
496 technical variations that confound the quantitative measurement of the projection strength. Surgical
497 targeting (one injection into each of the target regions) may introduce technical variations that confound
498 the quantitative measurement of the projection strength and might limit the multiplexing capacity. Despite
499 these technical challenges, barcode detection is relatively simple and flexible and can be achieved using
500 various commercial assays. Projection-TAGs enable multiplex neuroanatomical studies and high-
501 throughput multi-modal profiling of projection neurons. AAVs have minimal cellular toxicity, which is ideal
502 for incorporating Projection-TAGs with additional experimental paradigms for studying activity-dependent
503 recruitment of distinct cell populations and projections in the stimulus of interest.

504 **Consideration for Projection-TAG experimental design and result interpretation**

505 Projection-TAGs allow multiplex projection tracing and multi-modal profiling of projection neurons.
506 In retrograde viral tracing experiments, the cargo gene expression in a single cell depends on a series of
507 biological processes such as viral attachment and internalization, transduction and trafficking to the cell
508 body, and escape and entry into the nucleus[103]. rAAV2-retro has been widely adopted in neuroscience
509 laboratories, Projection-TAG rAAV2-retro performance would be similar to other rAAV2-retro viruses, and
510 prior experience using rAAV2-retro may be used to guide the planning of Projection-TAG multiplex tracing
511 experiment. Efficient and accurate labeling of Projection-TAGs relies on successful targeting of desired
512 brain regions, transduction efficacy of rAAV2-retro, titer (Projection-TAG AAVs with the titer of
513 ~2e+12vg/ml label 14.7 ± 8.4% of MOp cells, compared 3.7 ± 3.0% by those with the titer of ~5e+11vg/ml)
514 and volume of the viruses injected. The high correlation between Projection-TAG (RNA level) and GFP
515 (protein level) expressed using the Projection-TAG AAVs provides a simple strategy for investigators to
516 estimate retrograde tracing efficiency and accuracy of target labeling, which are essential for a successful
517 Projection-TAG experiment. While Projection-TAG plasmids have the potential to be flexibly packaged
518 into other AAV serotypes, the transduction efficiency and tissue/celltype tropisms related to AAV
519 serotypes may lead to false negatives and should be considered while designing the experiments and
520 interpreting the results. While Projection-TAG AAVs can be multiplexed and used interchangeably
521 (Extended Data Fig. 3d-h, Extended Data Fig. 4h), AAV transduction efficiency and Projection-TAG
522 expression may not be uniform across projection pathways (Extended Data Fig. 4f,g), which may
523 introduce confounds to quantitative measurement of the projection strength.

524 After the tracing experiments, Projection-TAG detection and multi-modal profiling can be achieved
525 using various commercial assays. Projection-TAG detection is specific and sensitive in single-cell
526 sequencing and multiplexed FISH, as previously described (Fig. 1). Multiplexed FISH is superior at the
527 Projection-TAG detection efficiency but is labor-intensive and difficult to scale up. Sun1-GFP expressed
528 by Projection-TAG AAVs reliably labels both whole cells and nuclear suspensions, enabling enrichment

529 of projection neurons for unbiased single-cell and single-nucleus transcriptional and epigenetic profiling
530 in a high-throughput manner. Enzyme-free nuclear extraction works for various tissue samples (fresh,
531 frozen, or fixed) and introduces minimal dissociation-induced transcriptional stress response, and thus
532 ideal for activity-dependent circuit mapping. Whole-cell sequencing has the ability to sequence both
533 cytoplasmic and nuclear transcripts, advantageous for recovering medium- to low-expressing transcripts
534 and may potentially increase Projection-TAG detection. Sequencing involves many biochemical
535 processes that may confound the Projection-TAG detection. Artifacts such as ambient RNA
536 contamination, doublets, and analytical errors may lead to false positives, which can be addressed
537 experimentally (e.g. reduce ambient RNA contamination by adding additional wash steps and
538 incorporating FACS) and computationally (e.g. ambient RNA correction, doublet removal, and setting up
539 appropriate FDR for the analysis). Projection-TAG detection rate is limited by the sequencing assays (e.g.
540 efficiency of reverse transcription and cDNA capture) and sequencing setups (Extended Data Fig. 4),
541 which may pose an upper limit for Projection-TAG detection and lead to false negatives. Consequently,
542 lack of Projection-TAG expression in a cell should not be simply interpreted as lack of projection, and the
543 neurons projecting to each target may be under-reported in snRNA-seq, which may result in a higher
544 degree of underestimation of the axonal collaterals. Projection-TAG detection using sequencing assays
545 with higher transcript capture efficiency and multiplexed FISH (or multi-color CTB tracing) may circumvent
546 false negatives.

547 **Future directions of Projection-TAGs**

548 Projection-TAGs could potentially be compatible with additional commercial platforms, such as
549 high-throughput spatial transcriptomics platforms like Xenium and Visium, that would allow spatially-
550 resolved investigation of diverse cell types and distinct anatomical organization of the projection neurons.
551 Another potential application of Projection-TAGs includes connecting neuronal activity to the distinct
552 projections by integrating the oScarlet version with in vitro and in vivo calcium imaging experiments. The
553 cell type and projection information from imaging-based molecular assays can be integrated with the real-
554 time neuronal activity information from the calcium imaging experiments via post-hoc imaging alignment
555 and registration[93,104-106].

556 The number of projections that can be labeled with Projection-TAG AAVs is not inherently
557 constrained, and Projection-TAGs could be easily scalable as we have screened 50 BCs that could
558 potentially be packaged to increase multiplexing of projection tagging. It has been well appreciated that
559 AAV serotypes have distinct tropism and selective labeling of the brain nuclei. With the discovery of
560 capsid selection using directed evolution[107-109] and advances in sequencing techniques and
561 computational tools to optimize exogenous transcript detection[110,111], we anticipate that the improved
562 retrograde labeling efficiency of viruses will enhance wider applications of Projection-TAGs. Given its
563 flexibility, usability, and compatibility with commercial platforms, we envision that Projection-TAGs can be
564 readily applied to study diverse projection types in the central and peripheral nervous system.

565

566 # Methods

567 **Generation of Projection-TAGs**

568 We used 100-bp BCs because they allow us to design HCR probes to detect the spatial
569 distribution of BC transcripts, and would improve their detection rate in RNA-seq compared to shorter
570 BCs that are commonly used. We first retrieved 60 previously reported BCs[55]. We then filtered out the
571 BCs that contain the sequence of restriction enzymes (BamHI, HpaI, and NotI). Next, we performed

572 sequence alignment using blastn suite searching in the "Nucleotide Collection (nr/nt)" database and
573 "refseq_representative_genomes" database against the genomes of human (taxid:9606), mouse
574 (taxid:10090), rat (taxid:10116), and Primates (taxid:9443) and filtered out the ones that showed
575 significant similarities[112]. We then calculated the Euclidean distance between any of the two BCs using
576 DistanceMatrix() function in R package DECIPHER. We then cataloged and reported the first 50 BCs
577 sorted based on the average Euclidean distances with other BCs.

578 **Cloning and viral packaging of Projection-TAG AAVs**

579 To generate the backbone of pAAV-CAG-Sun1-GFP-WPRE-pA, we first linearized two
580 previously reported plasmids: pAAV-CAG-H2B-GFP (Addgene, Plasmid #116869)[26] using HindIII and
581 pAAV-Ef1a-DIO-Sun1GFP-WPRE-pA (Addgene, Plasmid #160141)[113] using AscI, and end filled. We
582 then digested them with SpeI and NheI, respectively, followed by gel extraction of the fragments at
583 around 4,700 bp (pAAV-CAG backbone) and 2,400 bp (Sun1-GFP), respectively. We next generated
584 the plasmid pAAV-CAG-Sun1-GFP-WPRE-pA by ligating the two fragments together. To insert BCs into
585 pAAV-CAG-Sun1-GFP-WPRE-pA, we first generated a gene block fragment for each BC with the
586 following structure (5' overhang-NotI-BC-SV40 polyA-HpaI-3' overhang) and sequence
587 (AAGGAAAAAAgcggccgc-100bp sequence of BC-
588 ttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt
589 gctttatttgtaaccattataagctgcaataaacaagttaacAACCGCTGCCG). We then digested both pAAV-CAG-
590 Sun1-GFP-WPRE-pA and BC gene blocks with NotI and HpaI, and generated pAAV-CAG-Sun1-GFP-
591 WPRE-BC-pA by ligating the fragments. We generated 12 plasmids each expressing a unique BC (1 -
592 12). We packaged the plasmids into AAV (BrainVTA, China) and generated a set of 12 rAAV2-retro
593 samples each expressing a unique BC (titer range 1.1 - 3.5e+12 vg/ml). To generate a set of 12
594 plasmids pAAV-CAG-oScarlet-WPRE-BC-pA each expressing a unique BC, we generated BamHI-
595 oScarlet-WPRE-NotI gene fragment and replaced Sun1-GFP-WPRE fragment in pAAV-CAG-Sun1-
596 GFP-WPRE-BC-pA using BamHI and NotI. All restriction enzymes were purchased from New England
597 Biolabs, and all plasmids produced in this study are deposited to Addgene:
598 https://www.addgene.org/browse/article/28247961/.

599 **Testing Projection-TAGs in HEK cells**

600 We acquired human embryo kidney cell line HEK 293T/17 (ATCC) and cultured it using DMEM
601 (Corning), supplemented with 10% Fetal Bovine Serum (Gibco), and 1% Penicillin/Streptomycin
602 (Sigma), on either 60 mm dishes for RNA-seq or 8-well chambered slides (Nunc™ Lab-Tek™) pre-
603 coated with PDL-collagen for multiplexed FISH. Once the cells reached 70 – 80% confluency, we
604 transfected the DNA of one BC plasmid into each of the HEK samples using lipofectamine 2000
605 (Invitrogen) according to the manual. Cells were incubated for 24 – 36 hours before analysis. For RNA-
606 seq, around 1 million cells were used for each sample and RNA was extracted and purified using
607 RNAqueous™ Total RNA Isolation Kit (Invitrogen) following the manual. Library preparation and
608 sequencing were conducted by the McDonnel Institute of Genomics at Washington University School of
609 Medicine. For FISH, cells were washed with DPBS and fixed using 4% formaldehyde for 10 minutes at
610 room temperature. FISH was performed following the manufacturer's manual. See the following section
611 "Multiplexed fluorescent *in situ* hybridization with HCR" for experimental details.

612 **Animals**

613 All experiments were conducted in accordance with the National Institute of Health guidelines
614 and with approval from the Animal Care and Use Committee of Washington University School of
615 Medicine. Mice were housed on a 12-hour light-dark cycle (6:00 am to 6:00 pm) and were allowed free

616 access to food and water. All animals were bred onto C57BL/6J background, and no more than five
617 animals were housed per cage. Female littermates between 8 and 10 weeks old were used for
618 experiments.

619 **Stereotaxic surgeries**

620 Mice were given a single subcutaneous injection of 0.5ml 0.9% sterile sodium chloride and
621 Buprenorphine-SR, 1 hour prior to surgery to help rehydrate the mouse after anesthesia. Mice were
622 anesthetized with 1.5 - 2% isoflurane in an induction chamber using isoflurane/breathing air mix. Once
623 deeply anesthetized, mice were secured in a stereotactic frame (RWD Life Science) where surgical
624 anesthesia was maintained using 2% isoflurane. Mice were kept on a heating pad for the duration of
625 the procedure. Preoperative care included application of sterile eye ointment for lubrication,
626 administration of 1 mL of subcutaneous saline, and surgery-site sterilization with iodine solution. We
627 injected into the seven projection targets of the cortex, each with a Projection-TAG (rAAV2-retro
628 expressing a unique TAG), in single animals (Supplementary Table 2). All injections were made using a
629 Nanoject II auto injector (Drummond) with a glass microelectrode at a rate of 1 nl/s and the needle was
630 held in place for 10 minutes prior to needle withdrawal. Needles are changed between each Projection-
631 TAG injection to avoid contamination between Projection TAGs. Stereotaxic surgery injecting into SC_L
632 and SC_S was performed at week 0. For SC_L injection, the injections were located 400 μm lateral to the
633 center of the posterior artery (150–300 mm below the dura), and the virus was bilaterally injected
634 between T12 and T13 intravertebrally with 250 μl volume each injection. For SC_S injection, the dorsal
635 part of the L1 vertebra was gently excised using a high-speed micro-drill (RWD Life Science) to unveil
636 the L6-S1 spinal cord segments. The injections were located 100 μm left-lateral to the center of the
637 posterior artery with a 10-degree left-right tilt (550–580mm below the dura). Three distinct sites were
638 injected with 150 nl of virus at each site. Following a 2-week convalescent interval, the second
639 stereotaxic surgery was conducted by injecting into cMOp, cSSp, VP, PAG, and MY using the following
640 coordinates relative to Bregma: cMOp (anterior-posterior [AP] -0.52 mm, medial-lateral [ML] +1.81 mm,
641 dorsal-ventral [DV] -0.86 mm, left hemisphere), cSSp (AP -0.52 mm, ML +0.75 mm, DV -0.86 mm, left
642 hemisphere), VP (AP -1.5 mm, ML -1.52 mm, DV -4 mm, right hemisphere), PAG (AP -4.5 mm, ML -0.5
643 mm, DV -2.83 mm, right hemisphere), and MY (AP -6.2 mm, ML -0.5 mm, DV -5.93 mm, left
644 hemisphere). A small midline dorsal incision was performed to expose the skull. After leveling the head
645 relative to the stereotaxic frame, the specified injection coordinates were used to mark the locations on
646 the skull, and a small hole (approximately 0.5 mm diameter) was drilled for each. 500 nl of virus was
647 injected into each of the targets. After each surgery, the surgery site was sutured, and mice were
648 recovered from anesthesia on a heating pad and then returned to their home cage. Mice were given
649 carprofen (0.05 mg/ml in water) to minimize inflammation and discomfort and monitored for three
650 consecutive days.

651 To label neurons co-projecting to cSSp and VP, 500 nl of AAV2retro-hSyn-eGFP-Cre (Addgene
652 #105540) was injected into VP and 500 nl of AAV2retro-Ef1a-DIO-H2B-tdTomato (BrainVTA) into cSSp
653 of the same mice (n=2) using the stereotaxic coordinates described above. Retrograde labeling in the
654 SSp and MOp was examined two weeks after injection.

655 **Mouse model of visceral pain**

656 Stereotaxic surgeries were performed as described above. At week 5 post-viral injections, mice
657 were acclimated for at least 3 days prior to the experiment. Mice were administered an intraperitoneal
658 injection of either cyclophosphamide (Sigma-Aldrich, dissolved and diluted to 40 mg/mL in 0.9% sodium

659 chloride) with a dose of 200 mg/kg or saline (approximately the same volume as CYP). Animals were
660 monitored in their home cage for 30 minutes before perfusion.

661 **Tissue processing**

662 In the experiments reported in this paper, we traced the seven projection targets of the MOp and
663 SSp by performing two stereotaxic surgeries using undiluted AAVs in 8-10 weeks-old C57BL/6 female
664 mice. Individual Projection-TAG AAVs were injected into the SC_L and SC_S at week zero, followed by a
665 second surgery of AAV injections into the MOp, SSp, VP, PAG, and MY at week two. Animals were
666 perfused, cortical samples were collected for either FISH or sequencing analysis at week five.

667 To collect biological samples for FISH studies, animals were anesthetized with ketamine
668 cocktail, and perfused with DEPC-PBS followed by 4% DEPC-PFA. Brain and spinal cord tissues were
669 dissected and incubated in 4% DEPC-PFA at 4 C for 6 - 8 hours. Tissues were then incubated in 30%
670 sucrose in 1X DEPC-PBS at 4 C for 24 - 48 hours until they sank to the bottom of the tube. Tissues
671 were then embedded in OCT and stored at -80℃ before slicing. Tissues were sliced coronally into
672 sections with 30 μm thickness using a cryostat (Leica), and tissue slices were mounted on microscope
673 slides. Slides were stored at -20℃ for a least one hour before moving to -80℃ for long-term storage.
674 Slides containing the brain regions of injections were examined under a fluorescent microscope, and
675 the fluorescent signal from Sun1-GFP was used to confirm the injection sites. As an optional step, the
676 native fluorescence of Sun1-GFP can be photobleached from mouse brain sections by floating slices in
677 1×PBS with 24 mM NaOH and 4.5% H_2O_2 and exposing under the UV light (27 total watts, OPPSK) for
678 30 minutes at room temperature. Slices were immediately rinsed with DEPC-PBS twice before
679 mounting on a microscope slide.

680 To collect biological samples for molecular experiments such as qPCR or sequencing, animals
681 were anesthetized with ketamine cocktail, and perfused with ice-cold NMDG-based cutting solution
682 (NMDG 93 mM, KCl 2.5 mM, NaH_2PO_4 1.25 mM, $NaHCO_3$ 30 mM, HEPES 20 mM, Glucose 25 mM,
683 Ascorbic acid 5 mM, Tiourea 2 mM, Sodium Pyruvate 3 mM, $MgSO_4$ 10 mM, $CaCl_2$ 0.5 mM, N-
684 acetylcysteine 12 mM; pH adjusted to 7.3 with 12N HCl, and bubbled with 95% O_2 and 5% CO_2). The
685 spinal cord was dissected, fixed, and sectioned by following the sample preparation process mentioned
686 above in order to confirm the injection sites of SC_L and SC_S. The brain was submerged in ice-cold
687 NMDG-based cutting solution and sliced coronally into sections with 400 μm thickness using a
688 Compresstome (Precisionary, VF-210-0Z). MOp and SSp samples were prepared by micro-dissecting
689 the respective regions under a microscope and collected into a nucleus-free centrifuge tube placed on
690 dry ice. Samples were stored at -80℃ for long-term storage.

691 To assess the viral spread in the injection sites, the remaining brain and spinal cord sections
692 containing the injection sites were placed on glass slides. The fluorescent signal from Sun1-GFP was
693 examined under a fluorescent microscope. We first examined and confirmed the GFP expression in
694 each of the injection sites for every mouse included in this study. To assess off-target labelling, we next
695 examined if there was strong GFP signal displayed in the neighbor regions of the injection site that are
696 away from the intended seed region, and if there is leaked labelling in the needle track. We quantified
697 the off-target effect for each injection site by calculating the area of GFP signal displayed in the
698 neighbor brain regions divided by the area of the intended target brain region (Extended Data Fig. 5).
699 On average, the off-target effect was measured at 39.6 ± 16.4% (cMOp), 8.4 ± 5.6% (cSSp), 16.2 ±
700 8.5% (VP), 14.7 ± 10.6% (PAG), and 8.8 ± 6% (MY) across all mice used in this study. GFP signals in
701 the spinal cord are restricted to the target spinal cord segment for all spinal cord injections across all
702 mice used in this study.

703 **Multiplexed FISH with HCR**

704 HCR (hybridization chain reaction) v3.0 probes, amplifiers, and reagents were purchased from
705 Molecular Instruments[114]. Multiplexed FISH was performed according to the manual of HCR RNA-FISH
706 with minor modifications. For hybridization, samples were equilibrated in hybridization buffer for 30 min
707 at 37℃ and hybridized with probe sets (16 nM each probe) in hybridization buffer overnight at 37℃.
708 Samples were washed in probe wash buffer and gradually switched to 5×SSCT (5×SSC, 0.1%Tween-
709 20) at 37℃. HCR was carried out at room temperature. Samples were equilibrated in amplification
710 buffer for 30 min and the amplifier hairpins (conjugated to Alexa-488, Alexa-546, Alexa-647, and/or
711 Cyanine 7) were heated to 95℃ and snap-cooled in a dark drawer for 30 min. Hairpins were then mixed
712 and diluted to 0.6 nM each hairpin in amplification buffer before incubated with samples overnight.
713 Samples were washed with 5×SSCT for a total of three times and imaged in 2×SSCT with DAPI.
714 Imaging was carried out as described in the next section. After imaging, samples were washed with
715 5×SSCT, then HCR probes and hairpins were stripped by incubating with 0.25 U/ul of DNase I (Sigma
716 Aldrich) in 1×DNase incubation buffer for 90 minutes at 37℃ (Extended Data Fig. 2g, i). Samples were
717 washed with 5×SSCT for 5 minutes for a total of 5 times before the next round of HCR-FISH.

718 **Imaging**

719 After each FISH round, slices were imaged at 10x magnification on a fluorescence microscope
720 (Keyence, BZ-X800). Filter cubes used include DAPI (Chroma 49021, Excitation filter at 405 nm with 20
721 nm bandwidth, Emission filter at 460 nm with 50 nm bandwidth), GFP/AF488 (Chroma 49011, 480/40x,
722 535/50m), Alexa-546 (Chroma 49304, 546/10x, 572/23m), Alexa-647 (Chroma 49006, 620/60x,
723 700/75m), and Cyanine 7 (Chroma 49007, 710/75x, 810/80m). The DAPI signal and signals from FISH
724 probes were imaged for each FISH round. Images from each FISH round were stitched using the
725 Keyence BZ-X800 Analyzer. Images from multiple FISH rounds were loaded into FIJI/ImageJ and
726 aligned using the HyperStackReg plugin by choosing DAPI channels for transformation matrix
727 computation. Images were then downsampled by a factor of 2 and background was subtracted.

728 **Quantification of projection neurons using QUINT**

729 FISH images were segmented and quantified using the QUINT workflow which allows for semi-
730 automated quantification of cells in labeled brain regions[115]. For atlas registration, images from different
731 channels were merged and downsampled according to the manual. An XML file was generated using
732 Filebuilder and loaded into QuickNII for linear registration to the Allen Mouse Brain Atlas CCFv3[116,117].
733 After, user-guided nonlinear refinements were applied to brain sections using VisuAlign. For the
734 segmentation of positive cells, images of individual channels were prepared, and segmentation was
735 performed through the pixel classification and object classification pipelines in ilastik[118]. In pixel
736 classification, models were trained to distinguish signal from background using a small subset of
737 images and applied to the whole dataset. Individual machine-learning algorithms were trained for each
738 channel. Probability maps of signal and background were exported as HDF5 files to be used in object
739 classification. In object classification, models were trained to distinguish objects from non-specific
740 labeling based on features such as size and shape using a small subset of images applied to the whole
741 dataset.

742 The performance of illastik segmentation algorithms was validated against manual
743 segmentation. Illastik accurately identified 96.8% of manually segmented cells with a 6.6% false
744 positive rate (Extended Data Fig. 6b). There are 2.1% of the segmented objects identified as multiple
745 cells merged into one object and 1.9% of the segmented objects identified as one cell split into multiple
746 objects. Finally, segmentation and registration files were uploaded to the Quantifier feature in Nutil

747 which resulted in quantification of cells per brain region according to the reference atlas[119]. To identify
748 cells that co-labeled by multiple markers, objects positive for each marker were segmented individually
749 and the x and y coordinates of the center of each object were calculated using illastik. If the Euclidian
750 distance of two objects from separate channels is less than the average radius of all objects, the two
751 objects were identified as one cell that co-expresses both marker genes. For Projection-TAG detection
752 and specificity analysis, cSSp and cMOp were excluded from this analysis because they are known to
753 contain neurons projecting to other injection sites and may be labeled by TAGs injected into other
754 regions which could contaminate the final distribution of the TAG.

755 **Single-nuclei isolation and FACS**

756 Nuclear extraction was performed according to the protocol described previously with minor
757 modifications[120]. Mouse cortical tissues were transferred to a dounce homogenizer in homogenization
758 buffer (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl, pH 8.0, 5 ug/mL actinomycin, 1%
759 BSA, and 0.08 U/ul RNase inhibitor, 0.01% NP40) on ice. Samples were homogenized for 10 strokes
760 with the loose pestle in a total volume of 1 mL, followed by 10 additional strokes with the tight pestle.
761 The tissue homogenate was then passed through a 50 µm filter and diluted 1:1 with working solution
762 (50% iodixanol, 25 mM KCl, 5 mM MgCl2, and 10 mM Tris-HCl, pH 8.0). Nuclei were layered onto an
763 iodixanol gradient after homogenization and ultracentrifuged as described previously. After
764 ultracentrifugation, nuclei were collected between the 30 and 40% iodixanol layers and diluted with
765 resuspension buffer (1xPBS with 1% BSA, and 0.08 U/ul RNase inhibitor). Nuclei were centrifuged at
766 500 g for 10 min at 4°C and resuspended in resuspension buffer with 5 ng/ul of 7-AAD. For FACS,
767 gates on GFP and 7AAD were set using tissues collected from mice with no stereotaxic surgeries and
768 no 7AAD staining. GFP+/7-AAD+ events and GFP-/7-AAD+ events were sorted using 100 µm nozzle
769 on a BD FACSARIA II and collected separately into 1.5 mL microcentrifuge tubes containing 100 µl of
770 resuspension buffer.

771 **snRNA-seq and snATAC-seq**

772 For snRNA-seq, nuclei were further processed and sequenced according to the manufacturer's
773 manuals of Parse-biosciences Evercode WT mini Kit V2 (Parse) or 10X Genomics Chromium Single
774 Cell Gene Expression 3' V3.1 Assay (10X RNA). For combinatorial snRNA-seq and snATAC-seq, nuclei
775 were processed and sequenced according to the manufacturer's manual of 10X Genomics Chromium
776 Single Cell Multiome Assay (10X Multiome). Libraries were sequenced on a NovaSeq6000 with 150
777 cycles each for Read1 and Read2, targeting 100,000 paired reads/nucleus for snRNA-seq libraries and
778 50,000 paired reads/nucleus for snATAC-seq libraries. Raw sequencing data from individual libraries
779 were processed and mapped using 10X Genomics cellranger-6.1.2 (10X RNA), 10X Genomics
780 cellranger-arc-2.0.1 (10X Multiome), or Parse Computational Pipeline v1.1.2 (Parse). The reference
781 genome was generated by adding the Sun1-GFP and Projection-TAGs 1-12 [Supplementary Table 1])
782 to GRCm38. Projection-TAG feature was extended from the 100-nt Projection-TAG sequence by 140-nt
783 on each end, to ensure sequencing reads have at least 10-nt alignment to the Projection-TAGs . For
784 snATAC-seq, accessible peaks were identified using cellranger-arc aggr by analyzing the combined
785 fragment signals across all snATAC-seq libraries in the dataset.

786 **Target Amplification**

787 Parallel PCR reaction (KAPA HiFi HotStart ReadyMix, Roche) was carried out to amplify
788 Projection-TAG sequences from the cDNA libraries generated using 10X Genomics Chromium Single
789 Cell Gene Expression 3' V3.1 or Multiome Assay, First, forward primer (5'-
790 ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse primer (5'-

791 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCCGCATCGATACCG'-3') was used
792 separately in PCR reactions with 20 ng of cDNA library each reaction. PCR program: (1) 95 °C for 3
793 min, (2) 98° for 20 s, 70 °C for 15 s, 72 °C for 20 s (10 cycles), (3) 72 for 1 min. The reactions
794 containing forward and reverse primers were pooled and a second PCR reaction was done following
795 the PCR program above. As a result, the region containing the Projection-TAG sequence, 10X cell
796 barcodes, and UMI sequences will be preferentially amplified from the cDNA molecules. The target
797 amplification amplicons were labeled with sample indices (Dual Index Plate TT Set A, 10X Genomics)
798 and using PCR reaction with program (1) 95 °C for 3 min, (2) 98° for 20 s, 65 °C for 15 s, 72 °C for 20 s
799 (12 cycles), (3) 72 for 1 min. PCR products were purified using SPRIselect beads sequenced on a
800 NovaSeq6000 with 150 cycles each for Read1 and Read2, targeting 10M reads per library. Raw
801 sequencing data from individual libraries were processed and mapped to individual Projection-TAGs
802 using 10X Genomics cellranger-6.1.2.

803 **Quality control, clustering, and annotation of snRNA-seq**

804 The gene-cell count matrices from all snRNA-seq libraries were concatenated using R (V4.1.1)
805 package Seurat (V4.4.0)[121]. To be included in the snRNA-seq analysis, nuclei were required to contain
806 at least 500 unique genes, less than 15,000 total UMIs, and fewer than 5% of the counts deriving from
807 mitochondrial genes. There were 69,657 nuclei that met these criteria. Raw counts were scaled to
808 10,000 transcripts per nucleus and log-transformed using NormalizeData() function to control the
809 sequencing depth between nuclei. Counts were centered and scaled for each gene using ScaleData()
810 function. Highly variable genes were identified using FindVariableFeatures() and the top 20 principal
811 components were retrieved with RunPCA() using default parameters. For dimension reduction and
812 visualization, Uniform Manifold Approximation and Projection (UMAP) coordinates were calculated
813 using RunUMAP(). Nuclei clustering was performed using FindClusters() based on the variable features
814 from the top 20 principal components, with the resolution set at 0.6, and the marker genes for each
815 cluster was identified using FindAllMarkers() comparing nuclei in one cluster to all other nuclei. Doublet
816 or low-quality nuclei were identified if they meet any of the following criteria: 1). Assigned to a cluster
817 with no significantly enriched marker genes (FDR < 0.05, log2FC > 1); 2). Assigned to a cluster in which
818 five or more mitochondrial genes were identified among top 20 marker genes (sorted by avg_log2FC);
819 3). Identified as doublets using R package DoubletFinder with doublet expectation rate at 5%[122]. In total
820 5,991 nuclei were identified as doublet or low-quality and thus excluded from the dataset. The
821 remaining 63,666 nuclei were clustered as described above.

822 Transcriptional classes and cell types were assigned to each cluster based on the canonical
823 marker genes previously reported (Fig. 2e). Specifically, for classes, glutamatergic and GABAergic
824 neuronal clusters are annotated based on the expression of *Slc17a7* and *Slc32a1*, respectively. Non-
825 neuronal clusters are annotated based on the lack of expression of *Slc17a7* and *Slc32a1*. For cell
826 types, IT clusters are annotated based on the expression of *Slc30a3*, and PT, NP, CT clusters are
827 labeled by *Lratd2*, *Lypd1*, and *Syt6*, respectively. L6b is marked by *Nxph4*. Vascular cells express *Crh*
828 and/or *Uaca*. Microglial cells and astrocytes are marked by *C1qa* and *Emx2*, respectively.
829 Oligodendrocytes and OPC are labeled based on the expression of *Mbp* and *Sox10*. Subtypes were
830 assigned to each cluster based on the marker gene uniquely expressed in that specific cluster
831 (Extended Data Fig. 7,e,f).

832 **Projection feature annotation**

833 To study the projection feature of single neurons, we assigned the projection targets to
834 individual snRNA-seq nuclei based on the expression of Projection-TAGs. In snRNA-seq, a Projection-

TAG-cell matrix was generated from the gene-cell matrix. A projection target is assigned to a single nucleus if that nucleus expresses the Projection-TAG (> 0 UMIs) injected into the target, and each nucleus is evaluated for each projection target based on the expression of the corresponding Projection-TAG. In target amplification, we set an expression cutoff for each Projection-TAG in each library. The cutoff is 1 UMI or X percentile of the corresponding Projection-TAG expression in the corresponding library (where the X equals the percentage of nuclei not expressing the corresponding Projection-TAG in the corresponding snRNA-seq library plus one), whichever comes greater. A projection target is assigned to a single nucleus if that the Projection-TAG expression is greater than the cutoff, and each nucleus is evaluated for each projection target based on the expression of the corresponding Projection-TAG. Finally, to combine results from snRNA-seq and target amplification, a projection target is assigned to a single nucleus if that target is assigned in either snRNA-seq or target amplification. Of note, a nucleus can be assigned multiple projection targets if it expressed multiple Projection-TAGs. In the analysis where individual projection targets are analyzed separately (e.g. Fig. 2), nuclei were grouped based on whether they express the corresponding Projection-TAG. For example, snRNA-seq nuclei only expressing cMOp-TAG and co-expressing cMOp-TAG and other TAGs will be grouped for cMOp analysis. In the analysis where axonal collaterals were reported (e.g. Fig. 3), nuclei were grouped based on the expression of each of the Projection-TAGs. For example, snRNA-seq nuclei only expressing cMOp-TAG and co-expressing cMOp-TAG and cSSp-TAG will be grouped separately for analysis. Projection groups with sample size no less than 60 nuclei (corresponding to 0.34% of the total snRNA-seq nuclei generated from FACS-sorted libraries, the upper limit of FDR described in Fig. 1) were reported in all analyses.

Anchoring analysis of snRNA-seq

To validate the transcriptional cluster annotation, we directly compare our snRNA-seq data to a published scRNA-seq of mouse MOp and SSp previously described[123]. We used Seurat to anchor our snRNA-seq data to the published dataset. First, FindTransferAnchors() was used to identify anchors (conserved features) between datasets. TransferData() was run to transfer cell type labels described in the published dataset to each nucleus in the our snRNA-seq data.

Quality control and clustering of snATAC-seq

The peak-cell counts matrix was loaded and analyzed using R package Signac (V1.7.0)[124]. To be included for snATAC-seq analysis, nuclei were required to be present in snRNA-seq data and contain at least 1,000 fragments overlapping with peaks. There were 41,553 nuclei that met these criteria. Term frequency-inverse document frequency normalization was perfomed using RunTFIDF () function and variable peaks were identified using FindTopFeatures (). Dimension reduction was performed with singular value decomposition using RunSVD() function and Uniform Manifold Approximation and Projection (UMAP) coordinates were calculated using RunUMAP(). Nuclei clustering was performed using FindClusters() based on the top 20 dimensions with the resolution of 0.8.

Differential analyses of gene expression and chromatin accessibility

To identify marker genes and peaks that are enriched in distinct transcriptional subtypes or projections, differential expression/accessibility analysis was performed using findAllMarkers() in Seurat/Signac, comparing nuclei from one subtype or projection to all other nuclei. Genes and peaks with FDR < 0.05 were reported. To identify genes and peaks that are differentially expressed/accessible between any two subtypes or projections, differential expression/accessibility analysis was performed using findMarkers() in Seurat/Signac, comparing nuclei from one subtype or projection to nuclei from another subtype or projection. Genes and peaks with FDR < 0.05 were reported.

To identify genes that are differentially expressed in neurons projecting to only one target compared to neurons projecting to multiple targets while controlling for transcriptional cell types, differential expression analysis was performed using FindMarkers() in Seurat, comparing nuclei of the same transcriptional cell type expressing only one Projection-TAG to nuclei co-expressing the same Projection-TAG and other TAGs. Cell types that have at least 60 nuclei for each projection group were analyzed, and genes and peaks with FDR < 0.05 were reported.

To identify genes and peaks that are differentially regulated by CYP in individual subtypes or projections, differential expression/accessibility analysis was performed using findMarkers() in Seurat/Signac, comparing nuclei from CYP-treated animals to nuclei from Saline-treated animals of the same subtype or projection. Genes and peaks with avg_log2FC > 0.5 and FDR < 0.05 were reported. To identify IEGs that are activated by CYP in neurons projecting to IT targets, cMOp and cSSp, differential expression analysis was performed using findMarkers() in Seurat, comparing activated nuclei, defined by Act-seq, that are positive for the corresponding Projection-TAG from CYP-treated animals to the same number of randomly sampled nuclei (with matched subtype distribution) positive for the same Projection-TAG from Saline-treated animals. IEGs with avg_log2FC > 0.5 and FDR < 0.05 were reported.

Identification of putative genomic regulatory elements

To identify snATAC-seq peaks that are correlated with gene expression in cis and may act as genomic regulatory elements (GREs) to regulate the expression of target genes, we modified a computational pipeline previously described[54]. we first calculated, using snATAC-seq data, the average accessibility of each peak and, using snRNA-seq data, the average expression of each gene in individual Subtypes or Nuclei positive for individual TAGs. We then computed the Pearson correlation coefficient r between the accessibility of peaks and the expression of genes across Nuclei positive for individual TAGs or Subtypes for any peak-gene pair, in which the center of the peak is located within 5 mbps of the center of the TSS of the gene on the same chromosome. To identify putative GREs that may direct celltype-specific gene expression, we identified pairs of Celltype-specific snATAC-seq peaks (avg_log2FC > 0.5, FDR < 0.05, comparing peak accessibility in nuclei of one Celltype to all others) and Celltype-specific snRNA-seq genes (avg_log2FC > 0.5, FDR < 0.05, comparing gene expression in nuclei of one T- type to all others) in the same T- type. A peak-gene pair with a strong positive correlation (Pearson's r > 0.75) is identified as a putative enhancer (pu.Enhancer) and its putative regulated gene, whereas a pair with a strong negative correlation (Pearson's r < -0.75) is categorized as a putative silencer (pu.Silencer) and its regulated gene. To identify putative GREs that may direct Projection-specific gene expression, we identified pairs of Projection-specific snATAC-seq peaks (avg_log2FC > 0.5, FDR < 0.05, comparing peak accessibility in nuclei of one Projection to all others) and Projection-specific snRNA-seq genes (avg_log2FC > 0.5, FDR < 0.05, comparing gene expression in nuclei of one Projection to all others) in the same Projection. Projection-specific pu.Enhancers and pu.Silencers and their regulated genes are identified as described above.

Act-seq analysis

To identify cell populations that are activated by visceral pain, the IEG score for each nucleus in snRNA-seq data was calculated using AddModuleScore() in Seurat based on the expression of a set of 139 immediate early genes (IEG) previously described[90]. A transcriptional subtype or projection was considered transcriptionally "activated" if the IEG scores of CYP-treated nuclei in this population were significantly higher than the IEG scores of Saline-treated nuclei. A nucleus was considered transcriptionally ''activated'' if its IEG score was 2 standard deviations higher than the average IEG

scores across all Saline-treated nuclei in the same projection group (nuclei expressing the corresponding TAG).

Statistical analysis and visualization

Statistical analyses including the number of animals or cells (n) and *p* values for each experiment are noted in the figure legends. Statistics and visualization were performed using R version 4.0.1. Student's t-tests were performed using R package stats V4.2.2. ANOVA tests were performed using R package rstatix V0.7.2, followed by post-hoc t-tests with Bonferroni correction. Hypergeometric tests were used to test the significance of the overlap between two gene sets or between nuclei expressing two TAGs using by calling phyper() function in R package stats V4.2.2. Plots were generated using R packages ggplot2 V3.4.0, gplots V3.1.3, and UpSetR V1.4.0.

Data and code availability

Raw and processed data of snRNA-seq and snATAC-seq experiments included in this study are deposited to the NCBI Gene Expression (GEO) SRA with accession number GSE277718. Custom pipelines and scripts are available on GitHub: https://github.com/Samineni-Lab/Projection-TAGs.

Acknowledgements

This work was funded by National Institute on Drug Abuse (R01DA056829) and National Institute of Diabetes and Digestive and Kidney Diseases (R01DK128475) awarded to V.K.S. We thank Flow Cytometry & Fluorescence Activated Cell Sorting Core in the Department of Pathology and Immunology at Washington University School of Medicine for help with FACS. We thank the Genome Technology Access Center at McDonnell Genome Institute at Washington University School of Medicine for help with snRNA-seq and snATAC-seq library construction and sequencing. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. We would like to thank Dr. Robert W Gereau IV, Dr. William Renthal, and Dr. Lisa Fang for their helpful discussion of the manuscript and experimental design; we would like to thank all the Samineni and Gereau lab members and WashU Pain Center colleagues for their help with manuscript preparation.

Author contributions

V.K.S. and L.Y. conceptualized the project and designed the experiments; L.Y., V.K.S., and M.R.H. generated the plasmids; F.L. performed the anatomical tracing experiments; H.H., F.L., T.O., and Y.Z. collected the samples for the library preparation and *in situ* validation studies. L.Y. and H.H. performed multiplexed FISH and analyzed the data; L.Y. performed sequencing experiments; L.Y. and X.L. performed the sequencing analysis; V.K.S. and L.Y. wrote the manuscript with inputs and feedback from all authors.

Competing interest

The authors declare that they have no conflict of interest.

References

1 McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3** (2018). https://doi.org:10.21105/joss.00861

2 Golgi, C. Sulla fina anatomia degli organi centrali del sistema nervosa. *Revista sperimentale di Freniatria.* (1885).

3 Cajal, S. R. Estructura del cerebelo. *Gac. Med. Catalana*, 449-457 (1888).

4 Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739-2755 (2022). https://doi.org:10.1016/j.cell.2022.06.031

5 Mukamel, E. A. & Ngai, J. Perspectives on defining cell types in the brain. *Curr Opin Neurobiol* **56**, 61-68 (2019). https://doi.org:10.1016/j.conb.2018.11.007

6 Nelson, S. B., Sugino, K. & Hempel, C. M. The problem of neuronal cell types: a physiological genomics approach. *Trends Neurosci* **29**, 339-345 (2006). https://doi.org:10.1016/j.tins.2006.05.004

7 Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* **9**, 868-882 (2008). https://doi.org:10.1038/nrg2416

8 Scala, F. *et al.* Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* (2020). https://doi.org:10.1038/s41586-020-2907-3

9 Zeisel, A. *et al.* in *Cell* Vol. 174 999-1014.e1022 (2018).

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

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

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

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

14 Gouwens, N. W. *et al.* Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* **183**, 935-953 e919 (2020). https://doi.org:10.1016/j.cell.2020.09.057

15 Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317-332 (2023). https://doi.org:10.1038/s41586-023-06812-z

16 Zhang, M. *et al.* Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343-354 (2023). https://doi.org:10.1038/s41586-023-06808-9

17 Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333-342 (2023). https://doi.org:10.1038/s41586-023-06818-7

18 Liu, H. *et al.* Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* **624**, 366-377 (2023). https://doi.org:10.1038/s41586-023-06805-y

19 Zu, S. *et al.* Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378-389 (2023). https://doi.org:10.1038/s41586-023-06824-9

20 Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552-561 (2023). https://doi.org:10.1038/s41586-023-06569-5

21 Winter, C. C. *et al.* A transcriptomic taxonomy of mouse brain-wide spinal projecting neurons. *Nature* **624**, 403-414 (2023). https://doi.org:10.1038/s41586-023-06817-8

22 Zhou, J. *et al.* Brain-wide correspondence of neuronal epigenomics and distant projections. *Nature* **624**, 355-365 (2023). https://doi.org:10.1038/s41586-023-06823-w

23 Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195-202 (2019). https://doi.org:10.1038/s41586-019-1716-z

24 Munoz-Castaneda, R. *et al.* Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159-166 (2021). https://doi.org:10.1038/s41586-021-03970-w

25 Gao, L. *et al.* Single-neuron projectome of mouse prefrontal cortex. *Nat Neurosci* **25**, 515-529 (2022). https://doi.org:10.1038/s41593-022-01041-5

26 Economo, M. N. *et al.* Distinct descending motor cortex pathways and their roles in movement. *Nature* **563**, 79-84 (2018). https://doi.org:10.1038/s41586-018-0642-9

27 Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137-143 (2021). https://doi.org:10.1038/s41586-021-03705-x

28 Kim, E. J. *et al.* Extraction of Distinct Neuronal Cell Types from within a Genetically Continuous Population. *Neuron* **107**, 274-282 e276 (2020). https://doi.org:10.1016/j.neuron.2020.04.018

29 Kaspar, B. K. *et al.* Targeted retrograde gene delivery for neuronal protection. *Mol Ther* **5**, 50-56 (2002). https://doi.org:10.1006/mthe.2001.0520

30 Tervo, D. G. *et al.* A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron* **92**, 372-382 (2016). https://doi.org:10.1016/j.neuron.2016.09.021

31 Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat Methods* **4**, 47-49 (2007). https://doi.org:10.1038/nmeth999

32 Reardon, T. R. *et al.* Rabies Virus CVS-N2c(DeltaG) Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron* **89**, 711-724 (2016). https://doi.org:10.1016/j.neuron.2016.01.004

33 Soudais, C., Laplace-Builhe, C., Kissa, K. & Kremer, E. J. Preferential transduction of neurons by canine adenovirus vectors and their efficient retrograde transport in vivo. *FASEB J* **15**, 2283-2285 (2001). https://doi.org:10.1096/fj.01-0321fje

34 Stoeckel, K., Schwab, M. & Thoenen, H. Role of gangliosides in the uptake and retrograde axonal transport of cholera and tetanus toxin as compared to nerve growth factor and wheat germ agglutinin. *Brain Res* **132**, 273-285 (1977). https://doi.org:10.1016/0006-8993(77)90421-8

35 Vercelli, A., Repici, M., Garbossa, D. & Grimaldi, A. Recent techniques for tracing pathways in the central nervous system of developing and adult mammals. *Brain Res Bull* **51**, 11-28 (2000). https://doi.org:10.1016/s0361-9230(99)00229-4

36 Ju, G., Han, Z. S. & Fan, L. Z. Fluorogold as a retrograde tracer used in combination with immunohistochemistry. *J Neurosci Methods* **29**, 69-72 (1989). https://doi.org:10.1016/0165-0270(89)90109-x

37 Lanciego, J. L. & Wouterlood, F. G. A half century of experimental neuroanatomical tracing. *J Chem Neuroanat* **42**, 157-183 (2011). https://doi.org:10.1016/j.jchemneu.2011.07.001

38 Schulte, S. J., Fornace, M. E., Hall, J. K. & Pierce, N. A. HCR spectral imaging: 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent samples. *bioRxiv* (2023). https://doi.org:10.1101/2023.08.30.555626

39 Wang, Y. *et al.* Multimodal mapping of cell types and projections in the central nucleus of the amygdala. *Elife* **12** (2023). https://doi.org:10.7554/eLife.84262

40 Bienkowski, M. S. *et al.* Integration of gene expression and brain-wide connectivity reveals the multiscale organization of mouse hippocampal networks. *Nat Neurosci* **21**, 1628-1643 (2018). https://doi.org:10.1038/s41593-018-0241-y

41 Baek, M., Menon, V., Jessell, T. M., Hantman, A. W. & Dasen, J. S. Molecular Logic of Spinocerebellar Tract Neuron Diversity and Connectivity. *Cell Rep* **27**, 2620-2635 e2624 (2019). https://doi.org:10.1016/j.celrep.2019.04.113

42 Phillips, J. W. *et al.* A repeated molecular architecture across thalamic pathways. *Nat Neurosci* **22**, 1925-1935 (2019). https://doi.org:10.1038/s41593-019-0483-3

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

44 Zhang, Z. *et al.* Epigenomic diversity of cortical projection neurons in the mouse brain. *Nature* **598**, 167-173 (2021). https://doi.org:10.1038/s41586-021-03223-w

45 Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86-102 (2021). https://doi.org:10.1038/s41586-021-03950-0

46 Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975-987 (2016). https://doi.org:10.1016/j.neuron.2016.07.036

47 Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772-786 e719 (2019). https://doi.org:10.1016/j.cell.2019.09.023

48 Klingler, E. *et al.* Temporal controls over inter-areal cortical projection neuron fate diversity. *Nature* (2021). https://doi.org:10.1038/s41586-021-04048-3

49 Chen, Y. *et al.* High-throughput sequencing of single neuron projections reveals spatial organization in the olfactory cortex. *Cell* **185**, 4117-4134 e4128 (2022). https://doi.org:10.1016/j.cell.2022.09.038

50 Sun, Y. C. *et al.* Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24**, 873-885 (2021). https://doi.org:10.1038/s41593-021-00842-4

51 Kuramoto, E. Method for labeling and reconstruction of single neurons using Sindbis virus vectors. *J Chem Neuroanat* **100**, 101648 (2019). https://doi.org:10.1016/j.jchemneu.2019.05.002

52 Xu, P. *et al.* High-throughput mapping of single-neuron projection and molecular features by retrograde barcoded labeling. *Elife* **13** (2024). https://doi.org:10.7554/eLife.85419

53 Zhao, Q. *et al.* A multidimensional coding architecture of the vagal interoceptive system. *Nature* (2022). https://doi.org:10.1038/s41586-022-04515-5

54 Yang, L. *et al.* Human and mouse trigeminal ganglia cell atlas implicates multiple cell types in migraine. *Neuron* **110**, 1806-1821 e1808 (2022). https://doi.org:10.1016/j.neuron.2022.03.003

55 Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat Biotechnol* **39**, 865-876 (2021). https://doi.org:10.1038/s41587-021-00837-3

56 Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat Neurosci* **18**, 170-181 (2015). https://doi.org:10.1038/nn.3917

57 Beloozerova, I. N., Sirota, M. G. & Swadlow, H. A. Activity of different classes of neurons of the motor cortex during locomotion. *J Neurosci* **23**, 1087-1097 (2003). https://doi.org:10.1523/JNEUROSCI.23-03-01087.2003

58 Pennartz, C. M. *et al.* Corticostriatal Interactions during Learning, Memory Processing, and Decision Making. *J Neurosci* **29**, 12831-12838 (2009). https://doi.org:10.1523/JNEUROSCI.3177-09.2009

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

60 Yao, Z. *et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103-110 (2021). https://doi.org:10.1038/s41586-021-03500-8

61 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-346 (2016). https://doi.org:10.1038/nn.4216

62 Beine, Z., Wang, Z., Tsoulfas, P. & Blackmore, M. G. Single Nuclei Analyses Reveal Transcriptional Profiles and Marker Genes for Diverse Supraspinal Populations. *J Neurosci* **42**, 8780-8794 (2022). https://doi.org:10.1523/JNEUROSCI.1197-22.2022

63 Bishop, H. I. *et al.* Distinct Cell- and Layer-Specific Expression Patterns and Independent Regulation of Kv2 Channel Subtypes in Cortical Pyramidal Neurons. *J Neurosci* **35**, 14922-14942 (2015). https://doi.org:10.1523/JNEUROSCI.1897-15.2015

64 Hunnicutt, B. J. *et al.* A comprehensive thalamocortical projection map at the mesoscopic level. *Nat Neurosci* **17**, 1276-1285 (2014). https://doi.org:10.1038/nn.3780

65 J. L. MacDonald, R. M. F., E. Azim, S. J. Shnider, B. J. Molyneaux, P. Arlotta, J. D. Macklis. in *Patterning and Cell Type Specification in the Developing CNS and PNS* 475-502 (Elsevier, 2013).

66 Macias, M., Lopez-Virgen, V., Olivares-Moreno, R. & Rojas-Piloni, G. Corticospinal neurons from motor and somatosensory cortices exhibit different temporal activity dynamics during motor learning. *Front Hum Neurosci* **16**, 1043501 (2022). https://doi.org:10.3389/fnhum.2022.1043501

67 Kawaguchi, Y. Pyramidal Cell Subtypes and Their Synaptic Connections in Layer 5 of Rat Frontal Cortex. *Cereb Cortex* **27**, 5755-5771 (2017). https://doi.org:10.1093/cercor/bhx252

68 Buenrostro, J. D. *et al.* in *Nature* Vol. 523 486-490 (Nature Publishing Group, 2015).

69 Winnubst, J. *et al.* Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* **179**, 268-281 e213 (2019). https://doi.org:10.1016/j.cell.2019.07.042

70 Garcia, A. F., Crummy, E. A., Webb, I. G., Nooney, M. N. & Ferguson, S. M. Distinct populations of cortical pyramidal neurons mediate drug reward and aversion. *Nat Commun* **12**, 182 (2021). https://doi.org:10.1038/s41467-020-20526-0

71 Baker, A. *et al.* Specialized Subpopulations of Deep-Layer Pyramidal Neurons in the Neocortex: Bridging Cellular Properties to Functional Consequences. *J Neurosci* **38**, 5441-5455 (2018). https://doi.org:10.1523/JNEUROSCI.0150-18.2018

72 Zemke, N. R. *et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390-402 (2023). https://doi.org:10.1038/s41586-023-06819-6

73 Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-154 (2015). https://doi.org:10.1038/nrm3949

74 Halfon, M. S. Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends Genet* **36**, 149-151 (2020). https://doi.org:10.1016/j.tig.2019.12.005

75 Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**, 1 (2012). https://doi.org:10.1186/1756-8935-5-1

76 Hrvatin, S. *et al.* in *eLife* Vol. 8 (eLife Sciences Publications Ltd, 2019).

77 Mich, J. K. *et al.* Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep* **34**, 108754 (2021). https://doi.org:10.1016/j.celrep.2021.108754

78 Graybuck, L. T. *et al.* Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449-1464 e1413 (2021). https://doi.org:10.1016/j.neuron.2021.03.011

79 Lambert, J. T. *et al.* Parallel functional testing identifies enhancers active in early postnatal mouse brain. *Elife* **10** (2021). https://doi.org:10.7554/eLife.69479

80 Salimando, G. J. *et al.* Human OPRM1 and murine Oprm1 promoter driven viral constructs for genetic access to mu-opioidergic cell types. *Nat Commun* **14**, 5632 (2023). https://doi.org:10.1038/s41467-023-41407-2

81 Lawler, A. J. *et al.* Machine learning sequence prioritization for cell type-specific enhancer design. *Elife* **11** (2022). https://doi.org:10.7554/eLife.69571

82 Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007-1012 (2014). https://doi.org:10.1126/science.1246426

83 Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**, D882-D889 (2020). https://doi.org:10.1093/nar/gkz1062

84 Zhu, X. *et al.* Somatosensory cortex and central amygdala regulate neuropathic pain-mediated peripheral immune response via vagal projections to the spleen. *Nat Neurosci* **27**, 471-483 (2024). https://doi.org:10.1038/s41593-023-01561-8

85 Ziegler, K. *et al.* Primary somatosensory cortex bidirectionally modulates sensory gain and nociceptive behavior in a layer-specific manner. *Nat Commun* **14**, 2999 (2023). https://doi.org:10.1038/s41467-023-38798-7

86 Chen, T. *et al.* Top-down descending facilitation of spinal sensory excitatory transmission from the anterior cingulate cortex. *Nat Commun* **9**, 1886 (2018). https://doi.org:10.1038/s41467-018-04309-2

87 Liu, Y. *et al.* Touch and tactile neuropathic pain sensitivity are set by corticospinal projections. *Nature* **561**, 547-550 (2018). https://doi.org:10.1038/s41586-018-0515-2

88 Gan, Z. *et al.* Layer-specific pain relief pathways originating from primary motor cortex. *Science* **378**, 1336-1343 (2022). https://doi.org:10.1126/science.add4391

89 Taub, D. G. *et al.* The secondary somatosensory cortex gates mechanical and heat sensitivity. *Nat Commun* **15**, 1289 (2024). https://doi.org:10.1038/s41467-024-45729-7

90 Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. in *Neuron* Vol. 96 313-329.e316 (Cell Press, 2017).

91 Hrvatin, S. *et al.* in *Nature Neuroscience* Vol. 21 120-129 (2018).

92 Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat Biotechnol* **39**, 30-34 (2021). https://doi.org:10.1038/s41587-020-0605-1
93 Condylis, C. *et al.* Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022). https://doi.org:10.1126/science.abl5981
94 Sorensen, A. T. *et al.* A robust activity marking system for exploring active neuronal ensembles. *Elife* **5** (2016). https://doi.org:10.7554/eLife.13918
95 Mukherjee, D. *et al.* Salient experiences are represented by unique transcriptional signatures in the mouse brain. *Elife* **7** (2018). https://doi.org:10.7554/eLife.31220
96 Gonzales, B. J., Mukherjee, D., Ashwal-Fluss, R., Loewenstein, Y. & Citri, A. Subregion-specific rules govern the distribution of neuronal immediate-early gene induction. *Proc Natl Acad Sci U S A* **117**, 23304-23310 (2020). https://doi.org:10.1073/pnas.1913658116
97 Bottai, D. *et al.* Synaptic activity-induced conversion of intronic to exonic sequence in Homer 1 immediate early gene expression. *J Neurosci* **22**, 167-175 (2002). https://doi.org:10.1523/JNEUROSCI.22-01-00167.2002
98 French, P. J. *et al.* Subfield-specific immediate early gene expression associated with hippocampal long-term potentiation in vivo. *Eur J Neurosci* **13**, 968-976 (2001). https://doi.org:10.1046/j.0953-816x.2001.01467.x
99 Brakeman, P. R. *et al.* Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 284-288 (1997). https://doi.org:10.1038/386284a0
100 Szumlinski, K. K. *et al.* Behavioral and neurochemical phenotyping of Homer1 mutant mice: possible relevance to schizophrenia. *Genes Brain Behav* **4**, 273-288 (2005). https://doi.org:10.1111/j.1601-183X.2005.00120.x
101 Huang, L. *et al.* BRICseq Bridges Brain-wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals. *Cell* **182**, 177-188 e127 (2020). https://doi.org:10.1016/j.cell.2020.05.029
102 Bhattacherjee, A. *et al.* Spatial transcriptomics reveals the distinct organization of mouse prefrontal cortex and neuronal subtypes regulating chronic pain. *Nat Neurosci* **26**, 1880-1893 (2023). https://doi.org:10.1038/s41593-023-01455-9
103 Pillay, S. & Carette, J. E. Host determinants of adeno-associated viral vector entry. *Curr Opin Virol* **24**, 124-131 (2017). https://doi.org:10.1016/j.coviro.2017.06.003
104 Bugeon, S. *et al.* A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330-338 (2022). https://doi.org:10.1038/s41586-022-04915-7
105 Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370** (2020). https://doi.org:10.1126/science.abb2494
106 von Buchholtz, L. J. *et al.* Decoding Cellular Mechanisms for Mechanosensory Discrimination. *Neuron* **109**, 285-298 e285 (2021). https://doi.org:10.1016/j.neuron.2020.10.028
107 Han, Z. *et al.* AAV11 enables efficient retrograde targeting of projection neurons and enhances astrocyte-directed transduction. *Nat Commun* **14**, 3792 (2023). https://doi.org:10.1038/s41467-023-39554-7
108 Deverman, B. E. *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat Biotechnol* **34**, 204-209 (2016). https://doi.org:10.1038/nbt.3440
109 Lin, K. *et al.* AAV9-Retro mediates efficient transduction with axon terminal absorption and blood-brain barrier transportation. *Mol Brain* **13**, 138 (2020). https://doi.org:10.1186/s13041-020-00679-1
110 Pool, A. H., Poldsam, H., Chen, S., Thomson, M. & Oka, Y. Recovery of missing single-cell RNA-sequencing data with optimized transcriptomic references. *Nat Methods* **20**, 1506-1515 (2023). https://doi.org:10.1038/s41592-023-02003-w
111 You, Y. *et al.* Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. *Genome Biol* **22**, 339 (2021). https://doi.org:10.1186/s13059-021-02552-3
112 Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**, 203-214 (2000). https://doi.org:10.1089/10665270050081478

113 Lawler, A. J. *et al.* Cell Type-Specific Oxidative Stress Genomic Signatures in the Globus Pallidus of Dopamine-Depleted Mice. *J Neurosci* **40**, 9772-9783 (2020). https://doi.org:10.1523/JNEUROSCI.1634-20.2020

114 Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145** (2018). https://doi.org:10.1242/dev.165753

115 Yates, S. C. *et al.* QUINT: Workflow for Quantification and Spatial Analysis of Features in Histological Images From Rodent Brain. *Front Neuroinform* **13**, 75 (2019). https://doi.org:10.3389/fninf.2019.00075

116 Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953 e920 (2020). https://doi.org:10.1016/j.cell.2020.04.007

117 Puchades, M. A., Csucs, G., Ledergerber, D., Leergaard, T. B. & Bjaalie, J. G. Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool. *PLoS One* **14**, e0216796 (2019). https://doi.org:10.1371/journal.pone.0216796

118 Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**, 1226-1232 (2019). https://doi.org:10.1038/s41592-019-0582-9

119 Groeneboom, N. E., Yates, S. C., Puchades, M. A. & Bjaalie, J. G. Nutil: A Pre- and Post-processing Toolbox for Histological Rodent Brain Section Images. *Front Neuroinform* **14**, 37 (2020). https://doi.org:10.3389/fninf.2020.00037

120 Yang, L., Tochitsky, I., Woolf, C. J. & Renthal, W. Isolation of Nuclei from Mouse Dorsal Root Ganglia for Single-nucleus Genomics. *Bio-protocol* **11**, e4102-e4102 (2021).

121 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048

122 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019). https://doi.org:10.1016/j.cels.2019.03.003

123 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226 (2021). https://doi.org:10.1016/j.cell.2021.04.021

124 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021). https://doi.org:10.1038/s41592-021-01282-5

125 Science, A. I. f. B. (2011).

126 Consortium, E. P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020). https://doi.org:10.1038/s41586-020-2493-4

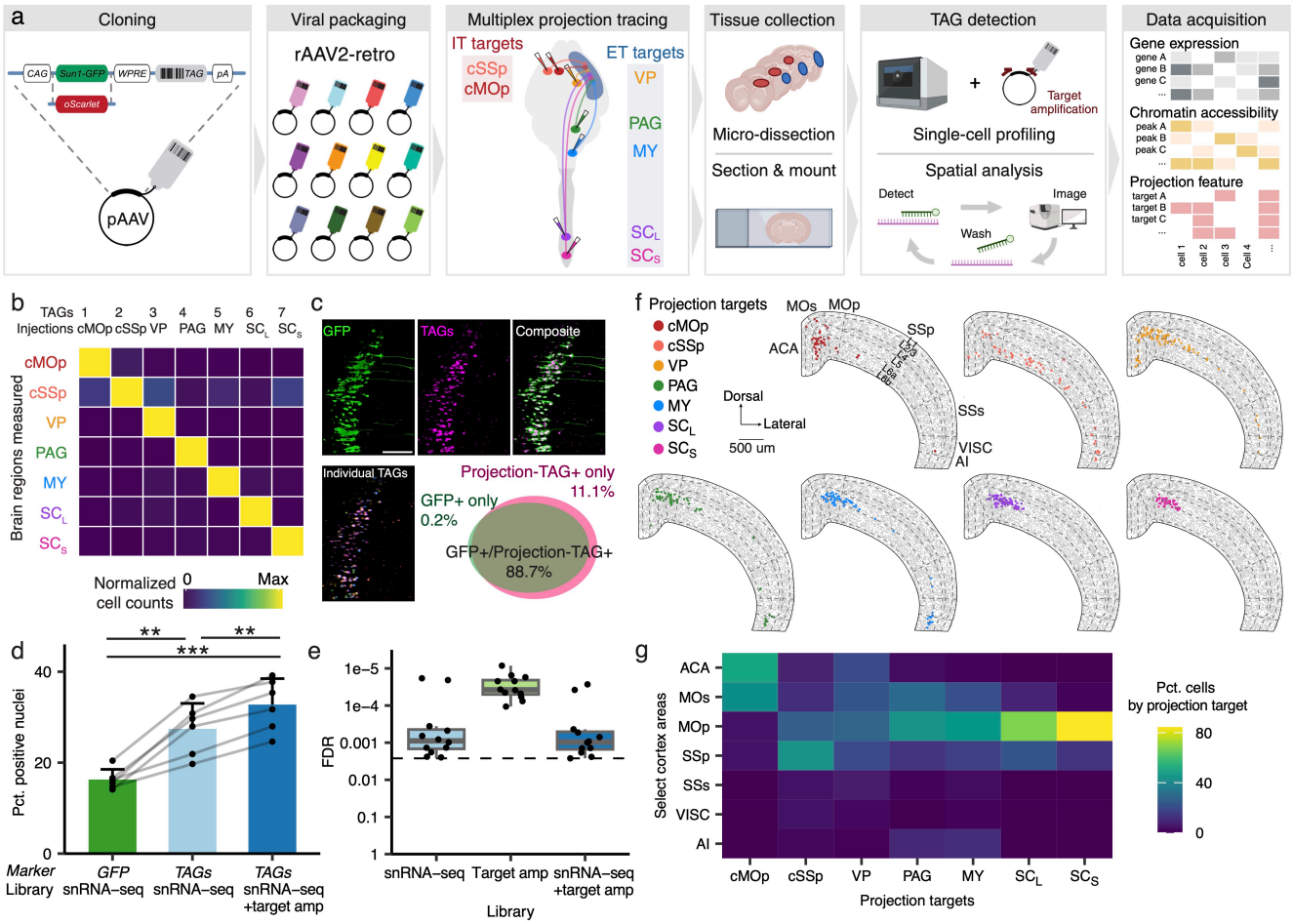


Figure 1. Multiplex projection tracing using Projection-TAGs
a, Experimental and analytical workflow of tagging neuronal projections of the MOp and SSp using Projection-TAGs. **b**, Heatmap showing the enrichment of cells labeled with individual Projection-TAGs by multiplexed FISH in brain regions of injections. Cell counts are normalized by the max values in each brain region. Individual TAGs and their injection sites are shown as on the column. **c**, Representative MOp images showing cells labeled with the fluorescent Sun1-GFP signal and fluorescent FISH signals from Projection-TAGs. Quantification shows the overlap between cells that are GFP+ and/or Projection-TAG+. FISH probes uniquely targeting Projection-TAGs (BCs 1-7) were visualized in separate imaging channels, shown as individual TAGs, and the signals were aggregated pos-hoc, shown as TAGs. Scale bar = 100 µm. **d**, Bar plot showing the percentage of nuclei expressing select marker genes (>0 UMIs for snRNA-seq, > cutoffs for target amplification, see Methods) in individual snRNA-seq libraries prepared by FACS and target amplification. Bar and error bar indicate the average and standard deviation across libraries. (p = 5.3e-06, $F(2,10)$ = 51.77, one-way ANOVA; **p = 0.006 between Projection-TAGs and *GFP* in snRNA-seq, **p = 0.003 for Projection-TAGs between snRNA-seq only and snRNA-seq with target amplification, ***p = 0.001 between *GFP* in snRNA-seq and Projection-TAGs in snRNA-seq with target amplification, post-hoc t-tests with Bonferroni correction). **e**, Box plot showing the false detection rate of Projection-TAGs due to ambient RNA contamination in snRNA-seq libraries. Dotted line shows the highest FDR due to ambient RNA contamination at 0.0026 among all libraries. **f**, Spatial map of Projection-TAG+ cells in select cortex areas from a brain section (-0.65 mm relative to bregma). Cells are downsampled to 200 per Projection-TAG and colored by the injection site of the corresponding TAG. Grey dots are cells segmented based on DAPI signal. **g**, Heatmap showing the distribution of projection neurons in select cortex areas from Fig. 1f. ACA: anterior cingulate area, MOs: secondary motor area, MOp: primary motor area, SSp: primary somatosensory area, SSs: supplemental somatosensory area, VISC: visceral area, AI: agranular insular area.

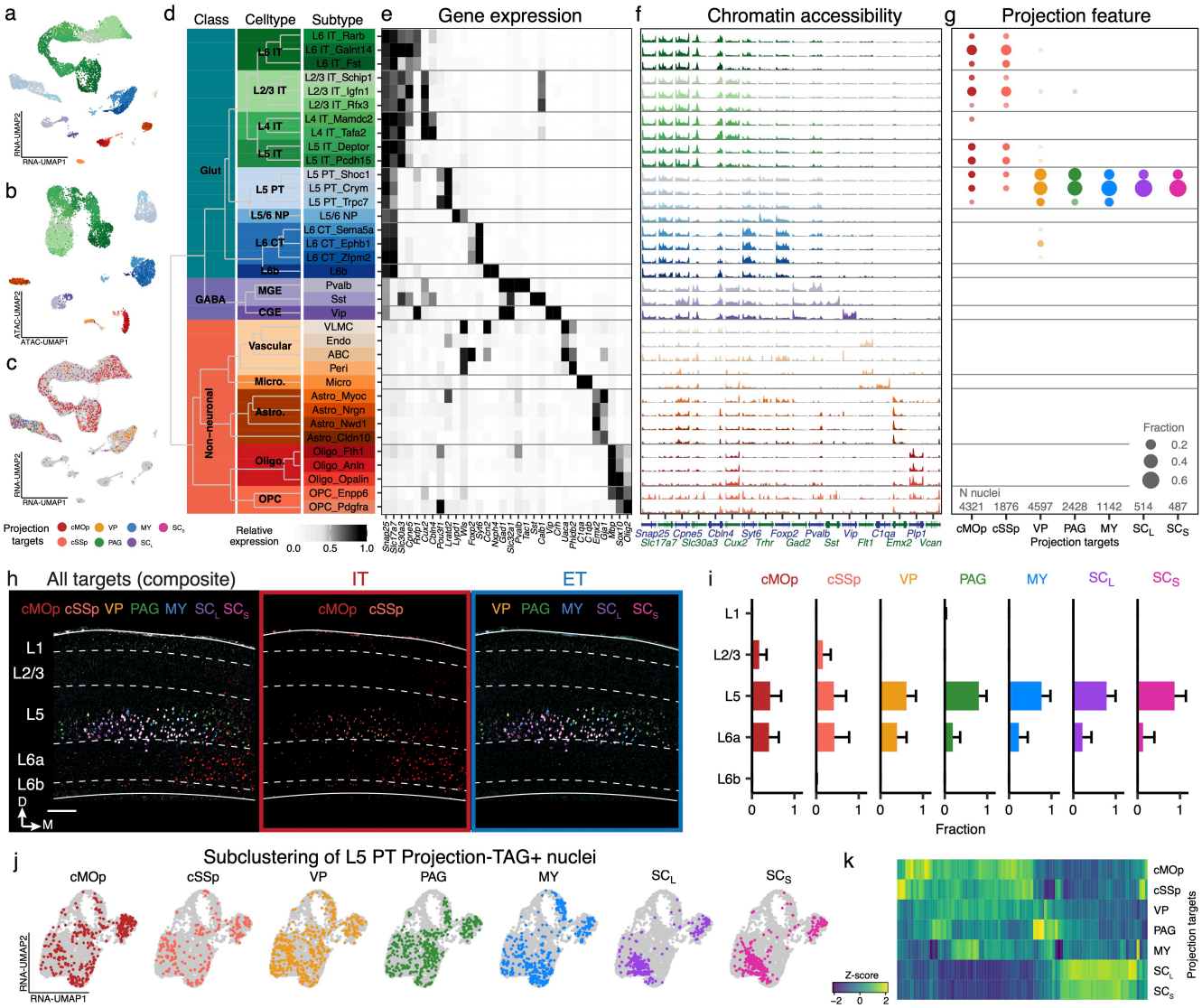


Figure 2. A multi-modal single-cell atlas of mouse cortex

a, Uniform manifold approximation and projection (UMAP)[1] visualization of snRNA-seq data showing 10,000 downsampled nuclei, colored by transcriptional subtypes. **b**, UMAP visualization of snATAC-seq data showing 10,000 downsampled nuclei, colored by transcriptional subtypes. **c**, UMAP visualization of snRNA-seq nuclei, colored by their projection targets. 10,000 downsampled Projection-TAG- nuclei are colored grey and displayed as the background. 500 nuclei (or all nuclei if total nuclei count of one projection is less than 500) expressing only the corresponding Projection-TAG are downsampled for each projection target and displayed. **d**, Hierarchical clustering of 35 snRNA-seq clusters based on the average expression of top 100 marker genes (by FDR) per cluster, and their annotations of class, celltype, and subtype. **e**, Heatmap showing the average expression of select marker genes in individual snRNA-seq clusters. **f**, Coverage plot of average accessibility of canonical celltype-specific marker genes in snATAC-seq nuclei, grouped by transcriptional clusters. The chromatin accessibility is displayed as the average frequency of sequenced DNA fragments per nucleus for each cluster, grouped by 50 bins per displayed genomic region. **g**, Dot plot showing fraction of snRNA-seq nuclei, positive for Projection-TAG of each projection, from each transcriptional cluster. **h**, A representative FISH image showing the layer distribution of projection neurons in MOp from a brain section (-0.91 mm relative to bregma). Colors denote the projection targets based on the expression of Projection-TAGs. Scale bar = 150 μm. Left: composite image of all FISH channels, middle and right: images of a subset of Projection-TAGs injected into IT targets and ET targets, respectively. **i**, FISH quantification showing the fraction of MOp cells retrogradely labeled by each Projection-TAG in each layer. Bar and error bar indicate the average and standard deviation across six slices from two mice. **j**, RNA-UMAP showing subclustering of L5 PT neurons labeled with each Projection-TAG. **k**, Heatmap showing the z-score of average expression of genes (column) differentially expressed in L5-PT neurons positive for each Projection-TAG (row). Data shown on the heatmap can be found in Supplementary Table 4.

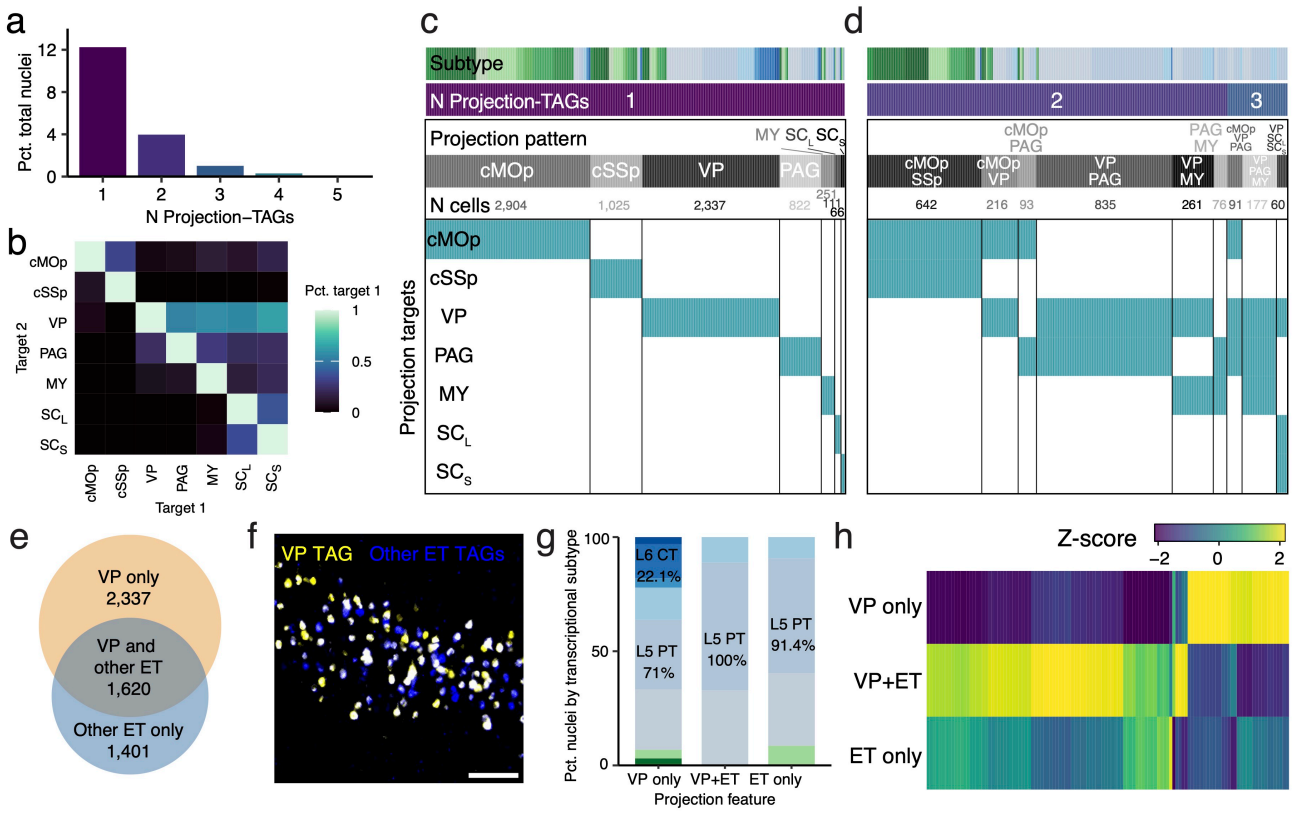


Figure 3. Projection-TAGs revealed axonal collaterals and complex wiring diagram
a, Bar plot showing the percentage of snRNA-seq nuclei, grouped by the number of unique Projection-TAGs detected in each nucleus. **b**, Heatmap demonstrating the pairwise overlapping of snRNA-seq nuclei positive for any two Projection-TAGs, shown as the percentage of nuclei positive for target 1 TAG (column) that are also positive for target 2 TAG (row). **c** and **d**, Heatmaps showing the Projection-TAG (row) detected in individual snRNA-seq nuclei (column) expressing only one Projection-TAG (c) and multiple Projection-TAGs (d). Nuclei are ordered based on the projection patterns, and only groups that passed the FDR cutoff (at least 60 nuclei) were shown on the heatmaps. **e**, Venn diagram showing the overlap between VP-TAG-expressing and other ET-TAG-expressing snRNA-seq nuclei. Other ET-TAGs include Projection-TAGs injected into PAG, MY, SC_L, and SC_S. **f**, Representative FISH image showing MOp cells labeled with VP- and other ET-TAGs. Scale bar = 100 µm. **g**, Distribution of transcriptional subtypes in each projection group. Transcriptional subtypes with < 60 nuclei from each projection feature group were excluded and celltypes that make up at least 10% nuclei in the projection group are annotated on the plot. **h**, Heatmap showing the z-score of average expression of 118 genes (row) differentially expressed in transcriptional L5 PT nuclei in each projection group (column). Data shown on the heatmap can be found in Supplementary Table 5.

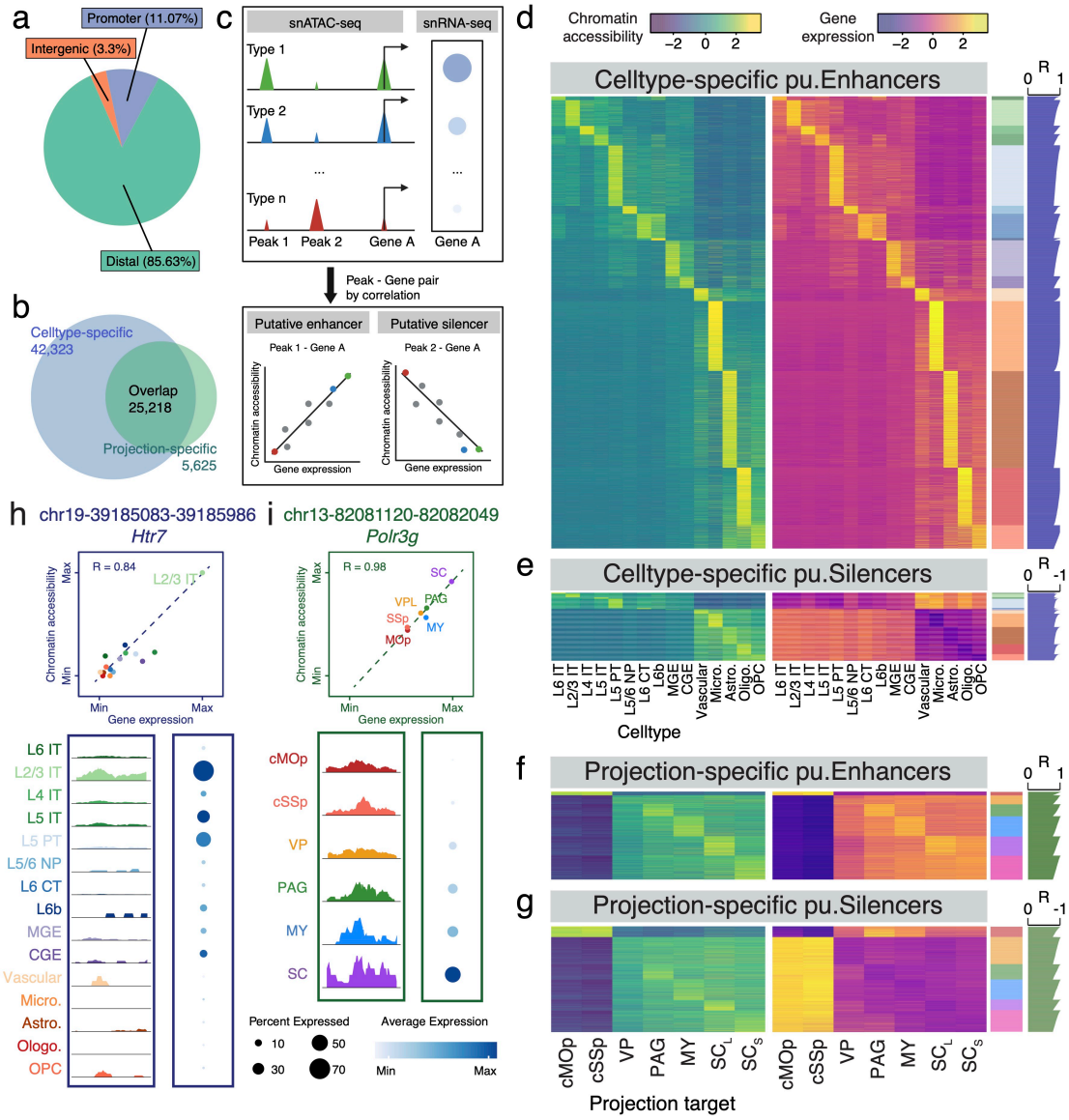


Figure 4. Genomic regulatory networks of cortical cell types using Projection-TAGs

a, Pie chart categorizing snATAC-seq peaks based on their genomic locations. Peaks are mapped to one of the three categories: promoter regions (-1,000 bp to +100 bp of transcription start site [TSS]), distal regions (<200 kb upstream or downstream of TSS or within gene body, excluding promoter), and intergenic regions (>200kb upstream or downstream of TSS, excluding gene body). **b**, Overlapping of celltype-specific peaks and projection-specific peaks in snATAC-seq data. **c**, Analytical workflow of identifying putative enhancers and silencers and their regulated genes. **d** and **e**, Heatmaps showing 18,088 pairs of celltype-specific pu.Enhancers and regulated genes (d) and 2,739 pairs of celltype-specific pu.Silencers and regulated genes (e). Left heatmaps show the average accessibility of pu.Enhancer peaks (row) in individual celltypes (column) in snATAC-seq data and right heatmaps show the average expression of regulated genes (row) in individual celltypes (column) in snRNA-seq data. On the right side next to the heatmaps, the color bar indicates the celltypes in which the peak and gene have highest accessibility and expression, and the bar plot shows the Pearson correlation coefficient (R) between the accessibility of the peak and the expression of the gene in individual celltypes. Data shown on the plots can be found in Supplementary Table 6. **f** and **g**, Heatmaps showing 3,545 pairs of projection-specific puEnhancers and regulated genes (f) and 4,200 pairs of projection-specific pu.Silencers and regulated genes (g). Left heatmaps show the average accessibility of pu.Enhancer peaks (row) in nuclei positive for each Projection-TAG (column) in snATAC-seq data and right heatmaps show the average expression of regulated genes (row) in nuclei positive for each Projection-TAG (column) in snRNA-seq data. On the right side next to the heatmaps, the color bar indicates the projection targets in which the peak and gene have highest accessibility and expression, and the bar plot shows the Pearson correlation coefficient (R) between the accessibility of the peak and the expression of the gene in nuclei positive for each Projection-TAG. Data shown on the plots can be found in Supplementary Table 6. **h**, Top: Scatter plot showing the correlation between the average accessibility of peak chr19-39,185,083-39,185,986 in snATAC-seq and the average expression of gene *Htr7* in snRNA-seq in individual celltypes (chromatin accessibility and gene expression are normalized to their max values). The dotted line represents the line of best

1316 fit. Bottom left: chromatin accessibility at the genomic locus of chr19-39,185,083-39,185,986, displayed as the average fraction
1317 of transposase-sensitive fragments per nucleus in each celltype (grouped by 50 bins per displayed genomic region).
1318 Accessibility at each locus (y axis) is scaled to the max value across all cell types. Bottom right: expression of *Htr7* in each
1319 celltype. **i**, Top: Scatter plot showing the correlation between the average accessibility of peak chr13-82,081,120-82,082,049
1320 and the average expression of gene *Polr3g* in nuclei positive for each Projection-TAG (chromatin accessibility and gene
1321 expression are normalized to their max values). Bottom left: chromatin accessibility at the genomic locus of chr13-82,081,120-
1322 82,082,049, displayed as the average fraction of transposase-sensitive fragments per nucleus in nuclei positive ach
1323 Projection-TAG (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Polr3g* in nuclei positive for
1324 each Projection-TAG. Colors indicate the injection site of the corresponding TAG. Nuclei positive for SC_L- and SC_S-TAGs are
1325 combined for visualization due to low cell number.

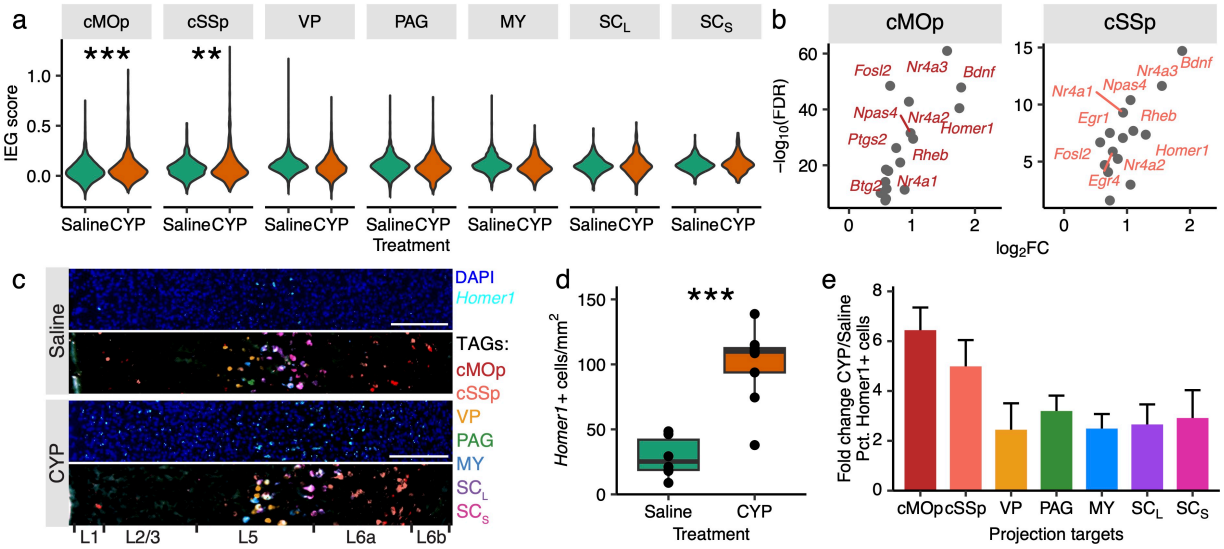


Figure 5. Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain
a, Violin plots showing IEG scores in snRNA-seq nuclei for each projection between treatments (***p = 4.6e-14 for cMOp, **p = 0.004 for cSSp, p = 1 for VP, p = 0.99 for PAG, p = 0.99 for MY, p = 0.4 for SC_L, p = 0.2 for SC_S, one-side t-test with Bonferroni correction). **b**, Volcano plots showing significantly induced IEGs (log2FC > 0.5) in activated nuclei from CYP-treated mice compared to the same number of randomly sampled nuclei (with matched transcriptional subtypes) from saline-treated mice for the two IT projections. **c**, Representative FISH images showing the expression of *Homer1* and Projection-TAGs in MOp (from brain section approximately -0.2 mm relative to bregma) from saline- and CYP-treated mice. Scale bar: 100 µm. **d**, Quantification of *Homer1*+ cells in MOp and SSp areas from CYP- and saline-treated mice (0 to -0.92 mm relative to bregma, nine slices from three mice per group, ***p = 1.9e-5, one-side t-test). **e**, Percentage of *Homer1*+ MOp and SSp cells in each projection. Fold change is calculated as the percentage in individual CYP slices, divided by the average percentage across saline slices. Bar and error bar indicate the mean and standard deviations across nine slices per group.

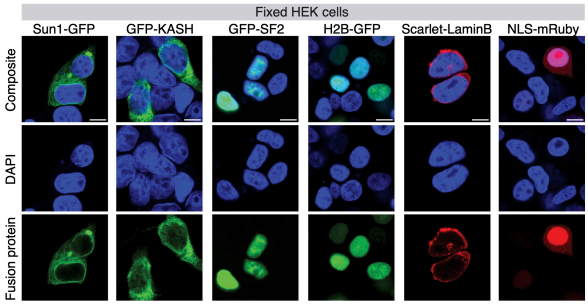
Extended Data Fig. 1

a

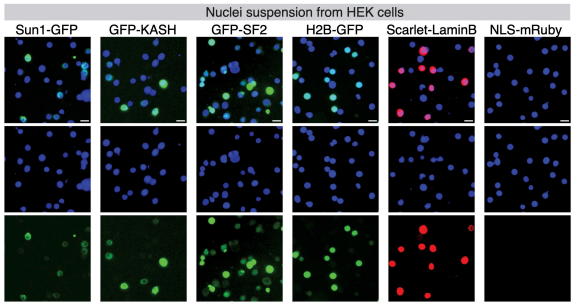
	Virus				BC		
	Viral species	Tracing scheme	Tracing timeline	Viral features	BC design	BC utility	Available BC viruses for projection tracing
MAP-seq/BAR-seq	Sinbis virus	Anterograde	Days	High replication and cargo gene expression rate	30-nt with random sequence Anterodrade transport to axonal terminals	Cell identifier	N/A
Projection-seq	rAAV2-retro	Retrograde	Weeks to months	Low toxicity Flexibility of packaging into other serotypes	300-400-nt truncated protein coding genes with known sequence	Projection identifier	7
MERGE-seq	rAAV2-retro	Retrograde			15-nt with known sequence		13
Projection-TAGs	rAAV2-retro	Retrograde			100-nt with known sequence		12

	Tracing experiment				BC detetion			
	Viral sample	Available BC viruses for projection tracing	Viral injection	Tissue collection	Sequencing-based detection strategy	Demonstrated scRNA-seq platform	BC detection rate in sn/scRNA-seq	Spatial assay
MAP-seq/BAR-seq	A sindbis viral library	N/A	One injection into ROI	ROI, projection targets	RNA-seq to quantify BC expression in projection targets, can be coupled with scRNA-seq of ROI	Fluidigm C1 system (Klingler et al. 2021)	16% (391 out pf 2,450 neurons whose barcodes are also identified in projection targets)	BARseq
Projection-seq	Individual AAV samples each expressing a unique BC	7	One injection into each of the projection targets pf ROI	ROI	scRNA-seq	10X Genomics gene expression 3' V3	32.8% (4,791 out of 14,590 Phox2b+ placode-derived nodose VSNs)	N/A
MERGE-seq		13			scRNA-seq coupled with BC target amplification	10X Genomics gene expression 3'	24% (2,116 out of 8,805 excitatory neurons in ventromedial prefrontal cortex)	N/A
Projection-TAGs		12			snRNA-seq and snMultiome-seq. Optional BC target amplification	10X Genomics gene expression 3' V3 10X Genomics Multiome V2	32.8±5.7% with target amplification (or 27.4±5.6% in snRNA-seq only)	Multiplexed FISH

b



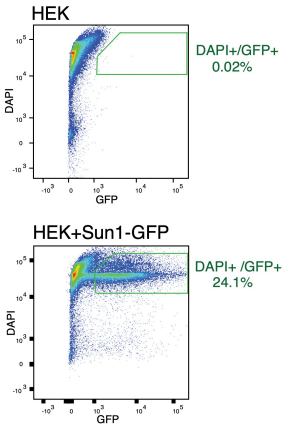
c



d

	Source	Marker gene	Localization domain	Localization domain properties	Label nuclei in intact cells?	Label nuclei in suspension and during FACS?
Sun1-GFP	Addgene #160141	Superfold GFP	Sad1 and UNC84 domain-containing 1 (Sun1)	localizes to the nuclear envelope in a manner that is resistant to detergent tissue homogenization	Yes	Yes
GFP-KASH	Addgene #60958	EGFP	Klarsicht, ANC-1, Syne Homology (KASH)	Interacts with SUN domains between the outer and inner nuclear membranes to bridge the nuclear envelope	Yes	Yes
GFP-SF2	Addgene #17990	EGFP	pre-mRNA-splicing factor SF2 or ASF1/SF2	An essential pre-mRNA splicing factor, which is concentrated in nuclear splicing factor compartments	Yes	Yes
H2B-GFP	Addgene #116869	mEGFP	Histone H2B	One of the five main histone proteins for the structure of chromatin	Yes	Yes
Scarlet-LaminB	Addgene #98831	mScarlet-1	Lamin B1 (LMNB1)	The lamin family proteins make up the nuclear laminar matrix. Lamin B is anchored to the inner surface of the nuclear membrane	Yes	Yes
NLS-mRuby	Addgene #99130	mRuby2	Nuclear localization sequence (NLS)	A sequence that signals a protein for import into the nucleus	Yes	No

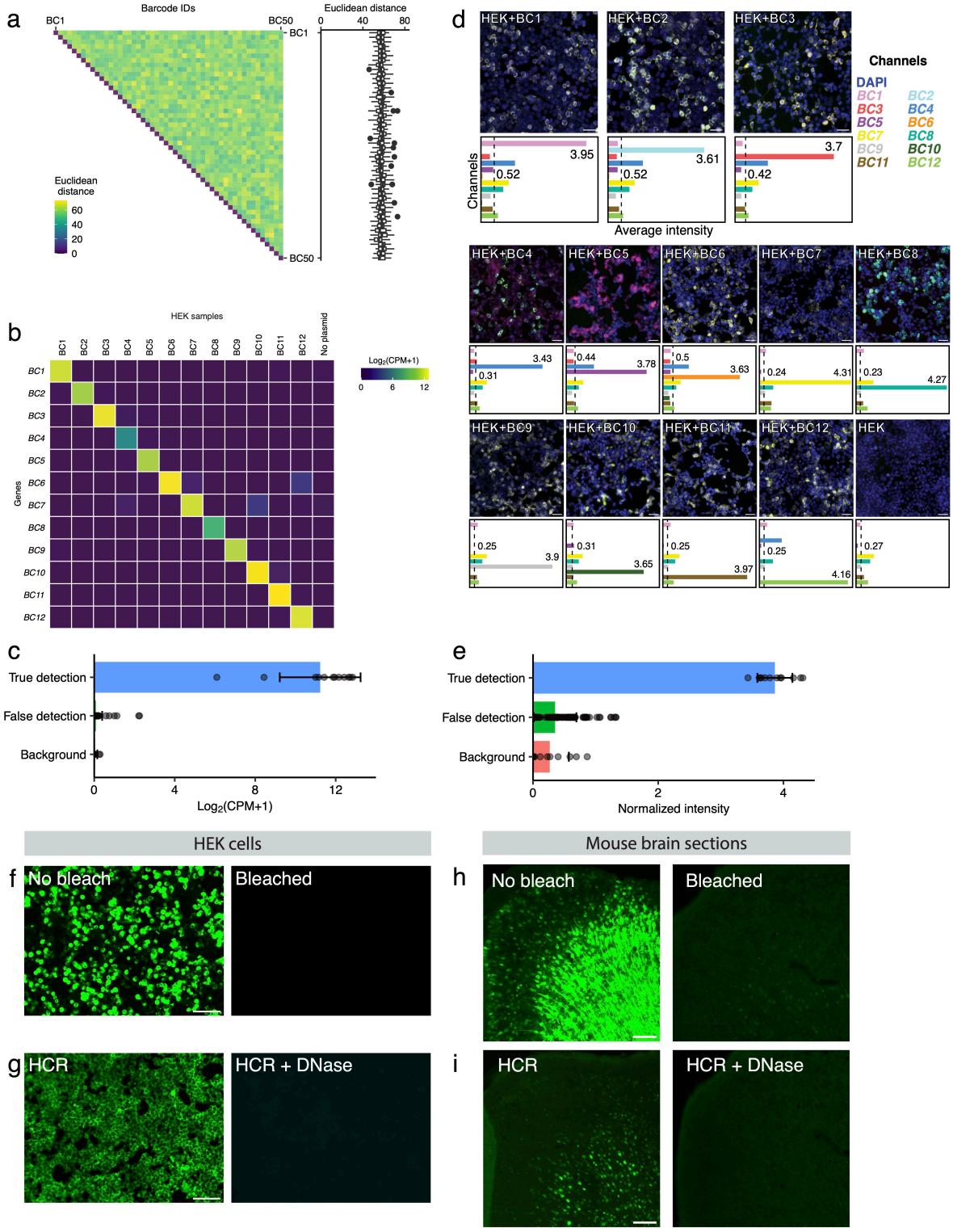
e



1338

1339 **Extended Data Fig. 1 – related to Fig. 1**
1340 **a**. Summary of available high-throughput neuroanatomical tools. **b**, Representative fluorescent images of the localization of
1341 fusion proteins in different cellular compartments in fixed HEK cells. Scale bar: 10 μm. **c**, Representative images of the
1342 fluorescent labeling using fusion proteins of nuclear resuspension extracted from HEK cells. Scale bar: 20 μm. **d**, Summary of
1343 the fluorescent proteins and their ability of labeling of cells and nuclei of HEK cells. **e**, Scatter plot showing the gate of sorting
1344 labeled HEK nuclei (DAPI+ and GFP+) extracted from HEK cells expressing Sun1-GFP.

Extended Data Fig. 2

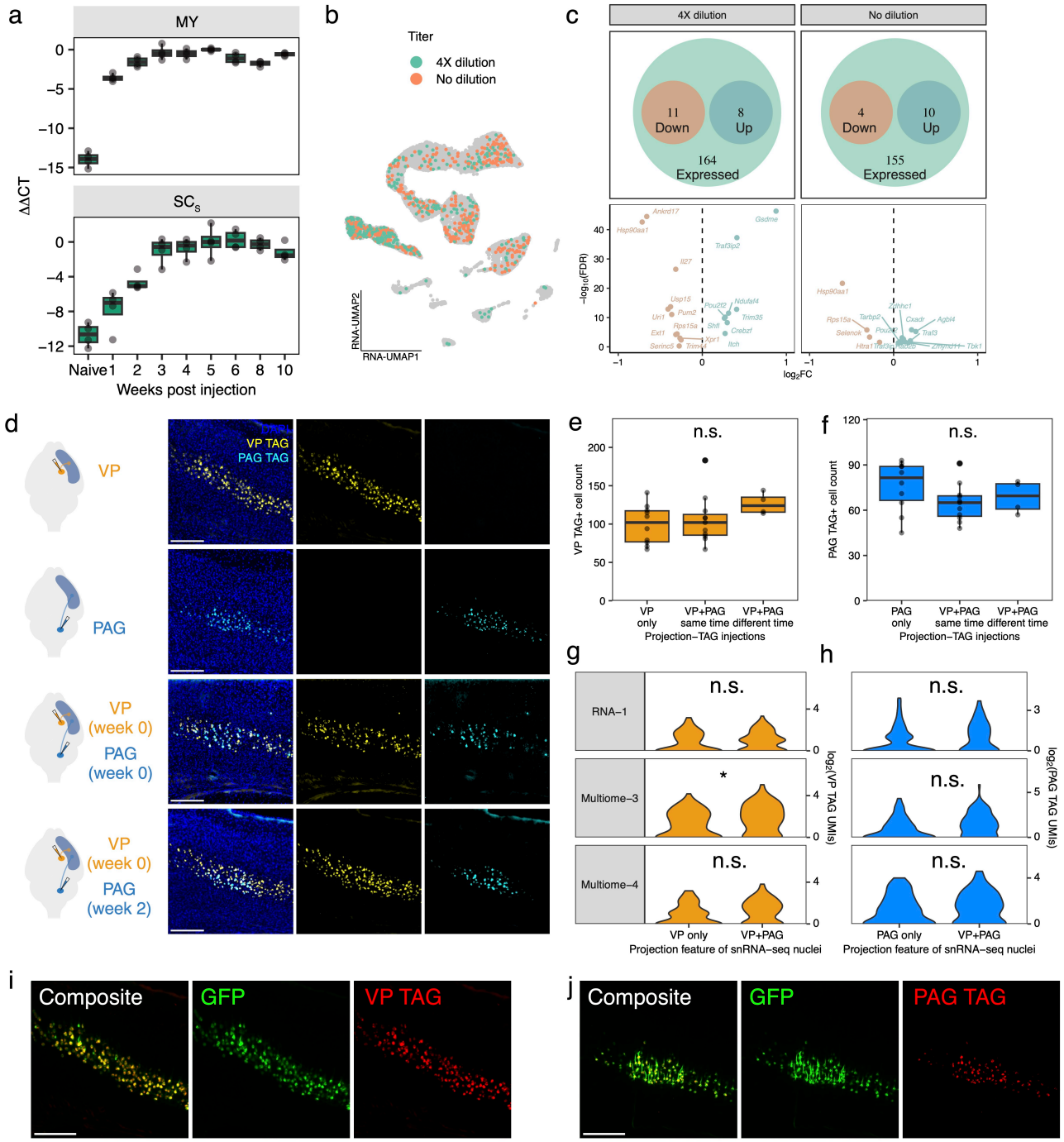


1345

Extended Data Fig. 2 – related to Fig. 1

a, Left: pairwise Euclidean distance between any two of the 50 BCs cataloged in this study. Right: box plot showing the average Euclidean distance of individual BCs to other BCs. **b**, Counts of sequencing reads that map to individual BCs (row) in individual HEK samples (column) in RNA-seq. Each HEK sample was either transfected with a plasmid expressing a unique Project-TAG BC or no transfection. **c**, Bar plot showing the average detection rate of Projection-TAGs in individual samples, error bars are standard deviation. True detection is the detection of the Projection-TAG expressed in a given sample transfected with one Projection-TAG plasmid, false detection is the detection of Projection-TAGs not expressed in the same samples, and the background is the detection of individual Projection-TAGs in the HEK sample with no transfection. **d**, Representative FISH images showing the detection of individual Projection-TAGs (BCs 1-12) in individual HEK samples each expressing a unique TAG or the HEK sample with no transfection. The bar plots show the average signal intensity from each Projection-TAG channel across a field of view with 1 mm by 1 mm. The average intensity is normalized by the max values of each sample and each channel. In HEK samples transfected with Projection-TAG plasmids, the dotted line and the associated number represent the false detection rate (average intensity of Projection-TAGs not expressed in a given sample), whereas the number on the right indicates the true detection rate (intensity from the Projection-TAG expressed in a given sample). In the HEK sample with no transfection, the dotted line and the associated number represent the background (average intensity of all Projection-TAGs). Scale bar: 20 μm. **e**, Bar plot showing average detection rate of Projection-TAGs across all samples, error bars are standard deviation. **f**, Representative images showing that photobleaching process removed the native Sun1-GFP fluorescence from HEK cells. Scale bar: 50 μm. **g**, Representative images showing that DNase incubation removed the fluorescent signal of Alexa-488 conjugated with *GAPDH* FISH probe in HEK cells. Scale bar: 50 μm. **h**, Representative images showing that photobleaching process significantly reduced the native Sun1-GFP fluorescence from mouse brain sections. Scale bar: 100 μm. **i**, Representative images showing that DNase incubation removed the fluorescent signals of Alexa-488 conjugated with the FISH probe targeting *GFP* in mouse brain sections. Scale bar: 100 μm.

Extended Data Fig. 3



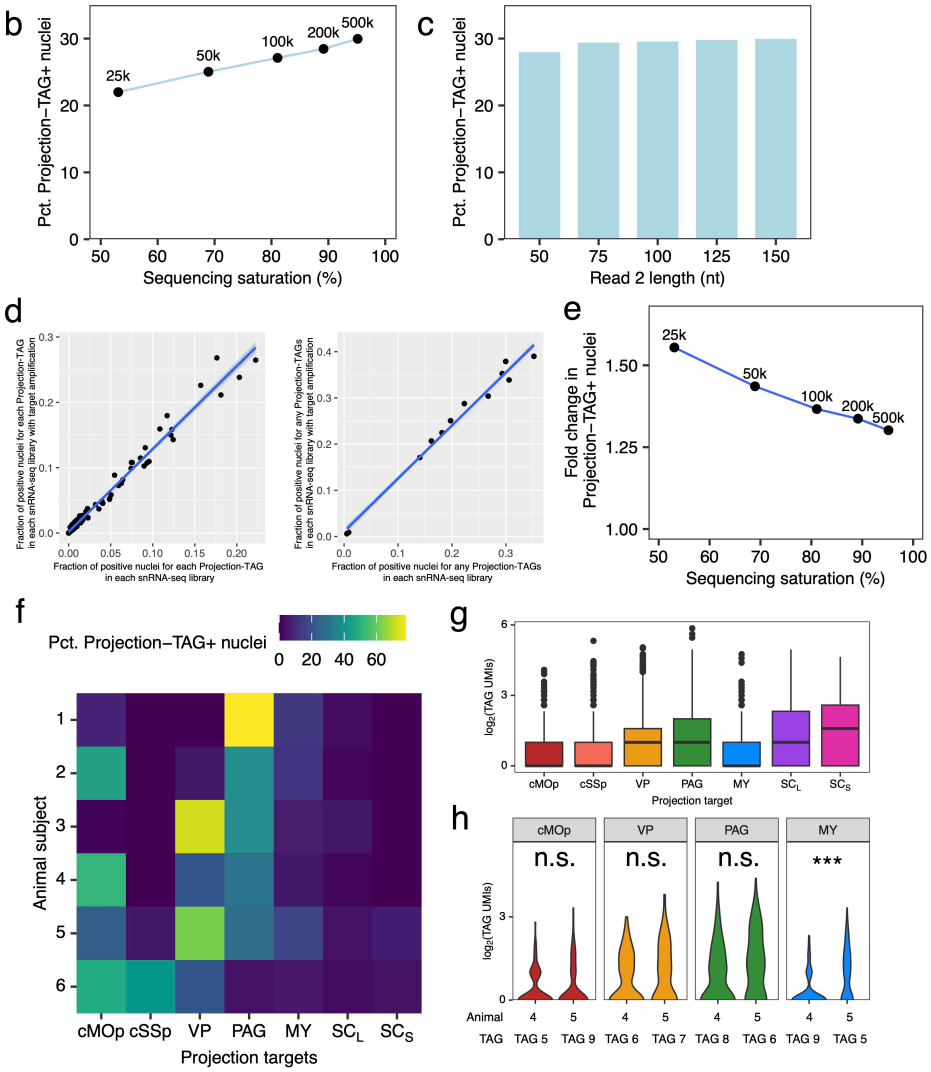
Extended Data Fig. 3 – related to Fig. 1

a, Box plot showing the time kinetics of Projection-TAG expression in the cortex. Animals received MY or SC_S injection with the Projection-TAG AAV (expressing TAG 2), cortical samples were collected at different time points post injection, and the Projection-TAG expression is measured using qPCR (n=4 for each time point and each injection). ΔCT is calculated as the CT value of TAG 2 minus that of GAPDH for each sample, ΔΔCT is calculated as the ΔCT at each time point minus the ΔCT at 4 weeks post-injection. **b**, RNA-UMAP of 10,000 nuclei (1,000 downsampled Projection-TAG+ nuclei from each titer group, colored by titer, and 8,000 downsampled BC- nuclei, colored grey). **c**, DE genes in response to viral infection. DE analysis was performed comparing the expression of genes annotated in the Gene Ontology list GO:0009615 (response to virus) between Projection-TAG+ nuclei and Projection-TAG- nuclei (downsampled to match the Celltype distribution and nuclei counts to the Projection-TAG+ nuclei) in each tier group. Top panels show the number of upregulated (avg_log2FC > 0) and downregulated (avg_log2FC < 0) viral responding genes among the viral responding genes expressed (average expression > 0.5) in snRNA-seq nuclei from each titer group. Fisher's exact test was performed to test if viral responding genes are significantly enriched among DE genes (p = 0.81 and 0.77 for undiluted and diluted group, respectively). Bottom panels show the individual DE viral responding genes. **d**, Representative FISH images of medial part of the posterior parietal association area of the cortex (PTLp) visualizing Projection-TAGs injected into VP and PAG in a mouse receiving only VP injection (first row), a mouse receiving only PAG injection (second row), a mouse receiving VP and PAG injections at the same time (third row), and a mouse receiving PAG injection two weeks after VP injection (bottom row). Scale bar: 100 μm. **e**, Quantification of VP-TAG+ cells in medial PTLp from different injection groups (nine slices from three mice each group). $p = 0.26$, $F(2,25) = 1.72$, one-way ANOVA. **f**, Quantification of PAG-TAG+ cells in medial PTLp from different injection groups (9 slices from three mice each group). $p = 0.27$, $F(2,25) = 1.4$, one-way ANOVA. **g**, Violin plot showing the VP-TAG UMIs from snRNA-seq nuclei expressing only VP-TAG and co-expressing VP/PAG-TAGs in three snRNA-seq libraries that contain at least 60 nuclei co-expressing VP/PAG-TAGs. $p = 0.99$ (RNA-1), 0.01 (Multiome-3), and 0.06 (Multiome-3), student t-tests with Bonferroni correction. **h**, Violin plot showing the PAG-TAG UMIs from snRNA-seq nuclei expressing only PAG-TAG and co-expressing VP/PAG-TAGs in the snRNA-seq libraries shown in g. $p = 0.5$ (RNA-1), 0.1 (Multiome-3), and 0.6 (Multiome-3), student t-tests with Bonferroni correction. **i** and **j**, representative images showing the co-localization of GFP fluorescent signals with VP-TAG signal (i) or PAG-TAG signal (j) from FISH. Scale bar: 100 μm.

Extended Data Fig. 4

a

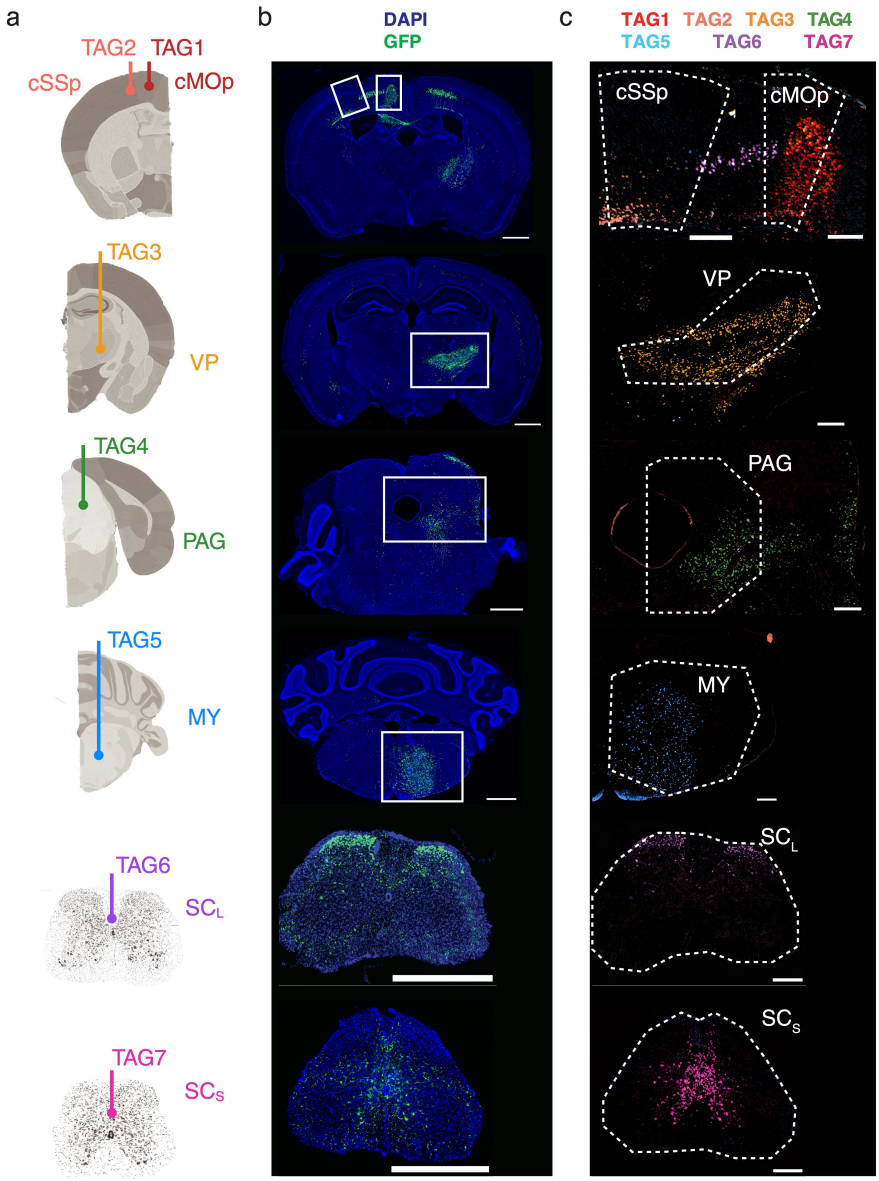
	Sequencing depth		Projection-TAG detection	
	Mean reads/nucleus	Mean UMIs/nucleus	Percentile by expression rank	Pct. of Projection-TAG+ nuclei
Chromium Single Cell Multiome Kit, 10X Genomics	63.2 ± 10.7	9.7 ± 2	63.8 ± 8.9	0.39 ± 0.3
Evercode WT Mini kit, Parse Biosciences	59.2 ± 2.1	20.3 ± 0.5	54.1 ± 6.2	0.04 ± 0.03



1400 **Extended Data Fig. 4 – related to Fig. 1**
1401 **a**, Key sequencing and Projection-TAG detection metrics from snRNA-seq libraries prepared using different commercial kits
1402 from nuclei of the same biological samples (n = 2 for each kit). **b**, Dot plot showing the effect of sequencing depth/saturation
1403 on the percentage of Projection-TAG+ nuclei in the snRNA-seq library that reached 500k reads/nucleus. Sequencing reads
1404 were randomly downsampled to match the sequencing reads/nucleus metric shown on the plot. **c**, Bar plot showing the effect
1405 of sequencing Read 2 length on the percentage of Projection-TAG+ nuclei in the snRNA-seq library shown in b. Sequencing
1406 reads were trimmed using "r2-length" parameter in cellranger. **d**, Detection of each Projection-TAG (left) and any Projection-
1407 TAGs (right) in snRNA-seq alone (x-axis) and in snRNA-seq library with target amplification (y-axis) in individual sequencing
1408 libraries. Blue line on each plot indicates the best fitted line. **e**, Dot plot showing the target amplification performance as fold
1409 change of Projection-TAG+ nuclei between snRNA-seq library with target amplification and snRNA-seq alone in the snRNA-
1410 seq library shown in b. Sequencing reads/nucleus metric is labeled on the plot **f**, Heatmap showing the percentage of
1411 Projection-TAG+ nuclei expressing each Projection-TAG (labeled as the injection site of the corresponding Projection-TAG on
1412 the row) from each animal. **g**, Box plot showing the Projection-TAG UMI count from snRNA-seq nuclei expressing each
1413 Projection-TAG (labeled as the injection site of the corresponding Projection-TAG on the row). **h**, Violin plots showing the
1414 Projection-TAG UMI count from snRNA-seq nuclei expressing the corresponding Projection-TAG from two animals (x-axis), in
1415 which different Projection-TAGs are injected into the same brain region for the regions shown on the plot. p = 0.08 cMOp, p =
1416 0.12 VP, p = 0.65 PAG, ***p = 1.5e-4 MY.

1417

Extended Data Fig. 5



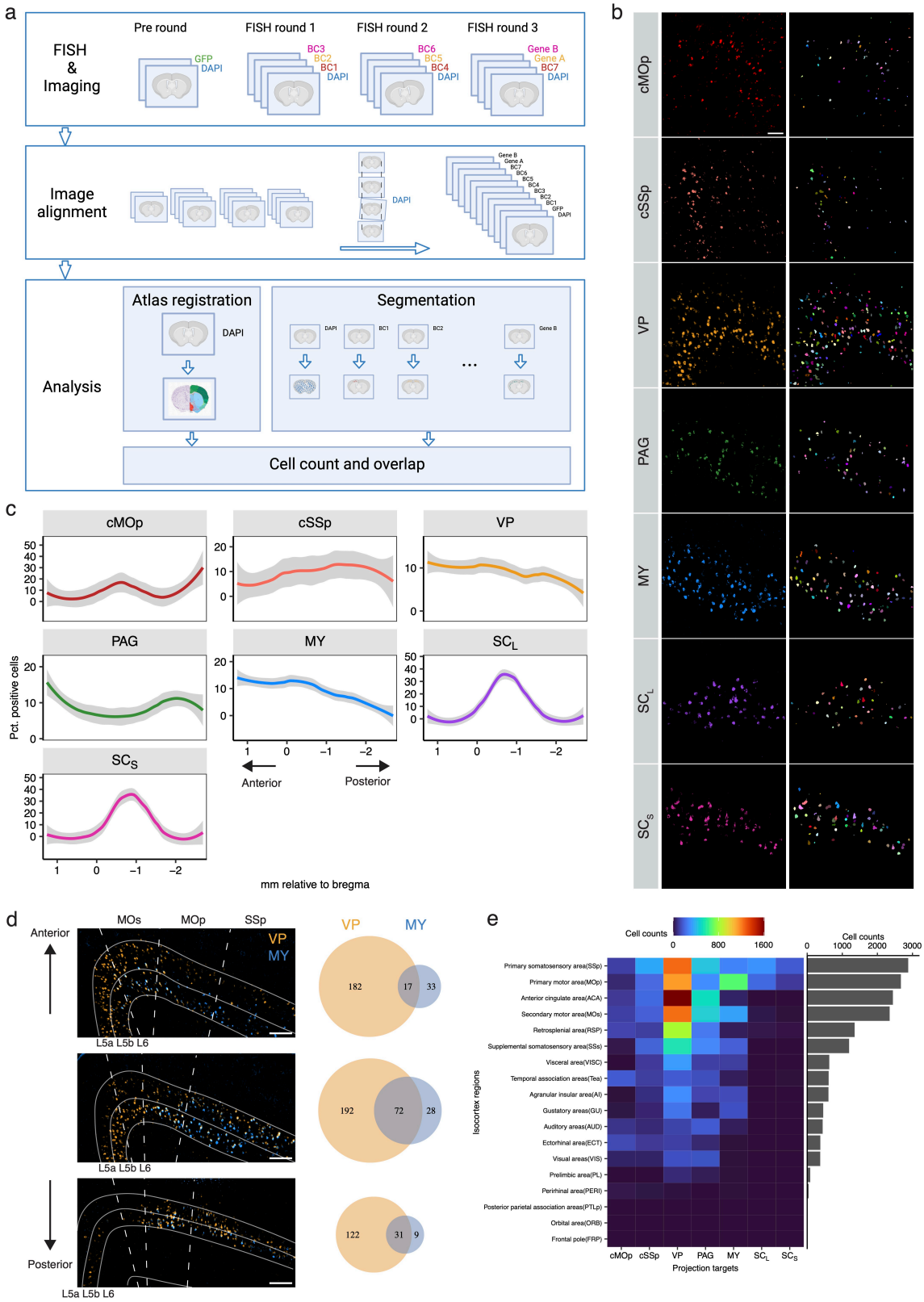
1418
1419

Extended Data Fig. 5 – related to Fig. 1

a, Schematics showing the stereotaxic injection of individual Projection-TAGs into each of the projection targets of the cortex. **b**, Representative images showing the native fluorescence of Sun1-GFP in the injection sites. Scale bar: 1 mm. **c**, Representative FISH images (zoom-in view of the region highlighted in b) showing the expression of Projection-TAGs in the injection sites. Multiplexed FISH was performed to visualize the Projection-TAGs (TAGs 1-7) in the projection targets of the cortex (highlighted with dotted lines). Scale bar: 300 μm.

1426

Extended Data Fig. 6



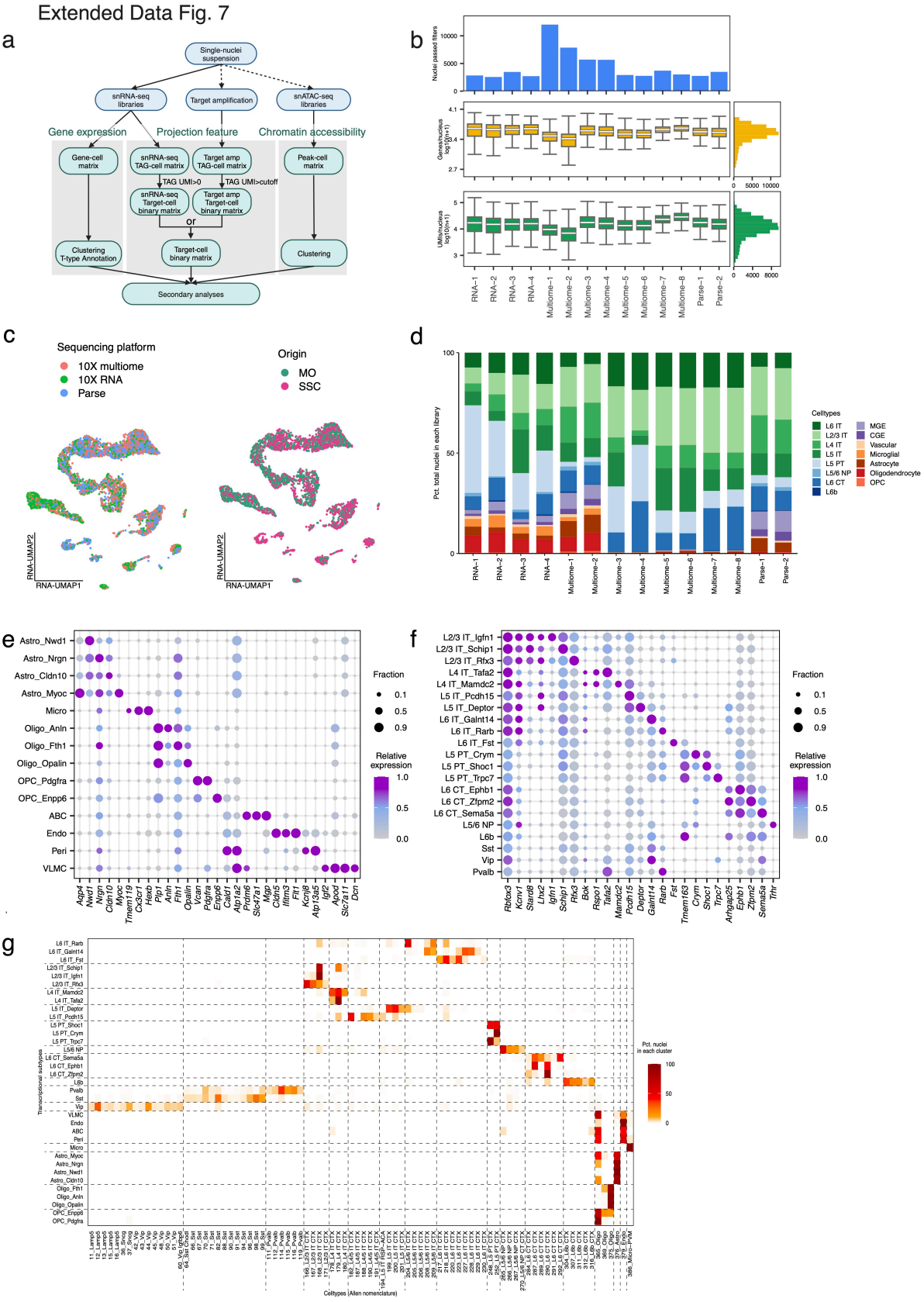
1427

Extended Data Fig. 6 – related to Fig. 1

a, Workflow of multiplexed FISH experiment and imaging analysis. **b**, Left: representative FISH images showing the Projection-TAGs labeled projection neurons to each target (row) in the MOp. Right: cell segmentation of the projection neurons from the images on the left panels using machine-learning-based algorithms. Colors denote individual objects identified by the algorithms. **c**, Rostro-caudal distribution of neurons projecting to each target in ipsilateral cortex areas. Smooth lines were fitted to the quantification of 22 brain sections from two mice, and the grey ribbons show the standard error of the mean (SEM). **d**, Representative images of VP- and MY-projecting neurons labeled with Projection-TAGs in the select cortex areas from three brain sections (top: 1.26 mm, middle: 0.23 mm, bottom -0.95 mm relative to bregma). Venn diagrams show overlapping of cells projecting to both targets (p = 7.57e-14, 1.23e-44, 3.30e-41 for top, middle, and bottom diagrams, respectively. Hypergeometric tests). Scale bar: 200 μm. **e**, Heatmap showing the cell counts of neurons projecting to each target in each select cortex area.

1439

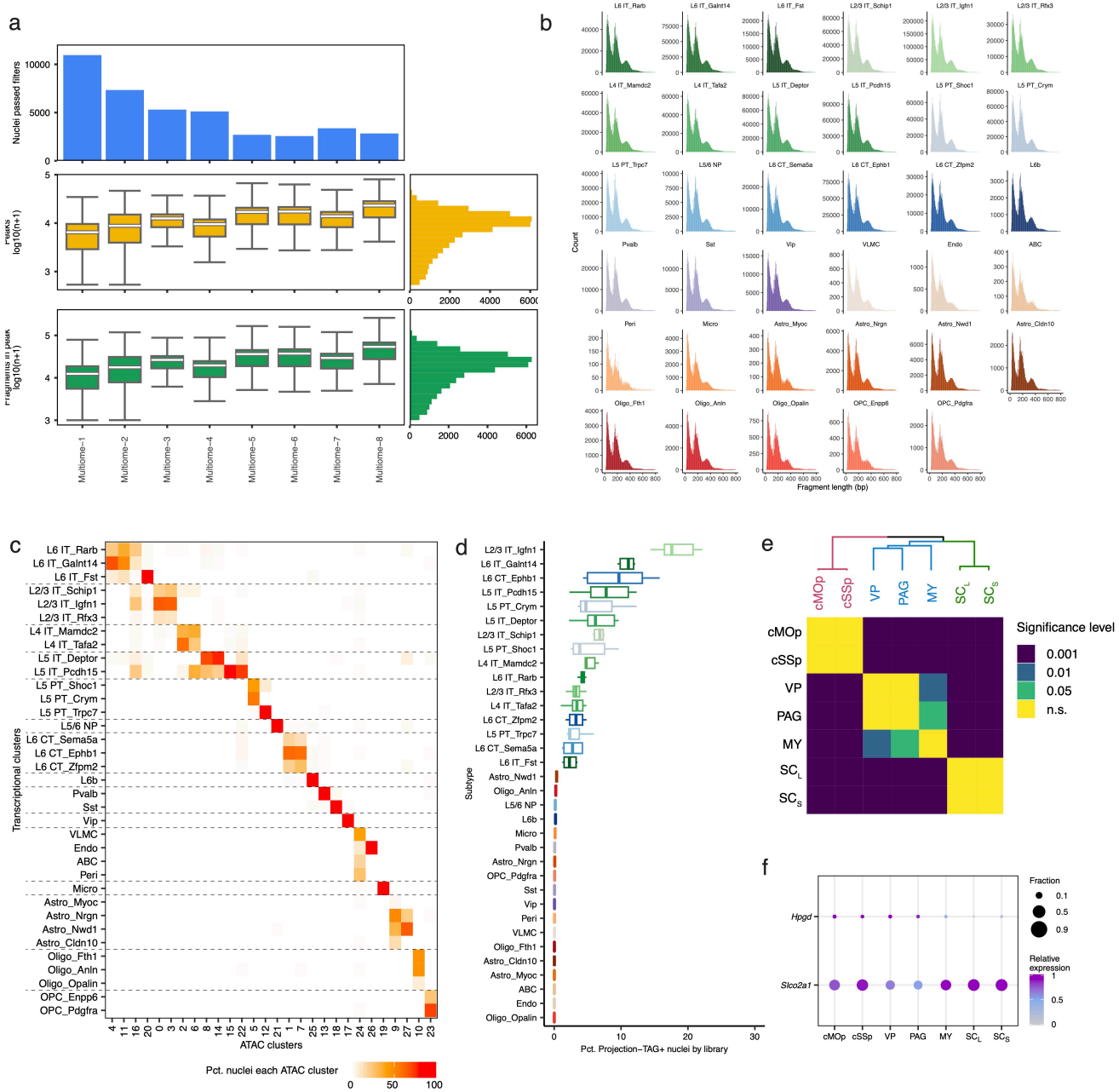
Extended Data Fig. 7



1440

1441 **Extended Data Fig. 7 – related to Fig. 2**
1442 **a**, Workflow of the data analysis of the snRNA-seq and snATAC-seq and celltype and projection annotation. **b**, snRNA-seq
1443 library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of genes per
1444 nucleus (log10 transformed) and the bottom row displays the number of UMIs per nucleus. Boxes indicate quartiles and
1445 whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is
1446 aggregated across all samples and displayed on the horizontal histogram. **c**, RNA-UMAP of 1,000 downsampled nuclei per
1447 group. Nuclei were colored by sequencing kits (left) or the origin of brain regions (right). **d**, Composition of transcriptional
1448 celltypes in each snRNA-seq library. **e** and **f**, Dot plots showing the marker genes used to annotate subtypes of non-neuronal
1449 cells (e) and neuronal cells (f). **g**, Heatmap showing the percentage of snRNA-seq nuclei in each cluster (row) assigned by
1450 anchoring analysis to the celltypes (column) of a published scRNA-seq dataset of mouse MOp and SSp[123].

Extended Data Fig. 8

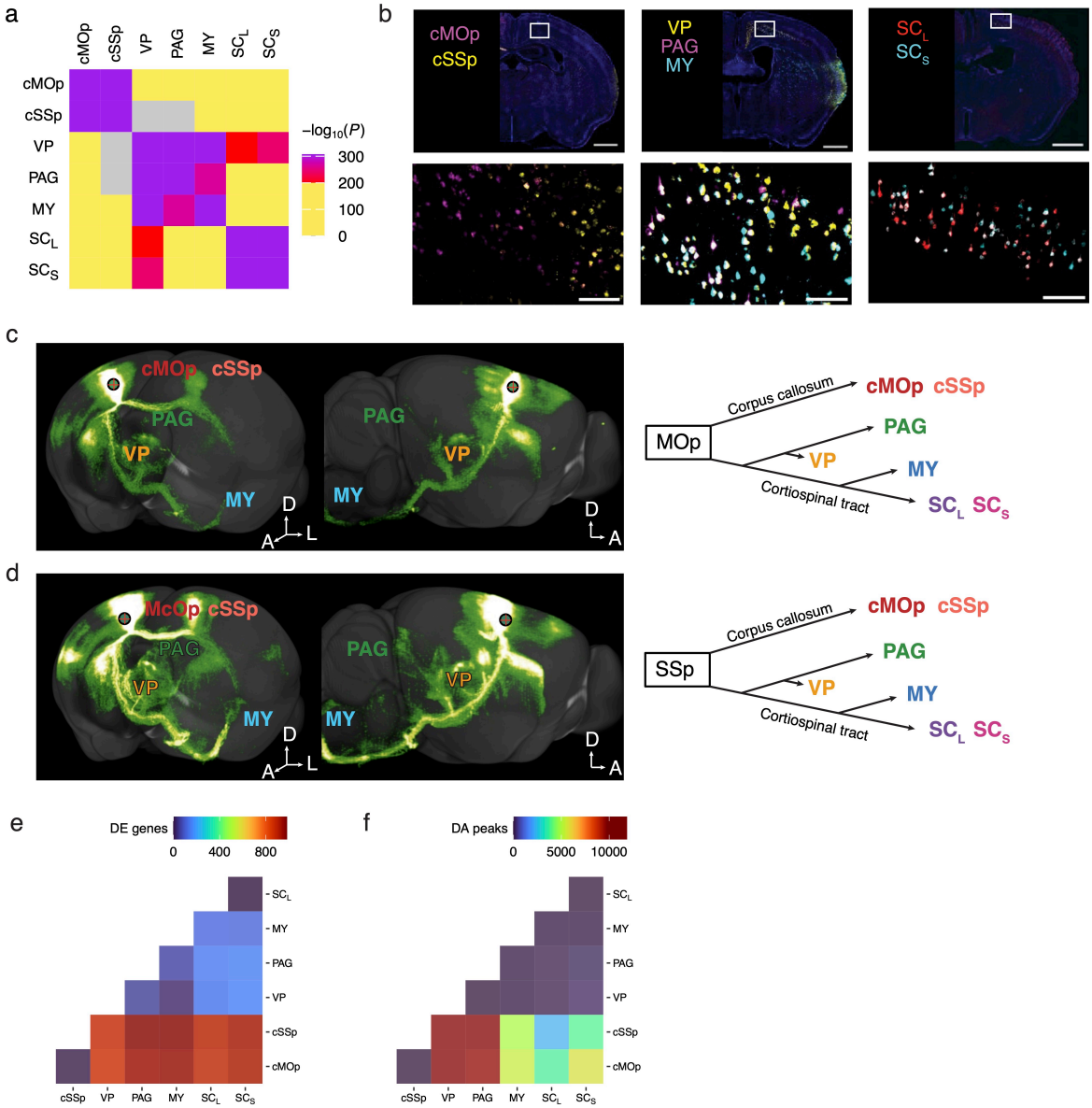


1451

Extended Data Fig. 8 – related to Fig. 2

a, snATAC-seq library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of peaks per nucleus (log10 transformed) and the bottom row displays the number of fragments that overlap with peaks per nucleus. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is aggregated across all samples and displayed on the horizontal histogram. **b**, Distribution of snATAC-seq fragment lengths in nuclei grouped by transcriptional subtypes. **c**, Correspondence of ATAC clusters (row) and transcriptional subtypes (column) of nuclei profiled combinatorially by snRNA-seq and snATAC-seq. Plot displays the percentage of nuclei within each ATAC cluster that is assigned to each transcriptional subtype. **d**, Box plot showing the percentage of Projection-TAG+ nuclei in each snRNA-seq library in each subtype. **e**, Heatmap showing the significance levels of the Bonferroni-corrected *p* values of pairwise hypergeometric tests of subtype distribution, shown in Fig. 2g, between any of the two projections. Hierarchical clustering was performed based on the subtype distribution for each projection. **f**, Dot plot showing the expression of *Hpgd* and *Slco2a1* in individual projections.

Extended Data Fig. 9

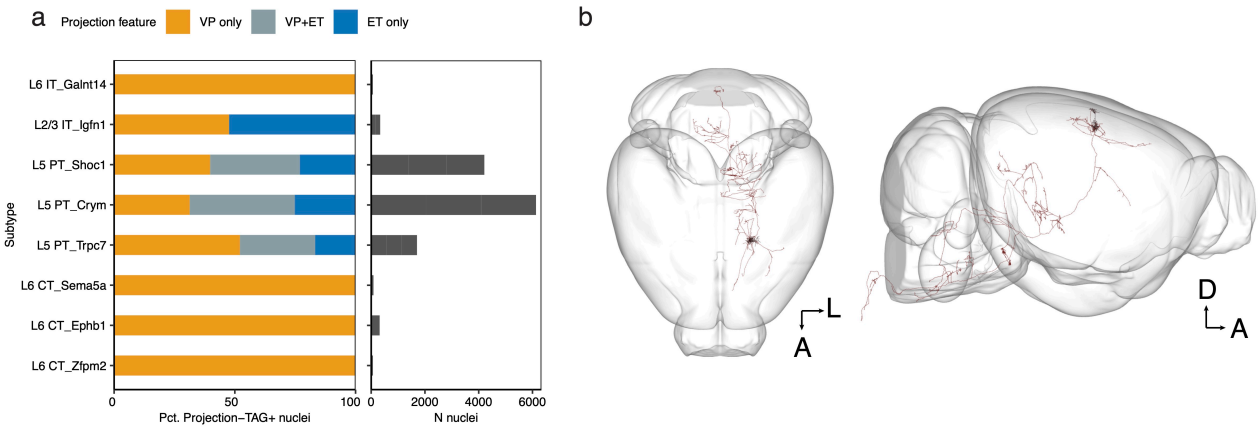


1465

Extended Data Fig. 9 – related to Fig. 3

a, Heatmap showing the Bonferroni-corrected *p* values from pairwise hypergeometric tests of overlapping between any of the two projections. Two projections were considered highly significantly overlapping with $p < 1e\text{-}200$ (shown as red to purple on the heatmap). **b**, Representative FISH images showing projection neurons labeled by Projection-TAGs in the coronal section showing motor cortex from different anterior-posterior sections (top) and zoomed-in images (bottom) of the highlighted region demonstrate expression of two unique Projection-TAGs injected into downstream cMOp and cSSp; middle: demonstrate expression of three unique Projection-TAGs injected into the VP, PAG, and MY; right: demonstrate expression of two unique Projection-TAGs injected into SC_L and SC_S). Scale bar: 1 mm (top), 100 μm (bottom). **c**, Tracing the brain-wide axon collaterals of MOp neurons. Screenshots acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/100141563)[125] were displayed on the left and a summary diagram of the anatomical hierarchy of axonal projections from MOp to the seven target regions was shown on the right. **d**, Tracing the brain-wide axon collaterals of SSp neurons. Images acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/112935169)[125]. were displayed on the left and a summary diagram of the anatomical hierarchy of axonal projections from SSp to the seven target regions was shown on the right. **e** and **f**, Heatmap showing the number of DE genes (FDR<0.05, e) and DA peaks (FDR<0.05, f) between any two projections.

Extended Data Fig. 10



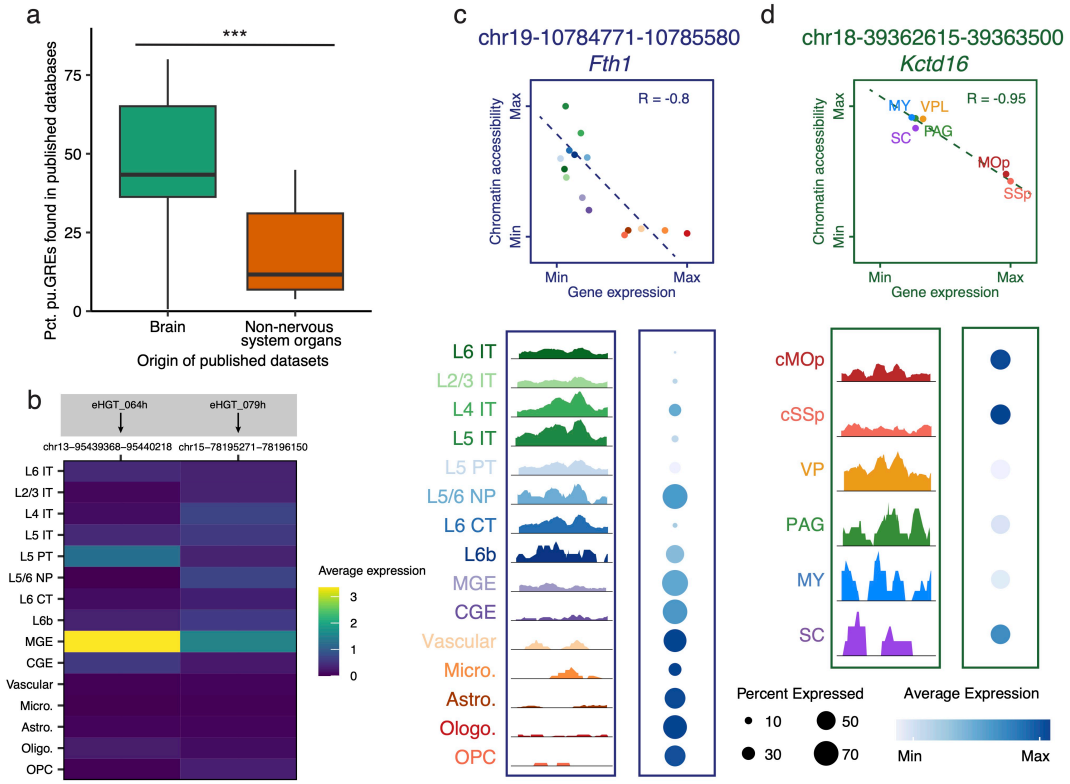
1481

1482

Extended Data Fig. 10 – related to Fig. 3

a, Distribution of projection feature to VP and/or other ET targets (defined in Fig. 3h-j) in Projection-TAG+ nuclei in each subtype.. **b**, A single neuron (MouseLight AA0923) whose cell body reside in the MOp L5 and axonal end points terminating in thalamus, PAG, MY, an axon follow through the corticospinal tract. Screenshots were acquired from the Janelia MouseLight data portal (http://ml-neuronbrowser.janelia.org)[69].

Extended Data Fig. 11

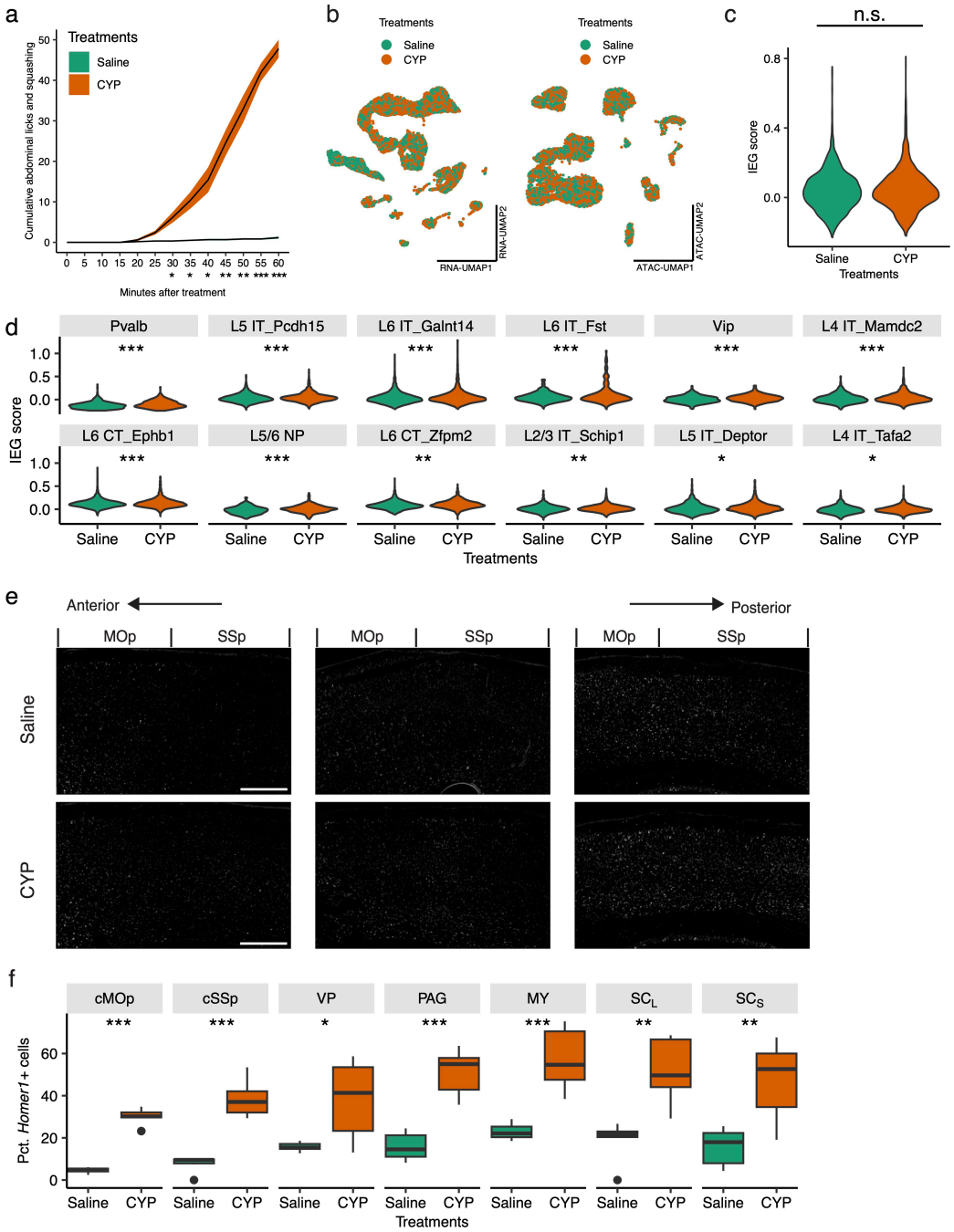


1488

Extended Data Fig. 11 – related to Fig. 4

a, Box plot showing the overlapping of putative GREs identified from this study to the putative GREs identified by studies from the ENCODE consortium[83,126]. Overlap is significantly higher with studies of the brain compared to those of non-brain organs (***p = 8.26e-05, t-test). Brain studies: ENCSR422CGZ, ENCSR694ZBB, ENCFF426KUB, and ENCSR410XMU. Studies of organs that are not the nervous system: ENCSR052AZT (stomach), ENCSR100TMZ (lung), ENCSR296UQL (liver), and ENCSR180OIF (limb). **b**, Heatmap showing the chromatin accessibility of two celltype-specific pu.Enhancers in individual Transcriptional subtypes. The peaks of the two pu.Enhancers overlap with the genomic coordinates liftover from two functional enhancers previously identified based on human snATAC-seq data[77]. Labels in the grey boxes show the functional enhancers and the cell types in which the enhancers exhibited highest activity in experimental validation from the previous study. **c**, Top: Scatter plot showing the correlation between the accessibility of peak chr19-10,784,771-10,785,580 and the expression of gene *Fth1* in individual transcriptional subtypes (chromatin accessibility and gene expression are normalized to their max values). The dotted line represents the line of best fit. bottom left: chromatin accessibility at the genomic locus of chr19-10,784,771-10,785,580, displayed as the average fraction of transposase-sensitive fragments per nucleus in each subtype (grouped by 50 bins per displayed genomic region). Accessibility at each locus (y axis) is scaled to the max value across all cell types. Bottom right: expression of *Fth1* in each celltype. Colors indicate Transcriptional subtypes. **d**, Top: Scatter plot showing the correlation between the accessibility of peak chr18-39,363,615-39,363,500 and the expression of the gene *Kctd16* in nuclei positive for individual TAGs (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr18-39,363,615-39,363,500, displayed as the average fraction of transposase-sensitive fragments per nucleus in each projection (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Kctd16* in each projection. Colors indicate nuclei positive for individual TAGs. Nuclei positive for SC_L- and SC_S-TAGs are combined for visualization due to low cell number.

Extended Data Fig. 12



1510

Extended Data Fig. 12 – related to Fig. 5

a, Cumulative scores of quantified behaviors between mice treated with saline and CYP. CYP treatment resulted in significant increases in the spontaneous behaviors like abdominal licking and squashing behaviors compared to saline treatment. Ribbons show SEM. ** p = 0.0034, one-way ANOVA. p = NaN (0 min), NaN (5 min), NaN (10 min), NaN (15 min), 0.25 (20 min), 0.07 (25 min), *0.05 (30 min), *0.04 (35 min), *0.04 (40 min), **0.001 (45 min), **0.006 (50 min), ***0.0005 (55 min), ***0.0003 (60 min), respectively, post-hoc one-sided t-test, n=5-6 mice/group. **b**, RNA-UMAP (left) and ATAC-UMAP (right) of 5,000 downsampled nuclei from each treatment group, colored by treatments. **c**, Violin plot showing IEG scores of all snRNA-seq nuclei from CYP and saline. p = 0.16. **d**, Violin plots showing IEG score of nuclei between treatments from CYP-activated transcriptional subtypes. ***p = 2.6e-7 Pvalb, ***p = 6.2e-7 L5 IT_Pcdh15, ***p = 1.3e-6 L6 IT_Galnt14, ***p = 1.5e-6 L6 IT_Fst, ***p = 4.6e-5 Vip, ***p = 1.5e-4 L4 IT_Mamdc2, ***p = 3.8e-4 L6 CT_Ephb1, ***p =3.8e-4 L5/6 NP, **p = 1.3e-3 L6 CT_Zfpm2, **p = 4.6e-3 L2/3 IT_Schip1, *p = 0.02 L5 IT_Deptor, *p = 0.04 L4 IT_Tafa2. **e**, Representative FISH images showing *Fos* expression in select cortex areas (top: 1.33 mm, middle: -0.95 mm, bottom -1.67 mm relative to bregma). Scale bar: 400 µm. **f**, Bar plot showing the percentage of Homer1+ cells in each projection group from FISH analysis. Nine slices from three mice per group, ***p = 6.7e-8 cMOp, ***p = 2.9e-6 cSSp, *p = 0.02 VP, ***p = 1.2e-5 PAG, ***p = 1.9e-4 MY, **p = 0.002 SC_L, **p = 0.006 SC_S.