

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

Received: 18 April 2024

Accepted: 19 May 2025

Published online: 01 July 2025

 Check for updates

Lite Yang  <sup>1,2</sup>, Fang Liu<sup>1</sup>, Hannah Hahm<sup>1</sup>, Takao Okuda  <sup>1</sup>, Xiaoyue Li<sup>1</sup>, Yufen Zhang<sup>1</sup>, Vani Kalyanaraman<sup>1</sup>, Monique R. Heitmeier<sup>1</sup> & Vijay K. Samineni  <sup>1</sup> 

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.

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<sup>23</sup>. 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<sup>4–20</sup>. 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<sup>21–28</sup>. 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 using fluorescence as the projection identifier, have been invaluable in mapping distinct neuronal projections<sup>29–37</sup>. 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<sup>38</sup>, common microscopy equipment in neuroscience labs can reliably detect only 3–4 fluorophores, limiting the number of projections to be examined simultaneously<sup>27,39,40</sup>. 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

<sup>1</sup>Department of Anesthesiology, Washington University Pain Center, Washington University School of Medicine, St. Louis, MO, USA. <sup>2</sup>Neuroscience Graduate Program, Division of Biology & Biomedical Sciences, Washington University School of Medicine, St. Louis, MO, USA.  e-mail: [vksamineni@wustl.edu](mailto:vksamineni@wustl.edu)

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<sup>41,42</sup>. Methods, such as retro-seq and epi-retro-seq, characterized the molecular features of projection neurons<sup>22,26,28,43-45</sup>, 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<sup>46-50</sup>. 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<sup>47,51</sup>. 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<sup>52,53</sup>. 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 (Supplementary 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 (Supplementary 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 (Supplementary Fig. 1e). Sun1-GFP labels the nuclear membrane without interfering with the chromosome structure and has been employed in numerous epigenetic studies (Supplementary Fig. 1d)<sup>44,54</sup>. To further enhance Projection-TAG functionality, we developed a protocol for photobleaching Sun1-GFP fluorescence (Supplementary Fig. 2f, h) and created a version of Projection-TAGs expressing a red fluorophore oScarlet (Methods). Third, to enable projection tracing using various commercial assays, we cataloged 50 unique 100-nt RNA barcodes<sup>55</sup> (Supplementary Fig. 2a and Supplementary Data 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 TAG 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 (Supplementary 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 (Supplementary 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)<sup>30</sup> 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 perform 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<sup>43,56-58</sup>. 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<sup>27,45,50,59-61</sup>, 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*, we identified seven downstream projection targets of MOp and SSp, including two intratelencephalic (IT) targets (contralateral MOp [cMOp] and contralateral SSp [cSSp]) and five extratelencephalic (ET) targets, which can be further classified into three subcortical targets (ipsilateral ventral posterior nucleus of the thalamus [VP], ipsilateral periaqueductal gray [PAG], ipsilateral medulla [MY]) and two corticospinal targets (lumbar spinal cord [SC<sub>L</sub>], and sacral spinal cord [SC<sub>S</sub>], Fig. 1a and Supplementary Fig. 5a).

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

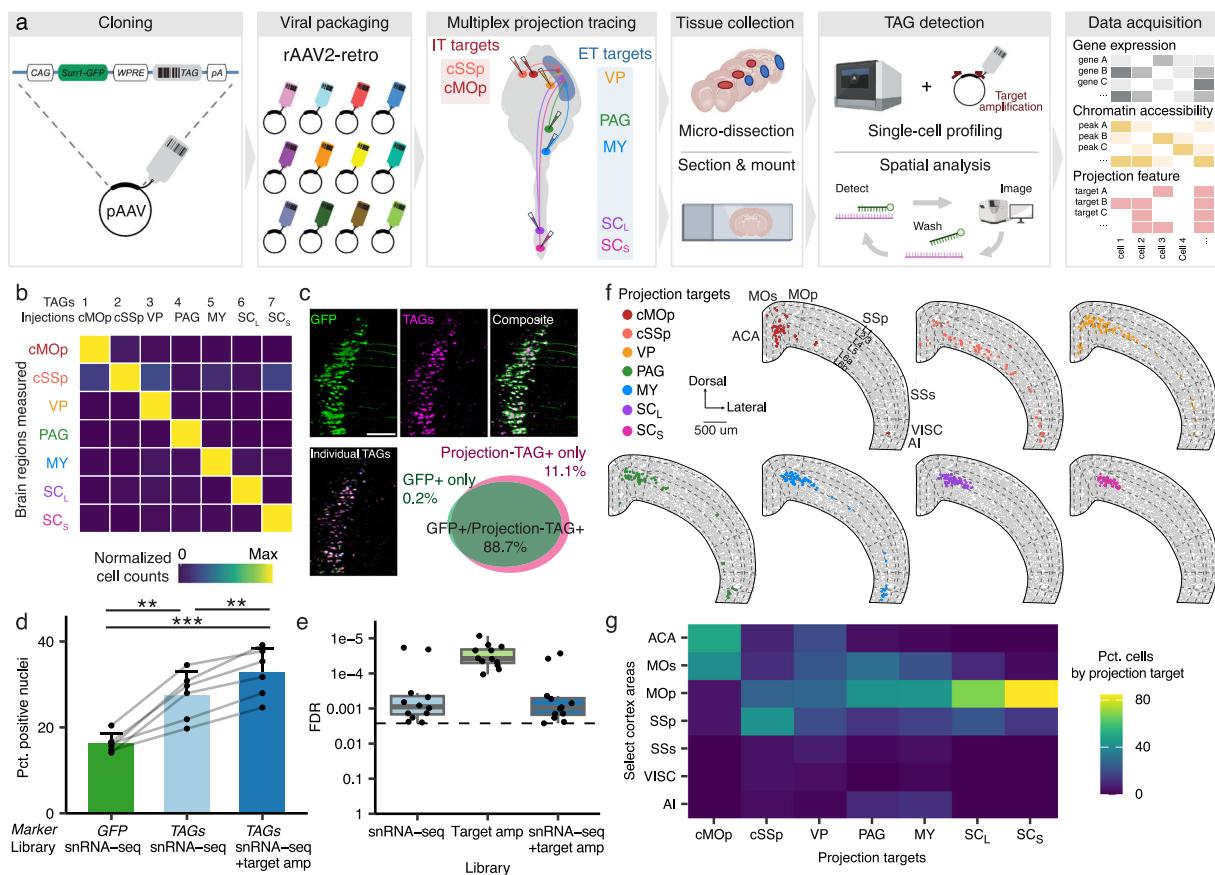
While previous studies using rAAV2-retro have shown that two weeks are typically sufficient for fluorescently labeling cortical projections<sup>21,22,43,44,62</sup>, a multiplex tracing experiment would benefit from understanding the temporal kinetics and stability of cargo gene expression across different cortical projections, which remains largely elusive. We thus investigated the *Projection-TAG* expression in two cortical projections with notably long axons that may represent the upper limit of waiting time: MY, one of the longest cortical projections in the brain, and SC<sub>S</sub>, one of the longest cortical projections in the central nervous system. We injected a *Projection-TAG* AAV into either MY or SC<sub>S</sub> and quantified the *Projection-TAG* expression over time in the cortex using qPCR (Supplementary Fig. 3a). *Projection-TAG* expression is detectable as soon as one week after injection and increased at week two, consistent with previous reports. However, the

expression *MY-TAG* increased rapidly and reached its peak at week three, whereas the expression of *SC<sub>S</sub>-TAG* increased more steadily and peaked at around week five. In both cases, expression plateaued until week ten, the endpoint of our study. While the initial peak timing may vary across projections, our results suggest that the stable *Projection-TAG* expression after peaking provides a flexible time window for synchronizing peak expression in different projections and coupling *Projection-TAGs* with other experimental paradigms.

To make sure *Projection-TAGs* can be unbiasedly used for multiplex projection tracing, we next investigated if significant viral competition exists among *Projection-TAG* AAVs that may reduce the tracing efficiency. Prior experiments showed that, in the medial part of the posterior parietal association area of the cortex (PTLp), more than 50% of neurons projecting to PAG also project to VP. We hypothesized that the competition between *Projection-TAG* AAVs would reduce tracing efficiency in mice receiving both VP and PAG injections compared to mice receiving only one of the injections. However, neither *VP*- nor *PAG-TAG* cell counts are significantly different from mice

receiving either injection than mice receiving both injections, regardless of whether the injections were made simultaneously or at different time (Supplementary Fig. 3d–f). Additionally, *VP*- and *PAG-TAG* UMI counts are not significantly different in most snRNA-seq libraries from nuclei expressing either *TAGs* than nuclei co-expressing both *TAGs* (Supplementary Fig. 3g, h). These results suggest that the *Projection-TAG* expression is largely unaffected by viral competition and thus can be safely used for multiplex projection tracing.

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



**Fig. 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 Sun1-GFP and *Projection-TAGs* in RNA-FISH. Quantification shows the overlap between cells positive for GFP and/or *Projection-TAGs*. FISH probes uniquely targeting each *Projection-TAG* (*TAGs* 1–7) were visualized in separate imaging channels, shown as individual *TAGs*. Their signals were aggregated post-hoc, shown as *TAGs*. Scale bar = 100  $\mu$ m. **d**, Bar plot showing the percentage of nuclei expressing select marker genes (> 0 UMIs for snRNA-seq, > cutoffs for target amplification, see Methods) in 10 snRNA-seq libraries prepared by FACS. 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 (on a reverse log scale) of *Projection-TAGs* due to ambient RNA contamination in all 12 snRNA-seq libraries. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1–Q3). The median is a gray line inside each box. Dotted line shows the highest FDR due to ambient RNA contamination at 0.0026 among all libraries. **f**, Spatial map of *Projection-TAG*-positive 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*. Gray 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.

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 (on a reverse log scale) of *Projection-TAGs* due to ambient RNA contamination in all 12 snRNA-seq libraries. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1–Q3). The median is a gray line inside each box. Dotted line shows the highest FDR due to ambient RNA contamination at 0.0026 among all libraries. **f**, Spatial map of *Projection-TAG*-positive 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*. Gray 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.

**Detection of *Projection-TAGs* with high specificity and efficiency**  
 Multiplex projection tracing relies on the demultiplexing of individual *Projection-TAGs* with various detection assays. In this section, we examined the specificity and efficiency of *Projection-TAG* detection in multiplexed FISH and snRNA-seq and highlighted some considerations that may potentially affect the interpretation of *Projection-TAG* experimental results.

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

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

To examine the efficiency of *Projection-TAG* detection, we performed snRNA-seq on FACS sorted nuclei, in which  $>95\%$  of the nuclei are GFP<sup>+</sup>, from mice receiving *Projection-TAG* AAV injections into the projection targets of the cortex (Fig. 1a). We identified significantly more *Projection-TAG*<sup>+</sup> nuclei ( $32.8 \pm 5.7\%$  with target amplification,  $27.4 \pm 5.6\%$  in snRNA-seq library alone) than Sun1-GFP<sup>+</sup> nuclei (Fig. 1d). To investigate the specificity of *Projection-TAG* detection, we assessed the *Projection-TAG* mismatching in snRNA-seq. *Projection-TAGs* used in experiments are highly expressed ( $75.9 \pm 6.6$  percentile by expression ranks among all genes) while *Projection-TAGs* not used yielded zero counts in both snRNA-seq library and target amplification. We also assessed the false positive rate of *Projection-TAG* detection due to ambient RNA contamination and clustering, and annotation errors. *Projection-TAGs* were detected in only  $0.1 \pm 0.08\%$  of the empty droplets per library (Fig. 1e) and 0.32% of the non-neuronal cells. Target

amplification marginally increased the average *Projection-TAG* detection in empty droplets and non-neuronal cells to 0.104% and 0.34%, respectively. These data suggest that in snRNA-seq, *Projection-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).

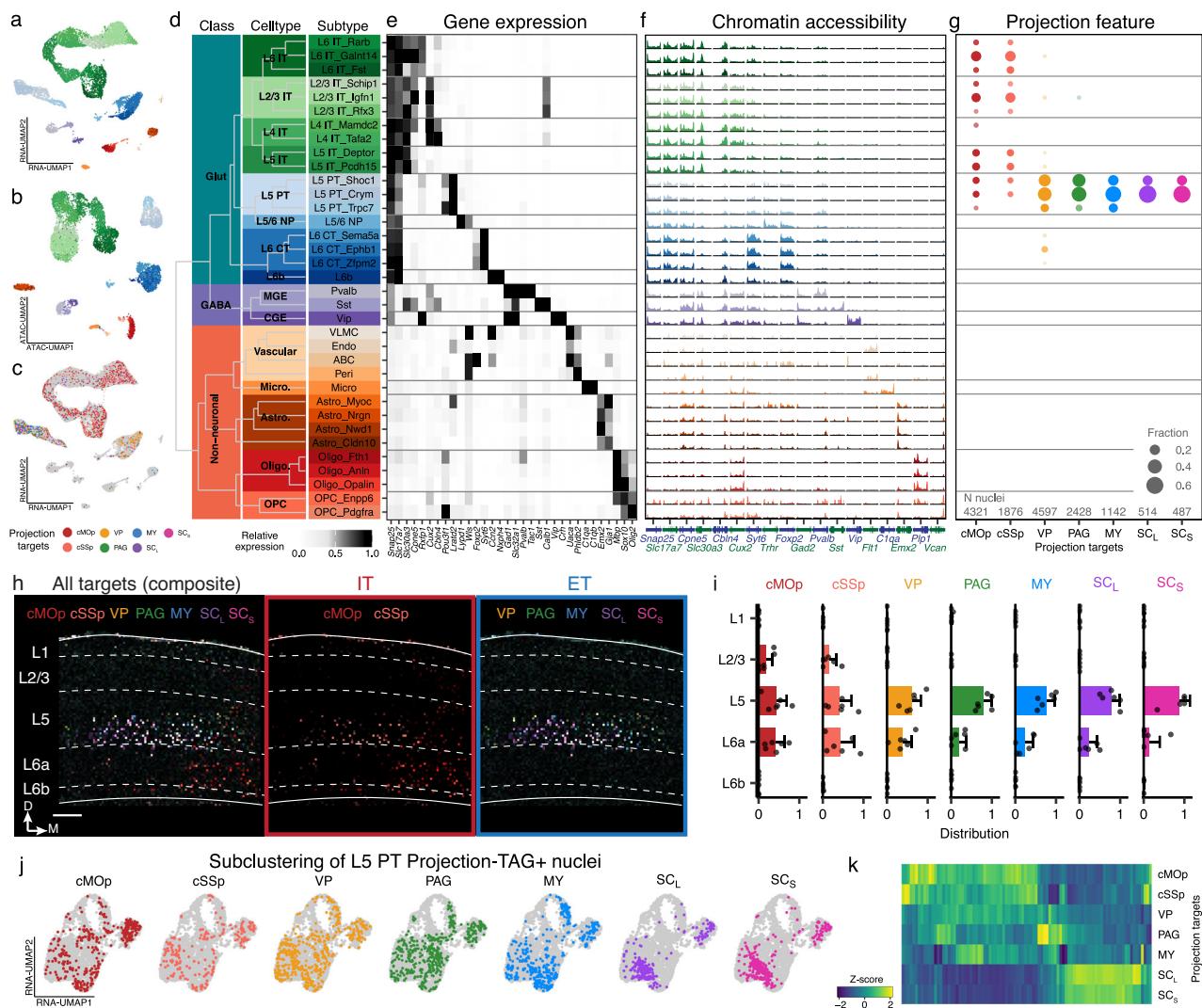
However, one should be mindful of several limitations when interpreting the *Projection-TAG* results. First, while the *Projection-TAG* detection in multiplexed FISH assay is highly efficient (only 0.2% GFP<sup>+</sup> cells do not express *Projection-TAGs*), the false negative rate of *Projection-TAG* detection is nontrivial in snRNA-seq ( $67.2\% \pm 5.7\%$  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 (Supplementary Fig. 4f), likely due to variations of stereotaxic injections, and the *Projection-TAG* expression is different across projection pathways (Supplementary Fig. 4g). However, the exact *TAGs* used to trace each projection pathway did not significantly affect their expression in most cases (Supplementary 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<sup>26,63</sup>, 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 (Supplementary Fig. 6c). We observed that neurons projecting to different targets are enriched in distinct areas across the cortex (Fig. 1f, g). For example, cMO<sup>+</sup> 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<sub>L</sub>- and SC<sub>S</sub>-projecting neurons are highly enriched around MOp and SSp. The spatial distribution of projection neurons is largely consistent with previous reports<sup>23,64–66</sup>. It has been reported that thalamus- and MY-projecting neurons in the anterior lateral motor cortex exhibit distinct spatial distribution in layer 5 (L5)<sup>26,67</sup>. 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 (Supplementary 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 they are particularly rich in projection neurons among all cortex areas analyzed and are highly enriched with neurons projecting to each of the seven targets examined (Supplementary 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, Supplementary 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 3621



**Fig. 2 | A multi-modal single-cell atlas of mouse cortex.** **a** Uniform manifold approximation and projection (UMAP)<sup>1</sup> 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 gray and displayed as the background. For each projection target, 500 *Projection-TAG*-nuclei (or all nuclei if total nuclei count is less than 500) are downsampled 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, cell type, 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 cell-type-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-TAG*s. Scale bar = 150  $\mu$ m. Left: composite image of all FISH channels, middle and right: images of a subset of *Projection-TAG*s 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 Data 4.

genes per nucleus (Supplementary Fig. 7b and Supplementary Data 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 (Supplementary 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 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<sup>27,43,60</sup>. Following this naming convention, we assigned glutamatergic cell-types 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 cell types based on developmental origins: medial ganglionic

eminence (MGE) neurons and caudal ganglionic eminence (CGE) neurons. We also identified five non-neuronal cell types: 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 (Supplementary Fig. 7e, f). While we observed some variability of celltype distribution across libraries, likely due to different nuclear preparation approaches (Supplementary Fig. 7d), the gene expression profiles of individual subtypes are largely consistent with previous reports<sup>15,26,61</sup> (Supplementary Fig. 7g and Supplementary Data 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 (Supplementary Fig. 8a)<sup>68</sup>. These snATAC-seq fragments captured the open chromatin regions of the genome as they followed the expected nucleosomal size distribution (Supplementary 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  $87.7 \pm 18.0\%$  nuclei in each snATAC-seq cluster assigned to the same transcriptional clusters as those made when independently analyzed for snRNA-seq data (Fig. 2b and Supplementary 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 Data 3). Consequently, we grouped snATAC-seq nuclei by transcriptionally defined cell types 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 and Supplementary 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 (Supplementary 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<sub>L</sub> and SC<sub>S</sub> projecting neurons enriched of L5 PT subtypes but are under-represented 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<sup>26,63</sup>.

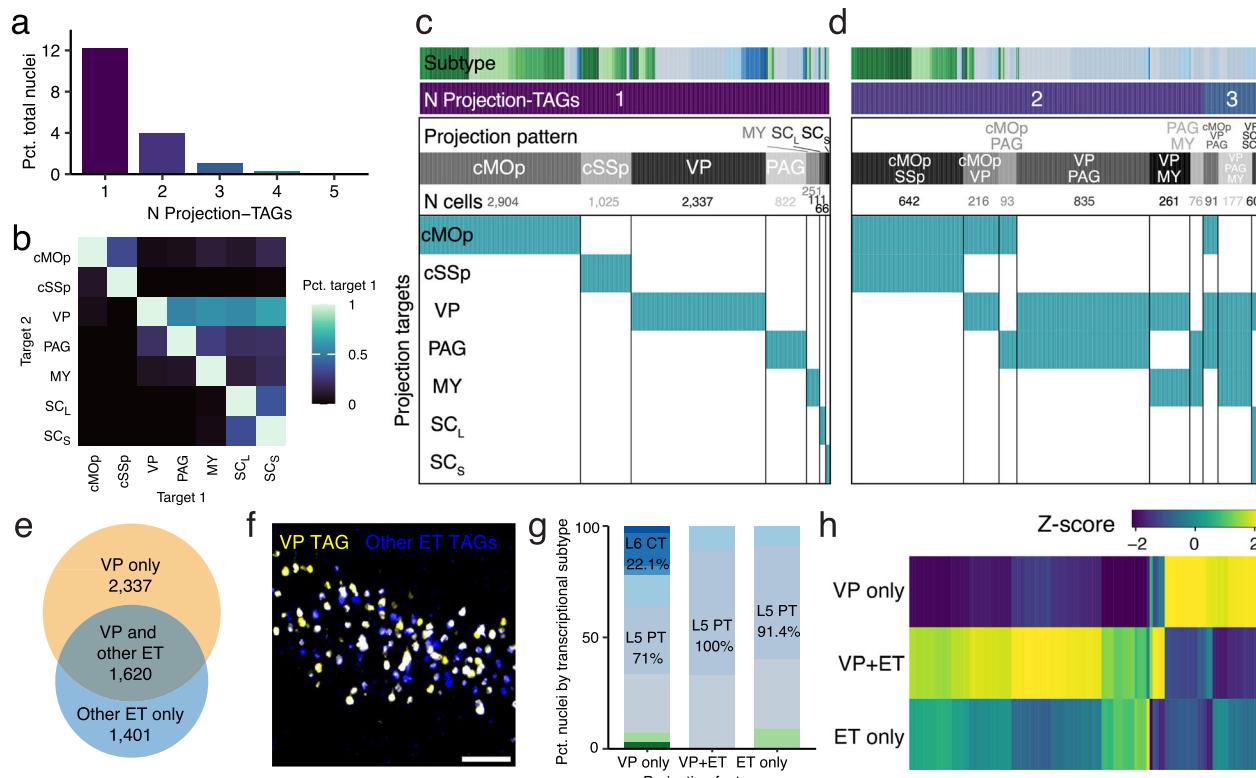
As transcriptionally defined L5 PT neurons are enriched of nuclei positive for *Projection-TAGs* of all seven projection targets, we next

asked whether L5 PT neurons projecting to different targets are transcriptionally distinct. Subclustering of 4,168 L5 PT *Projection-TAG*<sup>+</sup> 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 and Supplementary Data 4). *Hpgd* (*hydroxyprostaglandin dehydrogenase 15*) and *Slco2a1* (*solute carrier organic anion transporter family, member 2a1*) have been reported as marker genes for thalamus- and MY-projecting neurons, respectively<sup>26</sup>. Interestingly, we observed an expression gradient of those genes across L5 PT neurons projecting to different targets (Supplementary 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. Projection-TAGs enable multiplex projection tracing in single animals, thus allowing the examination of 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*<sup>+</sup> 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 (Supplementary Fig. 9a) with each other, with 16.8% *MOp-TAG*<sup>+</sup> nuclei or 36.8% *SSp-TAG*<sup>+</sup> 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. There are 56.3–66.7% nuclei labeled with other *ET-TAGs* (PAG, MY, SC<sub>L</sub>, and SC<sub>S</sub>) that are also co-labeled with *VP-TAG*. Additionally, PAG and MY also highly significantly overlap, as well as SC<sub>L</sub> and SC<sub>S</sub>. 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 (Supplementary Fig 9b). These observations are highly correlated with the anatomical hierarchy of axonal projections of the cortex previously demonstrated by anterograde bulk tracing techniques (Supplementary 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 (Supplementary 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 5358 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<sub>L</sub>, and SC<sub>S</sub>), 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 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 (Supplementary Fig. 10a, b)<sup>69</sup> 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 1971 differentially expressed (DE) genes and 2737 differentially accessible (DA) peaks in neurons with



**Fig. 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, 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 were shown on the heatmaps. **e** Venn diagram showing the overlap between *VP-TAG* only, *VP* and other *ET*, and Other *ET* only.

expressing and other *ET-TAG*-expressing snRNA-seq nuclei. Other *ET-TAGs* include *Projection-TAGs* injected into PAG, MY, SC<sub>L</sub>, and SC<sub>S</sub>. **f** Representative FISH image showing MOOp cells labeled with *VP*- and other *ET-TAGs*. Scale bar = 100  $\mu$ m. Experiments were repeated in three mice. **g** Distribution of transcriptional subtypes in each projection group. Transcriptional subtypes with <60 nuclei from each projection feature group were excluded, and cell types 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 Data 5.

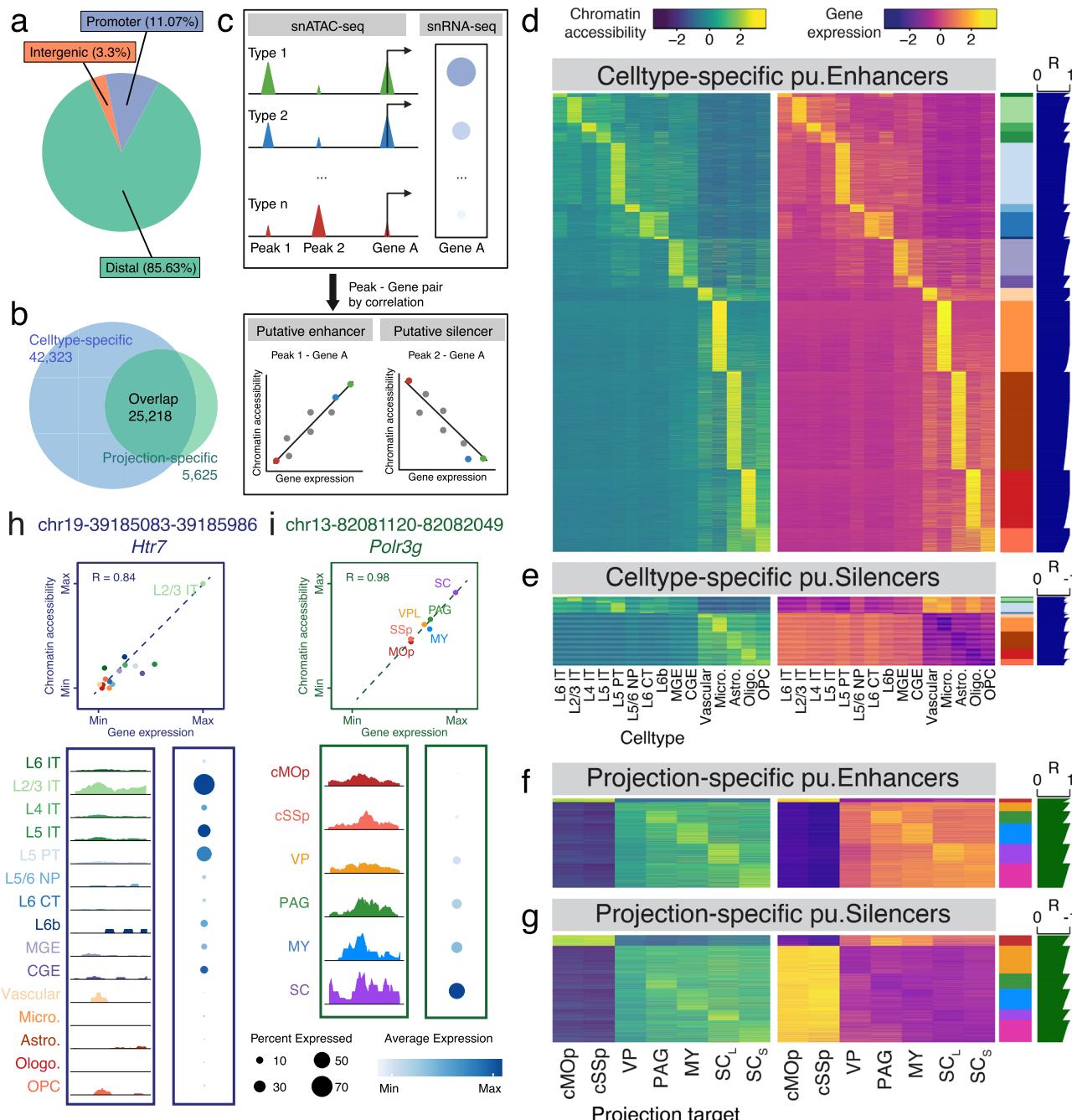
single projections compared to those with multiple projections (Fig. 3h and Supplementary Data 5).

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

While the gene expression program in the diverse MOOp 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 insufficiently investigated<sup>22,44,60,70</sup>. The genomic regulatory elements (GREs), mostly identified within the non-coding region of the genome and acting in a cell-type-specific and tissue-specific manner, fine-tune the expression level of the regulated genes<sup>71</sup>. 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<sup>72,73</sup>. 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<sup>74–79</sup>. Though recent progress has been made in characterizing the landscape of GREs for the projection neurons, the current experimental workflow is tedious and painstaking 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<sup>22,44</sup>. We

next performed high-throughput analysis on *Projection-TAG*-positive nuclei for identifying projection-specific GREs.

To this end, we integrated the snRNA-seq and snATAC-seq data from MOOp and SSp to produce a unified, multi-modal cell census. We first identified 166,540 peaks when aggregated across all snATAC-seq libraries. Peaks are mapped to one of the three categories: promoter regions (-1000 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 (>200 kb upstream or downstream of TSS, excluding gene body). 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). Among the differentially accessible peaks, we found 43.9% of them in individual transcriptional cell types and/or neurons projecting to different targets (Fig. 4b). To identify putative GREs and their correspondence to genes they may regulate, we first identified 67,541 D peaks ( $\log_2 FC > 1$ , FDR < 0.05) in transcriptional cell types 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 cell types 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 5 M 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



**Fig. 4 | Genomic regulatory networks of cortical cell types using Projection-TAGs.** **a** Pie chart categorizing snATAC-seq peaks based on their genomic locations. **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–g** Heatmaps showing celltype-specific pu. Enhancers (**d**) and pu. Silencers (**e**), projection-specific puEnhancers (**f**) and pu.Silencers (**g**) and their regulated genes. Left heatmaps show the average accessibility of peaks (row) in individual defined populations (column) in snATAC-seq data, and right heatmaps show the average expression of genes (row) in individual populations (column) in snRNA-seq data. Color bars indicate the population in which the peak and gene have highest accessibility and expression, and the bar plot shows the Pearson correlation coefficient (R) between the peak accessibility and gene expression across defined populations. **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 cell types (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 cell type (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 cell type. **i** Top: Scatter plot showing 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 for each *Projection-TAG* (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Polr3g* in nuclei positive for each *Projection-TAG*. Colors indicate the injection site of the corresponding TAG. Nuclei positive for *SC<sub>L</sub>*- and *SC<sub>S</sub>*-TAGs are combined for visualization due to low cell number.

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 2739 pu.Silencer-gene pairs that are celltype-specific (Fig. 4d, e and Supplementary Data 6), and 3545 pu. Enhancer-gene pairs and 4200 pu.Silencer-gene pairs that are projection-specific (Fig. 4f, g and Supplementary Data 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<sup>80</sup>. 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 (Supplementary Fig. 11a)<sup>81</sup>. Furthermore, our analysis reveals four experimentally validated functional enhancers that coincide with peaks identified in our snATAC-seq data<sup>75</sup>. Notably, two of these validated enhancers overlap with the cell-type-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 (Supplementary Fig. 11b).

Among identified putative GREs, the peak 39,185,083–39,185,986 on chromosome 19 is ~3.2 Mbp upstream of the TSS of *Htr7* (*5-hydroxytryptamine [serotonin] receptor 7*) 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* (*polymerase [RNA] III [DNA directed] polypeptide G*) 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 cell-type-specific and projection-specific pu.Silencers, respectively. While accessible in the genome, they may reduce the expression of *Fth1* (*ferritin heavy polypeptide 1*) in the IT neurons and *Kctd16* (*potassium channel tetramerisation domain containing 16*) in ET-projecting neurons, respectively (Supplementary 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

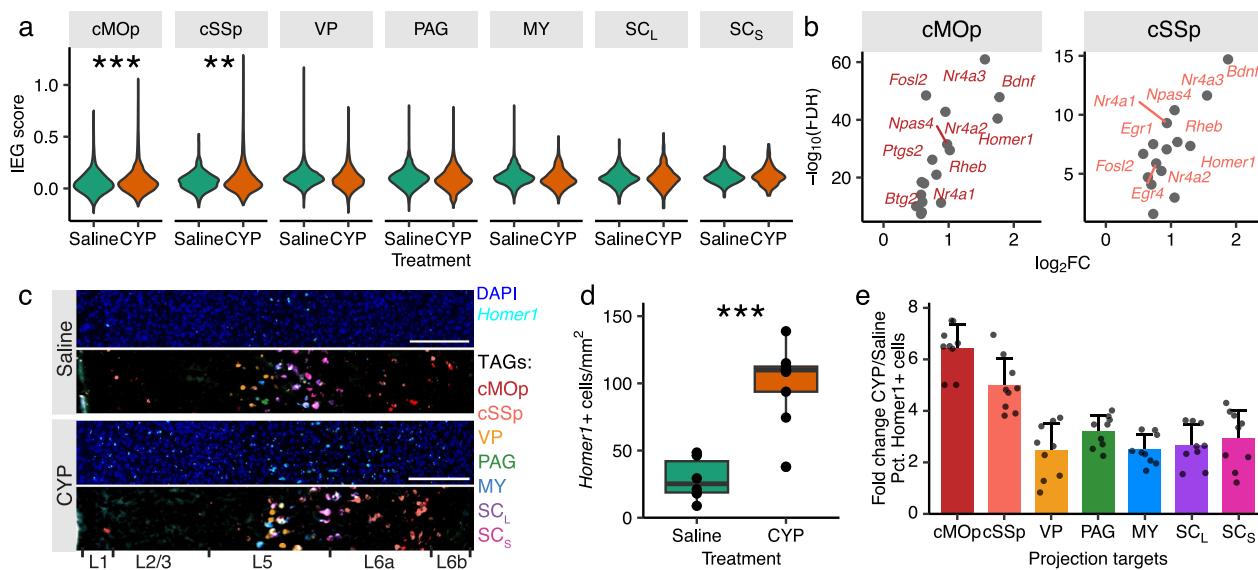
Projection-TAGs allow us to elucidate the spatial, gene expression, and chromatin accessibility profiles of diverse projection neurons at single-cell resolution. As Projection-TAG AAVs mediate stable transgene expression with minimal immune response, we sought to leverage it to obtain a snapshot of activity-dependent 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<sup>82–87</sup>, we wanted to gain insight into the gene expression changes in the cell types and their projection 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 injected saline as the control, followed by combinatorial snRNA-seq and snATAC-seq of the MOp and SSp 30 min after the stimulus, at which we observed significant spontaneous behaviors (Supplementary 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 (Supplementary Fig. 12b). To pinpoint the cell populations that are acutely activated by visceral pain, we next conducted Act-seq<sup>88–90</sup>, 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 all nuclei between treatments (Supplementary Fig. 12c), CYP significantly activated 12 transcriptional subtypes, including seven IT neuronal subtypes and two GABAergic neuronal subtypes (Supplementary Fig. 12d), suggesting their preferential recruitment following the acute visceral pain stimulus. Consistently, 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<sup>87,91–94</sup>. 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 ( $\log_2FC > 0.5$  in either cMOp or cSSp, Fig. 5b). Among them, *Nr4a3* (*nuclear receptor subfamily 4, group A, member 3*), *Bdnf* (*brain derived neurotrophic factor*), *Rheb* (*Ras homolog enriched in brain*), and *Homer1* (*homer scaffolding protein 1*) showed higher fold change ( $\log_2FC > 1$ ) in both cMOp- and cSSp-projecting neurons. Interestingly, CYP did not significantly induce *Fos* (*FBF osteosarcoma oncogene*) expression (Fig. 5b and Supplementary Fig. 12e). *Homer1* expression is directly linked to neuronal activity and synaptic plasticity<sup>95–98</sup>. 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 (Supplementary 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 show that Projection-TAGs could be readily implemented with transcriptional analysis, such as 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 multimodal 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 workflows 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



**Fig. 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 cMOOp,  $**p = 0.004$  for cSSp,  $p = 1$  for VP,  $p = 0.99$  for PAG,  $p = 0.99$  for MY,  $p = 0.4$  for SC<sub>L</sub>,  $p = 0.2$  for SC<sub>S</sub>, one-side  $t$ -test with Bonferroni correction). **b** Volcano plots showing significantly induced IEGs ( $\log_2\text{FC} > 0.5$ ) 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:  $100\text{ }\mu\text{m}$ . **d** Quantification of *Homer1*<sup>+</sup> 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). Boxes indicate quartiles, and whiskers are 1.5-times the interquartile range (Q1–Q3). The median is a line inside each box. **e** Percentage of *Homer1*<sup>+</sup> 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.

study, we applied Projection-TAGs to examine the transcriptional and epigenetic landscapes of the cortex using combinatorial snRNA-seq and snATAC-seq. 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 axonal transport and barcoding schemes. Anterograde tracing techniques, namely MAP-seq and its derivatives, such as BRICseq, BARseq, and BARseq2<sup>46,49,50,99</sup>, utilize 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<sup>48</sup> 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 several expert labs. The high replication rate of Sindbis virus results in cellular toxicity, posing a challenge to integrating their use with chronic experimental paradigms.

Retrograde tracing techniques include Projection-seq, MERGE-seq, and Projection-TAGs<sup>53,100</sup>. These techniques utilize a limited number of RNA barcodes with known sequences, which act as the projection barcode/identifier. A retrograde AAV expressing a unique BC is injected into each of the target regions, which retrogradely labels projection neurons in the source region. Multiplexed projection tracing is achieved by assaying the barcodes present in the source region, which can be simply read out by scRNA-seq and spatial assays. Due to the retrograde tracing scheme, Projection-TAGs and similar techniques can achieve high spatial resolution in the source region, but not necessarily the target regions. The surgery procedures are relatively complicated (one injection into each of the target regions), which may limit the multiplexing capacity, and the accuracy of targeting each target region may introduce technical variations that confound the quantitative measurement of the projection strength. Surgical targeting (one injection into each of the target regions) may introduce technical variations that confound the quantitative measurement of the projection strength and might limit the multiplexing capacity. Despite these technical challenges, barcode detection is relatively simple and flexible and can be achieved using various commercial assays. Projection-TAGs enable multiplex neuroanatomical studies and high-throughput multi-modal profiling of projection neurons. AAVs have minimal cellular toxicity, which is ideal for incorporating Projection-TAGs with additional experimental paradigms for studying activity-dependent recruitment of distinct cell populations and projections in the stimulus of interest.

#### Consideration for projection-TAG experimental design and result interpretation

Projection-TAGs allow multiplex projection tracing and multi-modal profiling of projection neurons. In retrograde viral tracing experiments, the cargo gene expression in a single cell depends on a series of biological processes such as viral attachment and internalization,

transduction and trafficking to the cell body, and escape and entry into the nucleus<sup>101</sup>. rAAV2-retro has been widely adopted in neuroscience laboratories. Projection-TAG rAAV2-retro performance would be similar to other rAAV2-retro viruses, and prior experience using rAAV2-retro may be used to guide the planning of Projection-TAG multiplex tracing experiment. Efficient and accurate labeling of Projection-TAGs relies on successful targeting of desired brain regions, transduction efficacy of rAAV2-retro, titer (Projection-TAG AAVs with the titer of  $\sim 2e+12$  vg/ml label  $14.7 \pm 8.4\%$  of MOP cells, compared  $3.7 \pm 3.0\%$  by those with the titer of  $\sim 5e+11$  vg/ml) and volume of the viruses injected. The high correlation between *Projection-TAG* and GFP expressed using the Projection-TAG AAVs provides a simple strategy for investigators to estimate retrograde tracing efficiency and accuracy of target labeling, which are essential for a successful Projection-TAG experiment. While Projection-TAG plasmids have the potential to be flexibly packaged into other AAV serotypes, the transduction efficiency and tissue/cell-type tropisms related to AAV serotypes may lead to false negatives and should be considered while designing the experiments and interpreting the results. While Projection-TAG AAVs can be multiplexed and used interchangeably (Supplementary Figs. 3d–h, 4h), AAV transduction efficiency and *Projection-TAG* expression may not be uniform across projection pathways (Supplementary Fig. 4f, g), which may introduce confounds to the quantitative measurement of the projection strength.

After the tracing experiments, *Projection-TAG* detection and multi-modal profiling can be achieved using various commercial assays. *Projection-TAG* detection is specific and sensitive in single-cell sequencing and multiplexed FISH, as previously described (Fig. 1). Multiplexed FISH is superior at the *Projection-TAG* detection efficiency, but is labor-intensive and difficult to scale up. Sun1-GFP expressed by *Projection-TAG* AAVs reliably labels both whole cells and nuclear suspensions, enabling enrichment of projection neurons for unbiased single-cell and single-nucleus transcriptional and epigenetic profiling in a high-throughput manner. Enzyme-free nuclear extraction works for various tissue samples (fresh, frozen, or fixed) and introduces minimal dissociation-induced transcriptional stress response, and thus is ideal for activity-dependent circuit mapping. Whole-cell sequencing has the ability to sequence both cytoplasmic and nuclear transcripts, advantageous for recovering medium- to low-expressing transcripts and may potentially increase *Projection-TAG* detection. Sequencing involves many biochemical processes that may confound the *Projection-TAG* detection. Artifacts such as ambient RNA contamination, doublets, and analytical errors may lead to false positives, which can be ameliorated experimentally (e.g., to reduce ambient RNA contamination by adding additional wash steps and incorporating FACS) and computationally (e.g., ambient RNA correction, doublet removal, and setting up appropriate FDR for the analysis). *Projection-TAG* detection rate is limited by the sequencing assays (e.g., efficiency of reverse transcription and cDNA capture) and sequencing setups (Supplementary Fig. 4), which may pose an upper limit for *Projection-TAG* detection and lead to false negatives. Consequently, lack of *Projection-TAG* expression in a cell should not be simply interpreted as lack of projection, and the neurons projecting to each target may be under-reported in snRNA-seq, which may result in an even higher degree of underestimation of the axonal collaterals. *Projection-TAG* detection using sequencing assays with higher transcript capture efficiency and multiplexed FISH may circumvent false negatives.

## Future directions of projection-TAGs

Projection-TAGs could potentially be compatible with additional commercial platforms, such as high-throughput spatial transcriptomics platforms like Xenium and Visium, that would allow spatially-resolved investigation of diverse cell types and distinct anatomical organization of the projection neurons. Another potential application of Projection-TAGs includes connecting neuronal activity

to the distinct projections by integrating the oScarlet version with *in vitro* and *in vivo* calcium imaging experiments. The cell type and projection information from imaging-based molecular assays can be integrated with the real-time neuronal activity information from the calcium imaging experiments via post-hoc imaging alignment and registration<sup>91,102-104</sup>.

The number of projections that can be labeled with Projection-TAG AAVs is not inherently constrained, and Projection-TAGs could be easily scalable as we have screened 50 BCs that could potentially be packaged to increase multiplexing of projection tagging. It has been well appreciated that AAV serotypes have distinct tropism and selective labeling of the brain nuclei. With the discovery of capsid selection using directed evolution<sup>105-107</sup> and advances in sequencing techniques and computational tools to optimize exogenous transcript detection<sup>108,109</sup>, we anticipate that the improved retrograde labeling efficiency of viruses will enhance wider applications of Projection-TAGs. Given its flexibility, usability, and compatibility with commercial platforms, we envision that Projection-TAGs can be readily applied to study diverse projection types in the central and peripheral nervous system.

## Methods

## Generation of projection-TAGs

We used 100-bp BCs because they allow us to design HCR probes to detect the spatial distribution of BC transcripts, and would improve their detection rate in RNA-seq compared to shorter BCs that are commonly used. We first retrieved 60 previously reported BCs<sup>55</sup>. We then filtered out the BCs that contain the sequence of restriction enzymes (BamHI, Hpal, and NotI). Next, we performed sequence alignment using blastn suite searching in the “Nucleotide Collection (nr/nt)” database and “refseq representative genomes” database against the genomes of human (taxid:9606), mouse (taxid:10090), rat (taxid:10116), and Primates (taxid:9443) and filtered out the ones that showed significant similarities<sup>110</sup>. We then calculated the Euclidean distance between any of the two BCs using DistanceMatrix() function in R package DECIIPHER. We then cataloged and reported the first 50 BCs sorted based on the average Euclidean distances with other BCs.

## Cloning and viral packaging of Projection-TAG AAVs

To generate the backbone of pAAV-CAG-Sun1-GFP-WPRE-pA, we first linearized two previously reported plasmids: pAAV-CAG-H2B-GFP (Addgene, Plasmid #116869)<sup>26</sup> using HindIII and pAAV-Efla-DIO-Sun1GFP-WPRE-pA (Addgene, Plasmid #160141)<sup>31</sup> using Ascl, and end-filled. We then digested them with SpeI and NheI, respectively, followed by gel extraction of the fragments at around 4700 bp (pAAV-CAG backbone) and 2400 bp (Sun1-GFP), respectively. We next generated the plasmid pAAV-CAG-Sun1-GFP-WPRE-pA by ligating the two fragments together. To insert BCs into pAAV-CAG-Sun1-GFP-WPRE-pA, we first generated a gene block fragment for each BC with the following structure (5' overhang-NotI-BC-SV40 polyA-Hpal-3' overhang) and sequence (AAGGAAAAAAgccccgc-100bp sequence of BC-ttcgagcagacatgataagatacattgtatggacaaaccacaactagaatgcgtgaaaaaatgttatttgtgaaatttgtatgtctatgtctttttgttaaccattataagctgcataaaacaaatgttaacACCCTGCCCC). We then digested both pAAV-CAG-Sun1-GFP-WPRE-pA and BC gene blocks with NotI and Hpal, and generated pAAV-CAG-Sun1-GFP-WPRE-BC-pA by ligating the fragments. We generated 12 plasmids, each expressing a unique BC (1–12). We packaged the plasmids into AAV (BrainVTA, China) and generated a set of 12 rAAV2-retro samples, each expressing a unique BC (titer range 1.1–3.5e + 12 vg/ml). To generate a set of 12 plasmids pAAV-CAG-oScarlet-WPRE-BC-pA, each expressing a unique BC, we generated BamHI-oScarlet-WPRE-NotI gene fragment and replaced Sun1-GFP-WPRE fragment in pAAV-CAG-Sun1-GFP-WPRE-BC-pA using BamHI and NotI. All restriction enzymes were purchased from New England

Biolabs, and all plasmids produced in this study are deposited to Addgene: <https://www.addgene.org/browse/article/28247961/>.

### Testing projection-TAGs in HEK cells

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

### Animals

All experiments were conducted in accordance with the National Institute of Health guidelines and with approval from the Animal Care and Use Committee of Washington University School of Medicine. Mice were housed on a 12-h light-dark cycle (6:00 a.m. to 6:00 p.m.) and were allowed free access to food and water. All animals were bred onto C57BL/6J background, and no more than five animals were housed per cage. Female littermates between 8 and 10 weeks old were used for all sequencing and RNA-FISH experiments.

### Stereotaxic surgeries

Mice were given a single subcutaneous injection of 0.5 ml 0.9% sterile sodium chloride and Buprenorphine-SR, 1 h prior to surgery to help rehydrate the mouse after anesthesia. Mice were anesthetized with 1.5–2% isoflurane in an induction chamber using isoflurane/breathing air mix. Once deeply anesthetized, mice were secured in a stereotactic frame (RWD Life Science) where surgical anesthesia was maintained using 2% isoflurane. Mice were kept on a heating pad for the duration of the procedure. Preoperative care included application of sterile eye ointment for lubrication, administration of 1 mL of subcutaneous saline, and surgery-site sterilization with iodine solution. We injected into the seven projection targets of the cortex, each with a Projection-TAG rAAV2-retro expressing a unique TAG, in single animals (Supplementary Data 2). All injections were made using a Nanoject II auto injector (Drummond) with a glass microelectrode at a rate of 1 nL/s, and the needle was held in place for 10 min prior to needle withdrawal. Needles are changed between each Projection-TAG AAV injection to avoid cross-contamination. Stereotaxic surgery, injecting into SC<sub>L</sub> and SC<sub>S</sub>, was performed at week 0. For SC<sub>L</sub> injection, the injections were located 400 µm lateral to the center of the posterior artery (150–300 mm below the dura), and the virus was bilaterally injected between T12 and T13 intravertebrally with 250 µl volume each injection. For SC<sub>S</sub> injection, the dorsal part of the L1 vertebra was gently excised using a high-speed micro-drill (RWD Life Science) to unveil the L6–S1 spinal cord segments. The injections were located 100 µm left-lateral to the center of the posterior artery with a 10-degree left-right tilt (550–580 mm below the dura). Three distinct sites were injected with 150 nL of virus at each site. Following a 2-week convalescent interval, the second stereotaxic surgery was conducted by injecting into cMOp, cSSp, VP, PAG, and MY using the following coordinates relative to Bregma: cMOp (anterior-posterior [AP] −0.52 mm, medial-lateral [ML] +1.81 mm, dorsal-ventral [DV] −0.86 mm, left hemisphere), cSSp (AP −0.52 mm, ML+ 0.75 mm, DV −0.86 mm, left hemisphere), VP

(AP −1.5 mm, ML −1.52 mm, DV −4 mm, right hemisphere), PAG (AP −4.5 mm, ML −0.5 mm, DV −2.83 mm, right hemisphere), and MY (AP −6.2 mm, ML −0.5 mm, DV −5.93 mm, left hemisphere). A small midline dorsal incision was performed to expose the skull. After leveling the head relative to the stereotaxic frame, the specified injection coordinates were used to mark the locations on the skull, and a small hole (~0.5 mm diameter) was drilled for each. 500 nL of virus was injected into each of the targets. After each surgery, the surgery site was sutured, and mice were recovered from anesthesia on a heating pad and then returned to their home cage. Mice were given carprofen (0.05 mg/ml in water) to minimize inflammation and discomfort and monitored for three consecutive days. A step-by-step protocol for stereotaxic surgeries can be found on protocols.io [<https://doi.org/10.17504/protocols.io.5qpv09ejxv4o/v1>].

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

### Mouse model of visceral pain

Stereotaxic surgeries were performed as described above. At week 5 post-viral injections, mice were acclimated for at least 3 days prior to the experiment. Mice were administered an intraperitoneal injection of either cyclophosphamide (Sigma-Aldrich, dissolved and diluted to 40 mg/mL in 0.9% sodium chloride) with a dose of 200 mg/kg or saline (approximately the same volume as CYP). Animals were monitored in their home cage for 30 min before perfusion.

### Tissue processing

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

To collect biological samples for FISH studies, animals were anesthetized with ketamine cocktail, and perfused with DEPC-PBS followed by 4% DEPC-PFA. Brain and spinal cord tissues were dissected and incubated in 4% DEPC-PFA at 4 °C for 6–8 h. Tissues were then incubated in 30% sucrose in 1X DEPC-PBS at 4 °C for 24–48 h until they sank to the bottom of the tube. Tissues were then embedded in OCT and stored at −80 °C before slicing. Tissues were sliced coronally into sections with 30 µm thickness using a cryostat (Leica), and tissue slices were mounted on microscope slides. Slides were stored at −20 °C for at least one hour before moving to −80 °C for long-term storage. Slides containing the brain regions of injections were examined under a fluorescent microscope, and the fluorescent signal from Sun1-GFP was used to confirm the injection sites. As an optional step, the native fluorescence of Sun1-GFP can be photobleached from mouse brain sections by floating slices in 1×PBS with 24 mM NaOH and 4.5% H<sub>2</sub>O<sub>2</sub> and exposing under the UV light (27 total watts, OPPSK) for 30 min at room temperature. Slices were immediately rinsed with DEPC-PBS twice before mounting on a microscope slide.

To collect biological samples for molecular experiments such as qPCR or sequencing, animals were anesthetized with ketamine cocktail, and perfused with ice-cold NMDG-based cutting solution (NMDG 93 mM, KCl 2.5 mM, Na<sub>2</sub>PO<sub>4</sub> 1.25 mM, NaHCO<sub>3</sub> 30 mM, HEPES 20 mM, Glucose 25 mM, Ascorbic acid 5 mM, Tiourea 2 mM, Sodium Pyruvate 3 mM, MgSO<sub>4</sub> 10 mM, CaCl<sub>2</sub> 0.5 mM, N-acetylcysteine 12 mM; pH adjusted to 7.3 with 12 N HCl, and bubbled with 95% O<sub>2</sub> and 5% CO<sub>2</sub>). The spinal cord was dissected, fixed, and sectioned by following the

sample preparation process mentioned above in order to confirm the injection sites of SC<sub>L</sub> and SC<sub>S</sub>. The brain was submerged in ice-cold NMDG-based cutting solution and sliced coronally into sections with 400  $\mu$ m thickness using a Comprestmome (Precisionary, VF-210-0Z). MOOp and SSp samples were prepared by micro-dissecting the respective regions under a microscope and collected into a nucleus-free centrifuge tube placed on dry ice. Samples were stored at -80 °C for long-term storage. A step-by-step protocol for fresh frozen tissue collection for sequencing can be found on protocols.io [<https://doi.org/10.17504/protocols.io.x54v9r94qv3e/v1>].

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

### Multiplexed FISH with HCR

HCR (hybridization chain reaction) v3.0 probes, amplifiers, and reagents were purchased from Molecular Instruments<sup>12</sup>. Multiplexed FISH was performed according to the manual of HCR RNA-FISH with minor modifications. For hybridization, samples were equilibrated in hybridization buffer for 30 min at 37 °C and hybridized with probe sets (16 nM each probe) in hybridization buffer overnight at 37 °C. Samples were washed in probe wash buffer and gradually switched to 5×SSCT (5×SSC, 0.1%Tween-20) at 37 °C. HCR was carried out at room temperature. Samples were equilibrated in amplification buffer for 30 min, and the amplifier hairpins (conjugated to Alexa-488, Alexa-546, Alexa-647, and/or Cyanine 7) were heated to 95 °C and snap-cooled in a dark drawer for 30 min. Hairpins were then mixed and diluted to 0.6 nM each hairpin in amplification buffer before incubated with samples overnight. Samples were washed with 5×SSCT for a total of three times and imaged in 2×SSCT with DAPI. Imaging was carried out as described in the next section. After imaging, samples were washed with 5×SSCT, then HCR probes and hairpins were stripped by incubating with 0.25 U/ul of DNase I (Sigma Aldrich) in 1×DNase incubation buffer for 90 minutes at 37 °C (Supplementary Fig. 2g, i). Samples were washed with 5×SSCT for 5 min for a total of 5 times before the next round of HCR-FISH.

### Imaging

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

### Quantification of projection neurons using QUINT

FISH images were segmented and quantified using the QUINT workflow, which allows for semi-automated quantification of cells in labeled brain regions<sup>13</sup>. For atlas registration, images from different channels were merged and downsampled according to the manual. An XML file was generated using Filebuilder and loaded into QuickNII for linear registration to the Allen Mouse Brain Atlas CCFv3<sup>14,15</sup>. After, user-guided nonlinear refinements were applied to brain sections using VisuAlign. For the segmentation of positive cells, images of individual channels were prepared, and segmentation was performed through the pixel classification and object classification pipelines in ilastik<sup>16</sup>. In pixel classification, models were trained to distinguish signal from background using a small subset of images and applied to the whole dataset. Individual machine-learning algorithms were trained for each channel. Probability maps of signal and background were exported as HDF5 files to be used in object classification. In object classification, models were trained to distinguish objects from non-specific labeling based on features such as size and shape, using a small subset of images applied to the whole dataset.

The performance of ilastik segmentation algorithms was validated against manual segmentation. Ilastik accurately identified 96.8% of manually segmented cells with a 6.6% false positive rate (Supplementary Fig. 6b). There are 2.1% of the segmented objects identified as multiple cells merged into one object, and 1.9% of the segmented objects identified as one cell split into multiple objects. Finally, segmentation and registration files were uploaded to the Quantifier feature in Nutil, which resulted in quantification of cells per brain region according to the reference atlas<sup>17</sup>. To identify cells that were co-labeled by multiple markers, objects positive for each marker were segmented individually, and the x and y coordinates of the center of each object were calculated using ilastik. If the Euclidean distance of two objects from separate channels is less than the average radius of all objects, the two objects are identified as one cell that co-expresses both marker genes. For *Projection-TAG* detection and specificity analysis, cSSp and cMOOp were excluded from this analysis because they are known to contain neurons projecting to other injection sites and may be labeled by TAGs injected into other regions, which could contaminate the final distribution of the TAG.

### Single-nuclei isolation and FACS

Nuclear extraction was performed according to the protocol described previously with minor modifications<sup>18</sup>. Mouse cortical tissues were transferred to a Dounce homogenizer in homogenization buffer (0.25 M sucrose, 25 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 8.0, 5  $\mu$ g/mL actinomycin, 1% BSA, and 0.08 U/ul RNase inhibitor, 0.01% NP40) on ice. Samples were homogenized for 10 strokes with the loose pestle in a total volume of 1 mL, followed by 10 additional strokes with the tight pestle. The tissue homogenate was then passed through a 50  $\mu$ m filter and diluted 1:1 with working solution (50% iodixanol, 25 mM KCl, 5 mM MgCl<sub>2</sub>, and 10 mM Tris-HCl, pH 8.0). Nuclei were layered onto an iodixanol gradient after homogenization and ultracentrifuged as described previously. After ultracentrifugation, nuclei were collected between the 30 and 40% iodixanol layers and diluted with resuspension buffer (1xPBS with 1% BSA, and 0.08 U/ul RNase inhibitor). Nuclei were centrifuged at 500  $\times$  g for 10 min at 4 °C and resuspended in resuspension buffer with 5 ng/ul of 7-AAD. For FACS, gates on GFP and 7AAD were set using tissues collected from mice with no stereotaxic surgeries and no 7AAD staining. GFP+/7-AAD+ events and GFP-/7-AAD+ events were sorted using 100  $\mu$ m nozzle on a BD FACSARIA II and collected separately into 1.5 mL microcentrifuge tubes containing 100  $\mu$ l of resuspension buffer. A step-by-step protocol for nuclei extraction and snMultiome-seq data processing for *Projection-TAGs* can be found on Protocols.io [<https://doi.org/10.17504/protocols.io.yxmvm9dm5l3p/v1>].

## snRNA-seq and snATAC-seq

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

## Target amplification

Parallel PCR reaction (KAPA HiFi HotStart ReadyMix, Roche) was carried out to amplify *Projection-TAG* sequences from the cDNA libraries generated using 10X Genomics Chromium Single Cell Gene Expression 3' V3.1 or Multiome Assay. First, forward primer (5'-ACACTCTTCCC-TACACGACGCTCTTCCGATCT-3') and reverse primer (5'-GTGACTGGAGTTCAGACGTGTGCTTCCGATCTCCTCCCGATCGATACCG'-3') were used separately in PCR reactions with 20 ng of cDNA library each reaction. PCR program: (1) 95 °C for 3 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 containing forward and reverse primers were pooled, and a second PCR reaction was done following the PCR program above. As a result, the region containing the *Projection-TAG* sequence, 10X cell barcodes, and UMI sequences will be preferentially amplified from the cDNA molecules. The target amplification amplicons were labeled with sample indices (Dual Index Plate TT Set A, 10X Genomics) 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 (12 cycles), (3) 72 for 1 min. PCR products were purified using SPRIselect beads and sequenced on a NovaSeq6000 with 150 cycles each for Read1 and Read2, targeting 10 M reads per library. Raw sequencing data from individual libraries were processed and mapped to individual *Projection-TAGs* using 10X Genomics cellranger-6.1.2.

## Quality control, clustering, and annotation of snRNA-seq

The gene-cell count matrices from all snRNA-seq libraries were concatenated using R (V4.1.1) package Seurat (V4.4.0)<sup>119</sup>. To be included in the snRNA-seq analysis, nuclei were required to contain at least 500 unique genes, <15,000 total UMIs, and fewer than 5% of the counts deriving from mitochondrial genes. There were 69,657 nuclei that met these criteria. Raw counts were scaled to 10,000 transcripts per nucleus and log-transformed using NormalizeData() function to control the sequencing depth between nuclei. Counts were centered and scaled for each gene using ScaleData() function. Highly variable genes were identified using FindVariableFeatures(), and the top 20 principal components were retrieved with RunPCA() using default parameters. For dimension reduction and visualization, Uniform Manifold Approximation and Projection (UMAP) coordinates were calculated using RunUMAP(). Nuclei clustering was performed using FindClusters() based on the variable features from the top 20 principal components, with the resolution set at 0.6, and the marker genes for each cluster were identified using FindAllMarkers(), comparing nuclei in one cluster to all other nuclei. Doublet or low-quality nuclei were identified if they meet any of the following criteria: 1) Assigned to a

cluster with no significantly enriched marker genes (FDR < 0.05,  $\log_2FC > 1$ ); 2) Assigned to a cluster in which five or more mitochondrial genes were identified among top 20 marker genes (sorted by  $\log_2FC$ ); 3) Identified as doublets using R package DoubletFinder with doublet expectation rate at 5%<sup>120</sup>. In total 5991 nuclei were identified as doublet or low-quality and thus excluded from the dataset. The remaining 63,666 nuclei were clustered as described above.

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

## Projection feature annotation

To study the projection feature of single neurons, we assigned the projection targets to 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 X equals the percentage of nuclei not expressing the corresponding *Projection-TAG* in the corresponding snRNA-seq library plus one), whichever is greater. A projection target is assigned to a single nucleus if 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 expresses 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<sup>121</sup>. 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 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)<sup>122</sup>. To be included for snATAC-seq analysis, nuclei were required to be present in snRNA-seq data and contain at least 1000 fragments overlapping with peaks. There were 41,553 nuclei that met these criteria. Term frequency-inverse document frequency normalization was performed 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. The analysis uses Wilcoxon Rank Sum test, with *p* value adjustment based on Bonferroni correction using all features in the dataset. 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  $\log_2\text{FC} > 0.5$  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  $\log_2\text{FC} > 0.5$  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 GREs to regulate the expression of target genes, we modified a computational pipeline previously described<sup>54</sup>. 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 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

cell-type-specific gene expression, we identified pairs of cell-type-specific snATAC-seq peaks ( $\log_2\text{FC} > 0.5$ , FDR < 0.05, comparing peak accessibility in nuclei of one cell type to all others) and cell-type-specific snRNA-seq genes ( $\log_2\text{FC} > 0.5$ , FDR < 0.05, comparing gene expression in nuclei of one subtype to all others) in the same subtype. 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 ( $\log_2\text{FC} > 0.5$ , FDR < 0.05, comparing peak accessibility in nuclei of one projection to all others) and projection-specific snRNA-seq genes ( $\log_2\text{FC} > 0.5$ , 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 IEGs previously described<sup>88</sup>. 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 4.1.1. 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 Seurat V4.4.0, ggplot2 V3.4.0, and gplots V3.1.3. Figure 1a, Supplementary Fig. 3d, 5a, 6a, and 7a were generated using bioRender.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data 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](https://www.ncbi.nlm.nih.gov/sra/277718). The counts-per-million BC-sample matrix from the RNA-seq experiment reported in Supplementary Fig. 2b is available in the Source Data file. Source data are provided with this paper.

## Code availability

Custom pipelines and scripts are available on GitHub [<https://github.com/Samineni-Lab/Projection-TAGs>] and WashU Digital Commons@Becker Database [<https://doi.org/10.17632/d4htjwvcmg.2>].

## References

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

2. Golgi, C. Sulla fina anatomia degli organi centrali del sistema nervosa. *Rev. Sper. Freniatr.* (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).
5. Mukamel, E. A. & Ngai, J. Perspectives on defining cell types in the brain. *Curr. Opin. Neurobiol.* **56**, 61–68 (2019).
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).
7. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
8. Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2020).
9. Zeisel, A. et al. in *Cell* 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).
11. Fuzik, J. et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).
12. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 e1016 (2018).
13. Gouwens, N. W. et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* **22**, 1182–1195 (2019).
14. Gouwens, N. W. et al. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 935–953.e919 (2020).
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).
16. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
17. Langlieb, J. et al. The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333–342 (2023).
18. Liu, H. et al. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* **624**, 366–377 (2023).
19. Zu, S. et al. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378–389 (2023).
20. Shi, H. et al. Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552–561 (2023).
21. Winter, C. C. et al. A transcriptomic taxonomy of mouse brain-wide spinal projecting neurons. *Nature* **624**, 403–414 (2023).
22. Zhou, J. et al. Brain-wide correspondence of neuronal epigenomics and distant projections. *Nature* **624**, 355–365 (2023).
23. Harris, J. A. et al. Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
24. Munoz-Castañeda, R. et al. Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159–166 (2021).
25. Gao, L. et al. Single-neuron projectome of mouse prefrontal cortex. *Nat. Neurosci.* **25**, 515–529 (2022).
26. Economou, M. N. et al. Distinct descending motor cortex pathways and their roles in movement. *Nature* **563**, 79–84 (2018).
27. Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
28. Kim, E. J. et al. Extraction of distinct neuronal cell types from within a genetically continuous population. *Neuron* **107**, 274–282.e276 (2020).
29. Kaspar, B. K. et al. Targeted retrograde gene delivery for neuronal protection. *Mol. Ther.* **5**, 50–56 (2002).
30. Tervo, D. G. et al. A designer AAV variant permits efficient retrograde access to projection neurons. *Neuron* **92**, 372–382 (2016).
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).
32. Reardon, T. R. et al. Rabies virus CVS-N2c(DeltaG) strain enhances retrograde synaptic transfer and neuronal viability. *Neuron* **89**, 711–724 (2016).
33. Soudais, C., Laplace-Builhe, C., Kiss, 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).
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).
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).
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).
37. Lanciego, J. L. & Wouterlood, F. G. A half century of experimental neuroanatomical tracing. *J. Chem. Neuroanat.* **42**, 157–183 (2011).
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* <https://doi.org/10.1101/2023.08.30.555626> (2023).
39. Wang, Y. et al. Multimodal mapping of cell types and projections in the central nucleus of the amygdala. *Elife* **12** <https://doi.org/10.7554/elife.84262> (2023).
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).
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).
42. Phillips, J. W. et al. A repeated molecular architecture across thalamic pathways. *Nat. Neurosci.* **22**, 1925–1935 (2019).
43. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
44. Zhang, Z. et al. Epigenomic diversity of cortical projection neurons in the mouse brain. *Nature* **598**, 167–173 (2021).
45. Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
46. Kebschull, J. M. et al. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987 (2016).
47. Chen, X. et al. High-throughput mapping of long-range neuronal projection using *in situ* sequencing. *Cell* **179**, 772–786 e719 (2019).
48. Klingler, E. et al. Temporal controls over inter-areal cortical projection neuron fate diversity. *Nature* <https://doi.org/10.1038/s41586-021-04048-3> (2021).
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).
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).
51. Kuramoto, E. Method for labeling and reconstruction of single neurons using Sindbis virus vectors. *J. Chem. Neuroanat.* **100**, 101648 (2019).
52. Xu, P. et al. High-throughput mapping of single-neuron projection and molecular features by retrograde barcoded labeling. *Elife* **13** <https://doi.org/10.7554/elife.85419> (2024).
53. Zhao, Q. et al. A multidimensional coding architecture of the vagal interoceptive system. *Nature* <https://doi.org/10.1038/s41586-022-04515-5> (2022).

54. Yang, L. et al. Human and mouse trigeminal ganglia cell atlas implicates multiple cell types in migraine. *Neuron* **110**, 1806–1821 e1808 (2022).

55. Emert, B. L. et al. Variability within rare cell states enables multiple paths toward drug resistance. *Nat. Biotechnol.* **39**, 865–876 (2021).

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

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).

58. Pennartz, C. M. et al. Corticostriatal interactions during learning, memory processing, and decision making. *J. Neurosci.* **29**, 12831–12838 (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).

60. Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).

61. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).

62. Beine, Z., Wang, Z., Tsoulfas, P. & Blackmore, M. G. Single nucleus analyses reveal transcriptional profiles and marker genes for diverse supraspinal populations. *J. Neurosci.* **42**, 8780–8794 (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).

64. Hunnicutt, B. J. et al. A comprehensive thalamocortical projection map at the mesoscopic level. *Nat. Neurosci.* **17**, 1276–1285 (2014).

65. J. L. MacDonald, R. M. F., E. Azim, S. J. Shrider, 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).

67. Kawaguchi, Y. Pyramidal cell subtypes and their synaptic connections in layer 5 of rat frontal cortex. *Cereb. Cortex* **27**, 5755–5771 (2017).

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

69. Winnubst, J. et al. Reconstruction of 1000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell* **179**, 268–281.e213 (2019).

70. Zemke, N. R. et al. Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390–402 (2023).

71. 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).

72. Halfon, M. S. Silencers, enhancers, and the multifunctional regulatory genome. *Trends Genet.* **36**, 149–151 (2020).

73. 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).

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

75. Mich, J. K. et al. Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).

76. Graybuck, L. T. et al. Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449–1464 e1413 (2021).

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

78. 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).

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

80. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012 (2014).

81. Luo, Y. et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).

82. 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).

83. Ziegler, K. et al. Primary somatosensory cortex bidirectionally modulates sensory gain and nociceptive behavior in a layer-specific manner. *Nat. Commun.* **14**, 2999 (2023).

84. Chen, T. et al. Top-down descending facilitation of spinal sensory excitatory transmission from the anterior cingulate cortex. *Nat. Commun.* **9**, 1886 (2018).

85. Liu, Y. et al. Touch and tactile neuropathic pain sensitivity are set by corticospinal projections. *Nature* **561**, 547–550 (2018).

86. Gan, Z. et al. Layer-specific pain relief pathways originating from primary motor cortex. *Science* **378**, 1336–1343 (2022).

87. Taub, D. G. et al. The secondary somatosensory cortex gates mechanical and heat sensitivity. *Nat. Commun.* **15**, 1289 (2024).

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

89. Hrvatin, S. et al. in *Nature Neuroscience*. 21 120–129 (2018).

90. Skinnider, M. A. et al. Cell type prioritization in single-cell data. *Nat. Biotechnol.* **39**, 30–34 (2021).

91. Condylis, C. et al. Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022).

92. Sorensen, A. T. et al. A robust activity marking system for exploring active neuronal ensembles. *eLife* **5**, <https://doi.org/10.7554/eLife.13918> (2016).

93. Mukherjee, D. et al. Salient experiences are represented by unique transcriptional signatures in the mouse brain. *eLife* **7**, <https://doi.org/10.7554/eLife.31220> (2018).

94. 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. USA* **117**, 23304–23310 (2020).

95. 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).

96. 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).

97. Brakeman, P. R. et al. Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 284–288 (1997).

98. Szumlinski, K. K. et al. Behavioral and neurochemical phenotyping of Homer1 mutant mice: possible relevance to schizophrenia. *Genes Brain Behav.* **4**, 273–288 (2005).

99. 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).

100. 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).

101. Pillay, S. & Carette, J. E. Host determinants of adeno-associated viral vector entry. *Curr. Opin. Virol.* **24**, 124–131 (2017).

102. Bugeon, S. et al. A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330–338 (2022).

103. Xu, S. et al. Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370**, <https://doi.org/10.1126/science.abb2494> (2020).

104. von Buchholtz, L. J. et al. Decoding cellular mechanisms for mechanosensory discrimination. *Neuron* **109**, 285–298.e285 (2021).

105. Han, Z. et al. AAV11 enables efficient retrograde targeting of projection neurons and enhances astrocyte-directed transduction. *Nat. Commun.* **14**, 3792 (2023).

106. 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).

107. Lin, K. et al. AAV9-Retro mediates efficient transduction with axon terminal absorption and blood-brain barrier transportation. *Mol. Brain* **13**, 138 (2020).

108. 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).

109. You, Y. et al. Benchmarking UMI-based single-cell RNA-seq pre-processing workflows. *Genome Biol.* **22**, 339 (2021).

110. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).

111. 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).

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

113. 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).

114. Wang, Q. et al. The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell* **181**, 936–953.e920 (2020).

115. 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).

116. Berg, S. et al. ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

117. 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).

118. Yang, L., Tochitsky, I., Woolf, C. J. & Renthal, W. Isolation of nuclei from mouse dorsal root ganglia for single-nucleus genomics. *Bio. Protoc.* **11**, e4102–e4102 (2021).

119. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).

120. 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).

121. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e3226 (2021).

122. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

## 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., 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 interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60360-w>.

**Correspondence** and requests for materials should be addressed to Vijay K. Samineni.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025