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

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

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

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

*Contact information: vksamineni@wustl.edu

Abstract

Single-cell multiomic techniques have sparked immense interest in developing a comprehensive multi-modal map of diverse neuronal cell types and their brain wide projections. However, investigating the 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 mouse cortex and high-throughput single-cell profiling of the transcriptional and epigenetic landscapes of the cortical projection neurons. Projection-TAGs can be leveraged to obtain a snapshot of activity-dependent recruitment of distinct projection neurons and their molecular features in the context of a specific stimulus. Given its flexibility, usability, and compatibility, we envision that Projection-TAGs can be readily applied to build a comprehensive multi-modal map of brain neuronal cell types and their projections.

Main

Understanding brain function requires the elucidation of the complex wiring diagram and constituent cell types across brain regions. Revolutionary work from Golgi and Cajal laid the foundation for understanding the diversity of neurons and their anatomical connections based on morphological features[3,4]. 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[5-21]. 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[22-29]. Despite the rapid integration of novel multiomic technologies in studying brain wide connectivity, the investigation of spatially mingled neuronal projections continues to be hampered by the lack of broadly available tools to simultaneously trace multiple neuronal projections and/or profile projection neurons for multi-modal investigations.

Traditional neuroanatomical tracing methods, performed often with tracers or with viral vectors use fluorescence as the projection identifier, have been invaluable in mapping distinct neuronal projections[30-38]. However, these methods are limited by their spectral properties of fluorophores that can be detected in a single experiment and 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[39], common microscopy equipment in neuroscience labs can reliably detect only 3-4 fluorophores, limiting the number of projections to be examined simultaneously[28,40,41]. Furthermore, such approaches are not suited for high-throughput sequencing-based assays such as single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq), as the detection of exogenous transcripts is usually low by short-read sequencing. Alternatively, one can still build a multi-modal atlas for projection neurons by tracing one or two projections at a time using fluorophores and incorporating this tracing paradigm with fluorescently-activated cell sorting (FACS) for single-cell profiling[42,43]. Methods, such as retro-seq and epi-retro-seq, characterized the molecular features of projection neurons[23,27,29,44-46], but they are often inefficient and come with a high experimental cost due to the multiplex strategy.

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. 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[47-51]. 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[48,52]. More recently, retrograde AAVs were used for multiplex projection tracing of peripheral sensory neurons innervating visceral organs by expressing a unique RNA barcode into each organ[53]. Although this is a very sophisticated approach, it does not label nuclei for epigenetic assays and may not be easily scalable and safely adaptable to different applications due to their RNA barcode design (truncated protein-coding sequences instead of randomly generated sequences that could interfere with gene expression or cellular function).

Here we introduce Projection-TAGs, a retrograde AAV platform that allows multiplex tracing of projection neurons by tagging each neuronal projection with a unique RNA barcode. The key component of Projection-TAGs is a set of engineered retrograde AAVs each expressing a unique barcode, which acts as the projection identifier. With this scheme, neurons projecting to a target region are uniquely tagged by a retrograde AAV-mediated RNA barcode and multiplex projection tracing can be achieved by injecting a unique barcode-expressing Projection-TAG AAVs into each of the target regions predefined for investigation. To map projection targets, RNA barcodes are demultiplexed using commercial assays, allowing multiplex neuroanatomical studies and high-throughput multi-modal profiling of projection neurons in single animals. Projection-TAGs can be utilized to study activity-dependent populations in response to a stimulus of interest. In this study, we utilized Projection-TAGs to study the 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 central nervous system.

Results

Design of Projection-TAGs

To make Projection-TAGs democratized and readily deployable to study the nervous system in neuroscience labs without any specialized equipment, we incorporated the following features into its design: 1) flexibility: can be used across multiple tissue types, 2) usability: easy to adapt to existing pipelines and optimized for commercial assays, and 3) compatibility: low toxicity and stable barcode (BC) expression to allow the potential to couple with additional experimental paradigms. To enhance the versatility of Projection-TAGs across different tissue types and research applications, we employed a chicken beta-actin (CAG) promoter, under an AAV plasmid backbone, to enable ubiquitous RNA expression in various tissue and cell types (Fig. 1a). We reasoned that the inclusion of a fluorescent marker could enable flexibility in downstream processes and confirming viral expression. Upon screening several fluorescent proteins fused with localization domains (Extended Data Fig. 1a,b), our final construct contains a fusion of Sad1 And UNC84 Domain Containing 1 (Sun1) to Green Fluorescent Protein (GFP). Sun1-GFP fusion protein reliably labels the nuclear membrane, allowing enrichment of target cell/nucleus populations using FACS while remaining relatively safe for epigenetic assays (Extended Data Fig. 1c,d). When the green channel is needed for imaging experiments, Sun1-GFP can be photobleached (Extended Data Fig. 2e,g). We also developed a version of Projection-TAGs expressing a red fluorophore oScarlet (Methods), that could facilitate combinatorial use with green and far-red reporters.

To optimize BC detection for various commercial assays, we cataloged 50 100-bp unique BCs[54], whose sequences are distinct from the genomes of human and select model animals (Extended Data Fig. 1e, Supplementary Table 1, Methods). To verify these BC sequences, we inserted them into the 3' untranslated region (UTR) of the *Sun1-GFP* transcript and generated 12 Projection-TAG plasmids each expressing a unique BC (BCs 1-12, Fig. 1a). To test if the Projection-TAG BCs can be expressed and reliably detected, we first performed RNA-seq and multiplexed fluorescent *in situ* hybridization (FISH) on human embryonic kidney (HEK) cells transfected with individual Projection-TAG plasmids (Extended Data Fig. 2a,c). We faithfully detected distinct BCs with high specificity, as the detection rate of BCs expressed in each HEK sample is at least 422.4 times (RNA-seq) and 6.9 times (multiplexed FISH) greater than that of BCs not expressed or the background (Extended Data Fig. 2b,d). To assess and confirm if Projection-TAGs work in multiplex tracing of neuronal projections in vivo, we packaged these

validated plasmids into recombinant AAV serotype 2 retro (rAAV2-retro)[31] and generated a set of 12 Projection-TAG AAVs each expressing a unique BC (BCs 1-12, Fig. 1a).

Optimization for retrograde tracing of cortical projection neurons using Projection-TAGs

To evaluate the ability of using Projection-TAGs in high-throughput multi-modal profiling of projection neurons, we applied it to performing multiplex projection tracing in the adult mouse primary

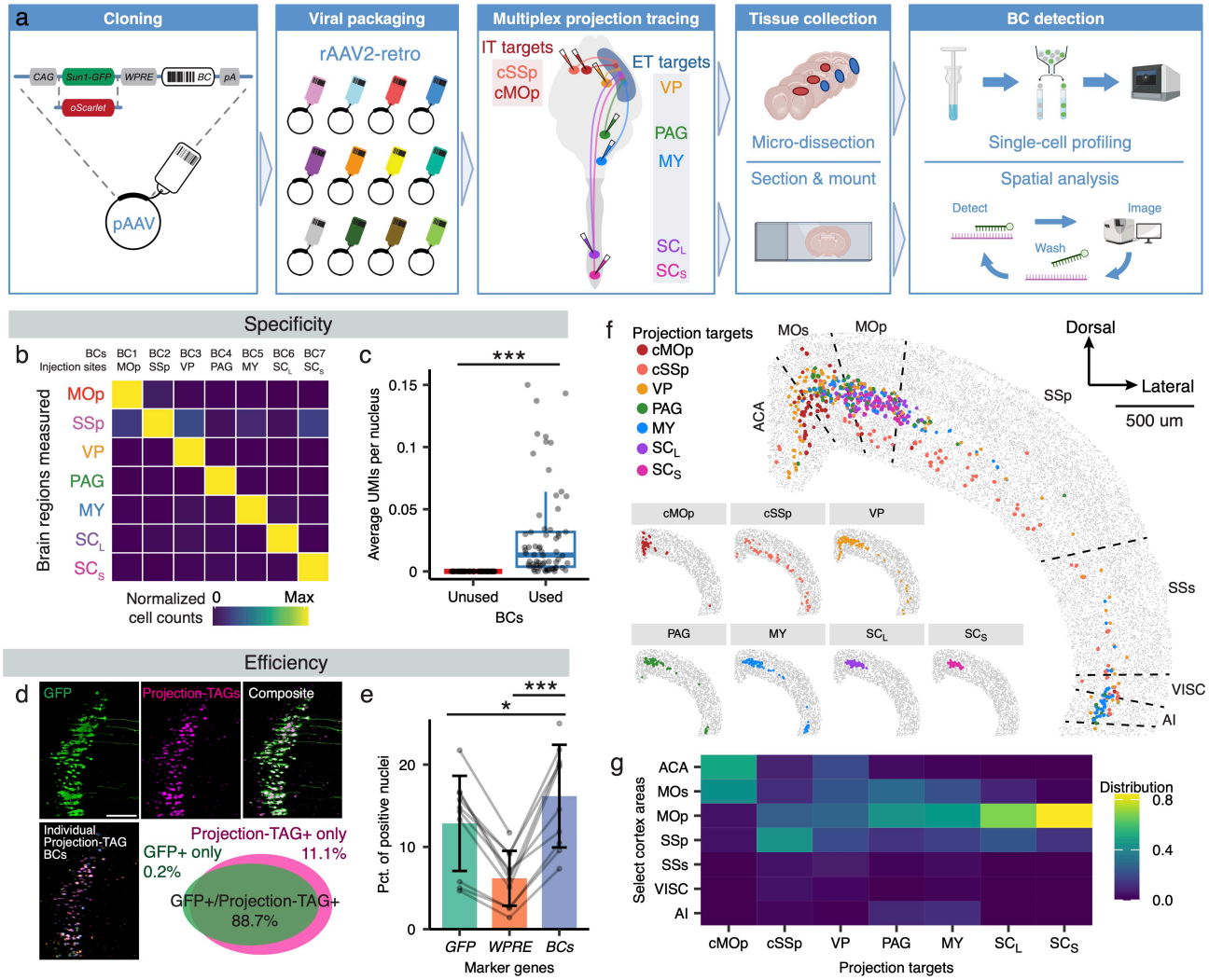


Figure 1. Multiplex projection tracing using Projection-TAGs
a, The experimental workflow of tagging neuronal projections of the MOp and SSp using Project-TAGs. **b**, Heatmap showing the enrichment of cells labeled with individual Projection-TAG BCs by multiplexed FISH in brain regions of injections. Cell counts are normalized by the max values in each brain region. Individual BCs and their injection sites are shown as on the column. **c**, Box plot showing the detection rate of Projection-TAG BCs in each snRNA-seq library prepared by FACS (Supplementary Table 2). Projection-TAG BCs are grouped by whether they were used in a given experiment. Detection rate is significantly higher for used BCs than unused ones (***p = 6.5e-08, one-sided student t-test). **d**, Representative MOp images showing cells labeled with the native fluorescence from Sun1-GFP protein and fluorescent signals from Projection-TAGs. Quantification shows the overlap between cells that are GFP+ and/or Projection-TAG+. FISH probes uniquely targeting Projection-TAGs (BCs 1-7) were visualized in separate imaging channels, shown as individual BCs, and the signals were aggregated pos-hoc, shown as Projection-TAGs. Scale bar = 100 µm. **e**, Dot plot showing the percentage of nuclei expressing select marker genes (>0 counts) in individual snRNA-seq libraries prepared by FACS. Bar and error bar indicate the average and standard deviation across libraries. There are significantly more Projection-TAG+ nuclei (nuclei expressing any of the BCs used in the experiment) compared to *GFP*+ and *WPRE*+ nuclei (***p = 7.98e-08, $F(2,18)$ = 46.32, one-way ANOVA; *p = 0.025 between BC and GFP, ***p = 5.13e-05 between BC and WPRE, post-hoc two-sided paired t-tests with Bonferroni correction). **f**, Spatial map of projection neurons in select cortex areas from a brain section (-0.65 mm relative to bregma). The projection targets are assigned to individual cells based on the Project-TAG BCs they express. Cells are downsampled to 200 per projection target and colored by their projection targets. Insets on the bottom left corner show the same cells separated by their individual projection targets. Grey dots are cells segmented based on DAPI signal. **g**, Heatmap showing the distribution of projection neurons in select cortex areas from Fig. 1f. Values in the heatmap are counts of neurons projecting to each target in each area normalized by total counts of Project-TAG+ cells for the given projection target. 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.

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[44,55-57]. 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[28,46,51,58-60], 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 periductal grey [PAG], ipsilateral medulla [MY]) and two corticospinal targets (lumbar spinal cord [SC_L], and sacral spinal cord [SC_S]).

To characterize the temporal kinetics and stability of Projection-TAGs, we injected a BC-expressing AAV into either MY or SC_S and quantified the BC expression overtime in the cortex using qPCR (Extended Data Fig. 3a). The BC expression in mice receiving MY injection rapidly increased, reaching its peak at week three and plateauing thereafter. Conversely, BC expression in mice receiving SC_S injection increased more steadily and plateaued at around week five. Despite different kinetics in the early weeks, BC expression in both cases remained stable during the time course profiled in the experiment. Next, we determined the optimal viral titer for the labeling efficiency of Projection-TAGs. We extracted nuclei from eight cortical samples collected from mice receiving injections into seven downstream projection targets mentioned above using either undiluted Projection-TAG AAVs (1.1 – 3.3 x 10^{12} viral genomes/ml) or those of 4X dilution, followed by quantifying the percentage of GFP+ nuclei using FACS. Undiluted AAVs labeled significantly more nuclei compared to diluted AAVs (14.7 ± 8.4% for undiluted AAVs and 3.7 ± 3.0% for diluted AAVs, p = 0.03, one-sided t-test). We next performed single-nucleus RNA-seq (snRNA-seq) on those FACS-sorted GFP+ nuclei. While snRNA-seq libraries from the two groups yielded similar transcripts per nucleus (p = 0.37, t-test), the detection rate of BCs is significantly higher in nuclei labeled by undiluted AAVs compared to those of diluted AAVs (p = 2.2e-16, Extended Data Fig. 3b). Moreover, we noticed that BC+ nuclei co-cluster well with BC- nuclei on the UMAP, indicating little gene expression alteration due to AAV infection (Extended Data Fig. 3c). Additional analysis into immune response to viral infection showed no significant differences between BC+ nuclei and BC- nuclei (Extended Data Fig. 3d). These results suggest Projection-TAG AAVs stably express BCs long-term with minimal immune response. Furthermore, to optimize Projection-TAGs for the multiplex labeling of the projections in the snRNA-seq experiment, we compared the BC detection rate in two different 3' snRNA-seq assays (Supplementary Table 2 and Extended Data Fig. 3e). Chromium Single Cell Multiome Kit from 10X Genomics (10X Genomics) yielded higher BC detection rate and identified proportionally more BC+ nuclei compared to Evercode WT Mini kit from Parse Biosciences (Parse), despite lower sequencing depth (29k reads/nucleus for 10X and 38k reads/nucleus for Parse). Therefore, we used the 10X Genomics kit for the sequencing experiments described in this study.

Projection-TAG AAVs label projection neurons with high specificity and efficiency

To examine the specificity of Projection-TAGs in vivo, we performed multiplex tracing of the MOp and the SSp downstream projections by injecting a Projection-TAG AAV expressing a unique BC into each projection target. We observed high correspondence in BC detection within the injection sites using multiplexed FISH with 19.5 – 393 times greater cells labeled with BCs injected into a given region than those injected into other regions (Fig. 1b, Extended Data Fig. 4c, and Methods). To further support this, we calculated the detection rate of individual BCs in snRNA-seq, grouped by whether they were used in

given experiments (Fig. 1c). BCs used in experiments are highly expressed (75.9 ± 6.6 percentile by expression ranks among all genes) while BCs not used yielded zero counts in all libraries. Those data showed that the BC detection is highly specific in both spatial validation and single-cell experiments.

Next, to compare the efficiency of Projection-TAGs to traditional fluorophore-based approaches, we examined if the MOp cells labeled with Projection-TAG AAV-mediated BCs track the expression of the Sun1-GFP driven by the same construct. We observed a high degree of overlap between cells labeled with BCs and those labeled with native GFP fluorescence (Fig. 1d), while BCs label slightly more cells (likely due to different imaging settings between the GFP channel and BC channels). Additionally, Sun1-GFP and BCs exhibit distinct subcellular compartmentalization. BCs reliably label the soma of retrogradely labeled cells, whereas Sun1-GFP broadly labels the soma, dendrites, and the proximal portion of the axons. Detection of exogenous marker genes is low by snRNA-seq, often because they are located distally from the 3' end of the transcript and thus unlikely to be captured by 3' snRNA-seq approaches. We reasoned that positioning BCs at the 3' UTR would improve their detection, and indeed we observed significantly more BC+ nuclei compared to nuclei positive for *Sun1-GFP* or *WPRE* (Fig. 1e). These data suggests that Projection-TAGs exhibit higher tracing efficiency compared to fluorophore-based AAV approaches.

Parallel detection of Projection-TAGs allows multiplex tracing of projection neurons

As it has been well established that neurons projecting the ET and IT targets exhibit distinct spatial distribution across the cortex[27,61], we next sought to validate the spatial distribution of Projection-TAG labeled neurons following AAV injection into the projection targets mentioned above. We detected individual Projection-TAG BCs using multiplexed FISH, followed by cell segmentation and we assigned projection targets to projection neurons based on the expression of respective Projection-TAG BCs. We then quantified the distribution of neurons projecting to each target in several neocortical areas (Extended Data Fig. 5a-c and Methods). We observed that neurons projecting to different targets are enriched in distinct areas across the cortex (Fig. 1f, g and Extended Data Fig. 5c). For example, cMOp-projecting neurons are located medially in the MOp, whereas cSSp-projecting neurons are dominantly found in MOp and SSp. While VP-, PAG-, and MY-projecting neurons are distributed across multiple cortex areas, SC_L- and SC_S-projecting neurons are highly enriched around MOp. The spatial distribution of projection neurons is largely consistent with previous reports[24,62-64]. It has been reported that thalamus- and MY-projecting neurons exhibit distinct spatial distribution in layer 5 (L5) of the anterior lateral motor cortex, where few neurons project to both thalamus and MY[27,65]. Indeed, we observed similarly that VP-projecting neurons were enriched in upper L5 whereas MY-projecting neurons were more abundant in lower L5 in the anterior part of the MOp with a significant portion of labeled neurons projecting to both targets (Extended Data Fig. 5d). Interestingly, their distinct sub-layer distribution appeared to attenuate towards the posterior MOp, accompanied by an increased proportion of neurons projecting to both targets. While numerous cortex areas contain neurons projecting to the seven targets examined in this study, our analysis highlights the MOp and the SSp as particularly rich in projection neurons among all cortex areas analyzed and are highly enriched with neurons projecting to each of the seven targets (Extended Data Fig. 5e).

A multi-modal single-cell atlas of mouse cortex

We next assessed whether Projection-TAGs can be used to obtain both the gene expression and chromatin accessibility features of projection neurons by simultaneous integration with multiomic analysis of snRNA-seq and single-nucleus ATAC-seq (snATAC-seq). We first performed snRNA-seq on four MOp

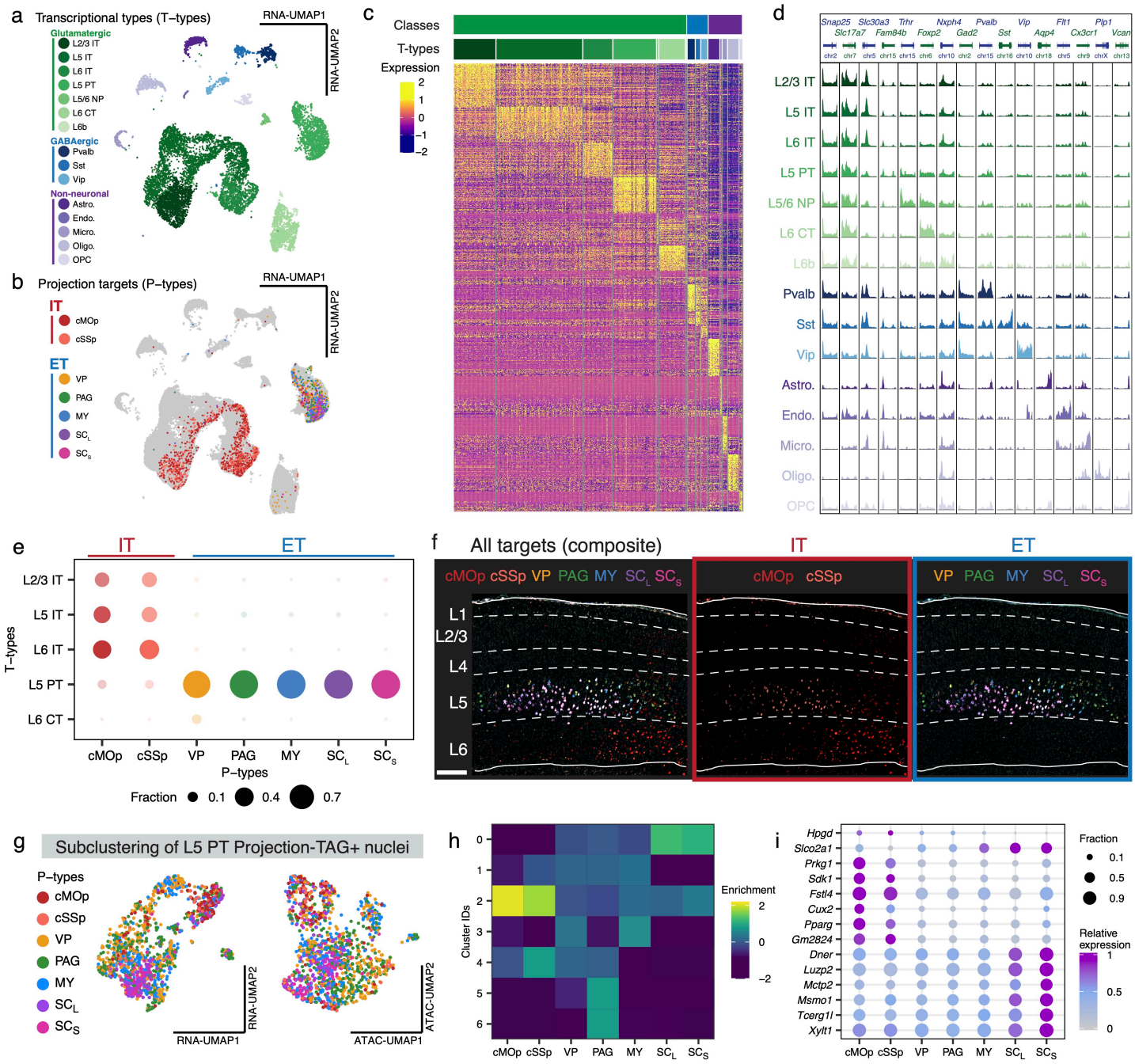


Figure 2. A multi-modal single-cell atlas of mouse cortex

a, Uniform manifold approximation and projection (UMAP)[1] visualization of snRNA-seq data showing 10,000 downsampled nuclei, colored by T-types. **b**, UMAP visualization of snRNA-seq data showing 17,000 nuclei (10,000 downsampled Projection-TAG- nuclei and 1,000 downsampled Projection-TAG+ nuclei for each Projection-TAG BC). Projection-TAG- nuclei are colored grey and Projection-TAG+ nuclei are colored by their projection targets (P-types). **c**, Heatmap showing expression of top 100 marker genes (sorted by Log2FC) per T-types in individual nuclei of snRNAs-seq. Nuclei are downsampled to 1,000 and ordered based on T-types. See Fig. 2a for the colors of classes and T-types. **d**, Coverage plot of average accessibility of canonical cell-type-specific marker genes in individual T-types of snATAC-seq. The chromatin accessibility is displayed as the average frequency of sequenced DNA fragments per nucleus for each T-type, grouped by 50 bins per displayed genomic region. **e**, Dot plot showing fraction of Projection-TAG+ nuclei annotated by individual T-types, normalized by the total counts of Projection-TAG+ nuclei per P-type. **f**, 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 BCs. Scale bar = 150 μm. Left: composite image of all FISH channels, middle and right: images of a subset of Projection-TAG channels labeling IT targets and ET targets, respectively. **g**, UMAP of L5 PT neurons that are Projection-TAG+. Nuclei are clustered based on snRNA-seq (left) and snATAC-seq (right), and colored by P-types. **h**, Heatmap showing the enrichment of P-types in each cluster. The enrichment indicates the Z-score for each P-type (by row) of the percentage of nuclei assigning to each cluster. **i**, Dot plot displaying the expression of select P-type-specific marker genes in nuclei of individual P-types. Dot size denotes the fraction of nuclei expressing a marker gene (>0 counts), and color denotes relative expression of a gene in each cell type (calculated as the mean expression of a gene relative to the highest mean expression of that gene across all P-types).

204

samples and six SSp samples from six mice receiving injections of Projection-TAG AAVs into seven downstream projection targets mentioned above, generating 14 libraries containing a total of 69,657 nuclei with an average of 3,621 genes per nucleus (Supplementary Table 2). After quality control and removal of low-quality nuclei and doublets, 63,666 nuclei were retained in the snRNA-seq dataset (Extended Data Fig. 6a,d). Notably, nuclei from individual libraries co-clustered together, indicating a low batch effect across libraries and largely shared gene expression profiles between nuclei from MOp and SSp (Extended Data Fig. 6c-e). Clustering analysis revealed 26 distinct transcriptional clusters, which were further classified into three major classes (Glutamatergic neurons, GABAergic neurons, and non-neuronal cells) including 15 transcriptional cell types (T-types, Fig. 2a). 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[28,44,59]. Following this naming convention, we assigned T-types to transcriptionally-defined glutamatergic neuronal clusters based on their layer enrichment and projection patterns previously described: IT neurons from layers 2/3 (L2/3 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 T-types based on the expression of canonical marker genes: *Pvalb*-expressing, *Sst*-expressing, and *Vip*-expressing neurons. We also identified five T-types of non-neuronal cells: astrocytes (Astro.), endothelial cells (Endo.), microglial cells (Micro.), oligodendrocytes (Oligo.), and oligodendrocyte progenitor cells (OPC). While there was some variability of T-type distribution across libraries, likely due to different nuclear preparation approaches (Extended Data Fig. 6h, Supplementary Table 2), individual T-types showed distinct gene expression patterns with enrichment of canonical marker genes in respective T-types [16,27,60] (Fig. 2c, Extended Data Fig. 6f,g, and Supplementary Table 3).

To enable simultaneous investigation of the gene expression and chromatin accessibility profiles of the same nuclei, we performed combinatorial snATAC-seq on eight snRNA-seq libraries reported above, generating a dataset of 41,553 high-quality nuclei with an average sequencing depth of 25,412 transposase-sensitive fragments per nucleus (Extended Data Fig. 7a)[66]. These snATAC-seq fragments captured the open chromatin regions of the genome as they followed the expected nucleosomal size distribution (Extended Data Fig. 7b). Clustering analysis of the snATAC-seq data revealed 26 distinct clusters. The chromatin accessibility and gene expression profiles of the sequenced nuclei are highly correlated, with 87.7 ± 18.0% nuclei in each snATAC-seq cluster assigned to the same T-type to those made when independently analyzed for snRNA-seq data (Extended Data Fig. 7c,d). In addition, nuclei in individual snATAC-seq clusters display distinct chromatin accessibility around the genomic loci of canonical marker genes for the corresponding T-types (Fig. 2d and Supplementary Table 3). Consequently, we grouped nuclei by T-types in subsequent snATAC-seq analysis.

Detection of Projection-TAGs by snRNA-seq enabled us to map the projection targets (P-types) to individual sequenced nuclei, thus allowing multi-modal profiling of projection neurons at single-cell resolution (Fig. 2b). To identify projection targets, we assigned P-types to each Projection-TAG+ nucleus based on the individual Projection-TAG BCs it expressed (Fig. 2b). Projection-TAGs labeled 6,261 nuclei (9.83% total, or 19.7% in FACS libraries, Extended Data Fig. 6h). Notably, we identified five glutamatergic neuronal T-types, L2/3 IT, L5 IT, L6 IT, L6 CT, and L6b, that have at least 5% of the nuclei being Projection-TAG+ (Extended Data Fig. 7e), which exhibited significantly higher enrichment of Projection-TAG+ nuclei compared to all other T-types (Extended Data Fig. 7f). These data suggest that these five T-types are enriched of individual neurons projecting to the seven targets examined in this study, consistent with what's been previously reported[27,28,44]. As prior reports have noted explicit correspondence between T-

types and P-types between neurons projecting to IT targets and ET targets[27,61], we next tested whether the neurons projecting to individual targets are transcriptionally distinct by investigating the composition of T-types in each P-type (Fig. 2e, Extended Data Fig. 5c). Hierarchical clustering separated seven P-types into three clusters. The first cluster contains two P-types of IT targets, cMOp and cSSp, with similar T-type composition. They both consist of transcriptionally-defined L2/3 IT, L5 IT, L6 IT, and L5 PT neurons (22.7%, 32%, 37.6%, and 7.3% for cMOp, 24.1%, 25.5%, 43.1%, and 6.8% for cSSp, respectively). The second and third clusters contain five P-types of ET targets. Neurons projecting to VP exhibit distinct T-type composition from all other P-types and consist of mostly L5 PT and minorly L6 CT neurons (88% and 8.7%, respectively), whereas neurons projecting to PAG, MY, SC_L, and SC_S are predominantly L5 PT neurons (95%, 97.1%, 97.2%, and 97.6%, respectively). Spatial analysis of multiplexed FISH further confirmed their layer distribution (Fig. 2f).

As nuclei expressing each of the seven Projection-TAGs are enriched in transcriptionally-defined L5 PT neurons, we next asked whether L5 PT neurons projecting to different targets are transcriptionally distinct. Subclustering of 3,589 L5 PT nuclei that are also Projection-TAG+ revealed distinct P-type enrichment across the UMAP space (Fig 2g). Subsequent snRNA-seq clustering analysis revealed seven transcriptionally distinct clusters. While nuclei of VP, PAG, and MY were distributed across all clusters, IT-projecting (cSSp and cMOp) nuclei were more enriched in cluster 2, and spinal cord-projecting (SC_L and SC_S) nuclei were more enriched in cluster 0, suggesting L5 PT neurons of different P-types consist of transcriptionally distinct nuclei (Fig. 2g,h). To identify molecular markers for L5-PT neurons projecting to each target, we identified differentially expressed (DE) genes and differentially accessible (DA) peaks in individual P-types (Fig. 2i and Supplementary Table 3). Among those DE genes, *Hpgd* and *Slco2a1* have been reported as marker genes for thalamus- and MY-projecting neurons, respectively[27]. Interestingly, we observed an expression gradient of those genes across different P-types, which appears to be correlated 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 identify cortical neurons projecting to multiple targets

As Projection-TAGs enable multiplex projection tracing in single animals, it opens new possibilities for studying neuronal collaterals projecting to multiple targets. A neuron sends its axonal collaterals to multiple targets may be labeled by multiple Projection-TAGs. Indeed, 22.6% of Projection-TAG+ nuclei in snRNA-seq express multiple Projection-TAG BCs (Fig. 3a). To investigate the collaterals of the individual projection neurons, we calculated the overlap of neurons projecting to any two targets (Fig. 3b). Two IT targets (cMOp and cSSp) exhibited significant overlap with each other but not with other P-types. PAG showed significant overlap with VP and MY, whereas VP, MY, SC_L, and SC_S demonstrated significant overlap among themselves. Those observations were further confirmed by examining overlapping across all P-types and multiplexed FISH (Extended Data Fig 8a,b).

The anatomical hierarchy of projections correlates distance in the molecular modalities of gene expression and chromatin accessibility

The pattern we identified by examining the overlap of P-types at single-cell resolution is striking and suggests that the axons of cortical neurons may be organized following a distinct anatomical pattern. To test the hypothesis, we examined the brain-wide axonal projection patterns of MOp and SSp using published datasets (Fig 3c,d, Extended Data Fig. 8c,d). Axon bundles innervating the contralateral IT targets (cMOp or cSSp) crossed the midline via the corpus callosum, while axon bundles innervating ET targets traveled down via the corticospinal tract. Axons innervating VP and PAG first separated with axons

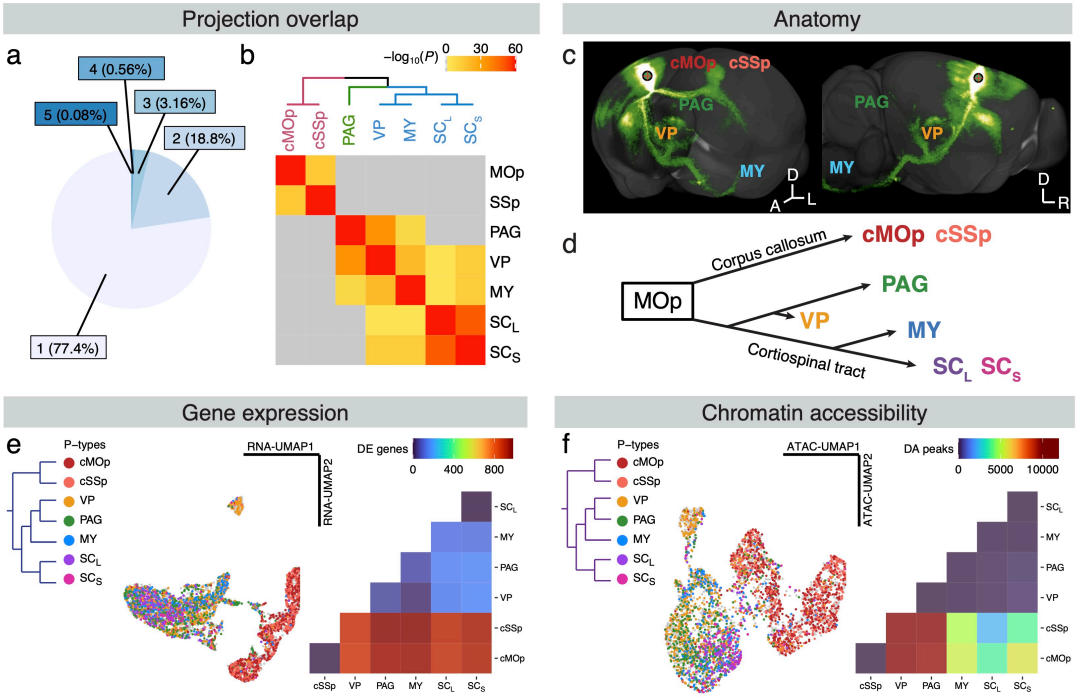


Figure 3. Projection-TAGs revealed the anatomical hierarchy of cortical projections **a**, Pie chart showing the percentage of Projection-TAG+ nuclei, grouped by the number of unique Projection-TAG BCs detected in each nucleus. **b**, Heatmap showing the Bonferroni-corrected *p* values from pairwise hypergeometric tests of overlapping between nuclei expressing any of the two P-types labeled with Projection-TAGs. Hierarchical clustering was performed based on the Bonferroni-corrected *p* values of each P-type to each of the other P-types. **c**, Tracing the brain-wide axon collaterals of MOp neurons. Screenshots were acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/100141563)[2]. **d**, Summary diagram of the anatomical hierarchy of axonal projections from MOp to seven target regions. **e**, Gene expression distance of P-types Left: hierarchical clustering was performed based on the average expression in individual P-types of each gene detected in snRNA-seq; middle: UMAP of Projection-TAG+ nuclei from snRNA-seq, colored by P-types; right: number of DE genes between any two P-types. **f**, Chromatin accessibility distance of P-types. Left: hierarchical clustering was performed based on the average accessibility in individual P-types of each peak detected in snATAC-seq data; middle: UMAP of Projection-TAG+ nuclei from snATAC-seq, colored by P-types; right: number of DA peaks between any two P-types.

reaching MY. While the spinal cord was not included in these experiments, SC-projecting axons reach the hindbrain via the corticospinal tract and continue to innervate SC[25,67]. These data suggest that the anatomical hierarchy of axonal projections of the cortex highly correlates with the overlapping patterns of the projections at the single-cell level identified by snRNA-seq.

Moreover, the hierarchical organization of projections is also reflected in the gene expression and chromatin accessibility profiles of individual P-types (Fig. 3e,f). Clustering of the snRNA-seq or snATAC-seq data by P-types revealed similar hierarchy, and their similarity in the omic spaces of gene expression or chromatin accessibility, shown as the number of DE genes or DA peaks between any two P-types, also exhibits similar patterns. In summary, the overlapping pattern of neurons projecting to targets identified in snRNA-seq data is consistent with the anatomical organization of these projections, which is further reflected in the transcriptional and epigenetic space. This complementary analysis provides a multi-modal and high-resolution view of the spatial and molecular intricacies of neuronal projections in the mouse brain.

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

While the gene expression program in the diverse MOp and SSp cell types has been largely elucidated by single-cell RNA-seq experiments, the regulatory networks that govern the distinct gene expression pattern in individual cell types are under investigated[23,45,59,68]. The genomic regulatory elements (GREs), mostly identified within the non-coding region of the genome and act in a cell-type-specific and tissue-specific manner, fine tune the expression level of the regulated genes[69]. 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[70,71]. 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[72-77]. Though recent progress has been made in characterizing the landscape of GREs for the projection neurons, the current experimental workflow is tedious and pain-staking as individual projection neurons are labeled via injection of retro Cre in floxed nuclear reporter mice, and regions of interest are pooled for analysis of potential GREs[23,45]. We tested whether Projection-TAGs could potentially be used to perform high throughput analysis of identifying projection-specific GREs.

To this end, we integrated the snRNA-seq and snATAC-seq data from MOp and SSp to produce a unified, multi-modal cell census. We first identified 166,540 peaks when aggregated across all snATAC-seq libraries. We believe many of those peaks are likely to contain functionally relevant GREs, as 85.6% of the peaks are located at the distal regions in the genome, and 66.3% of them are differentially accessible in individual T-types and/or P-types (Fig. 4b). To identify putative GREs and their correspondence to genes they regulate, we first calculated the average expression of genes from snRNA-seq data and the average accessibility of peaks from snATAC-seq data for individual T-types or P-types. We then calculated the Pearson's correlation between the accessibility of peaks with the expression of

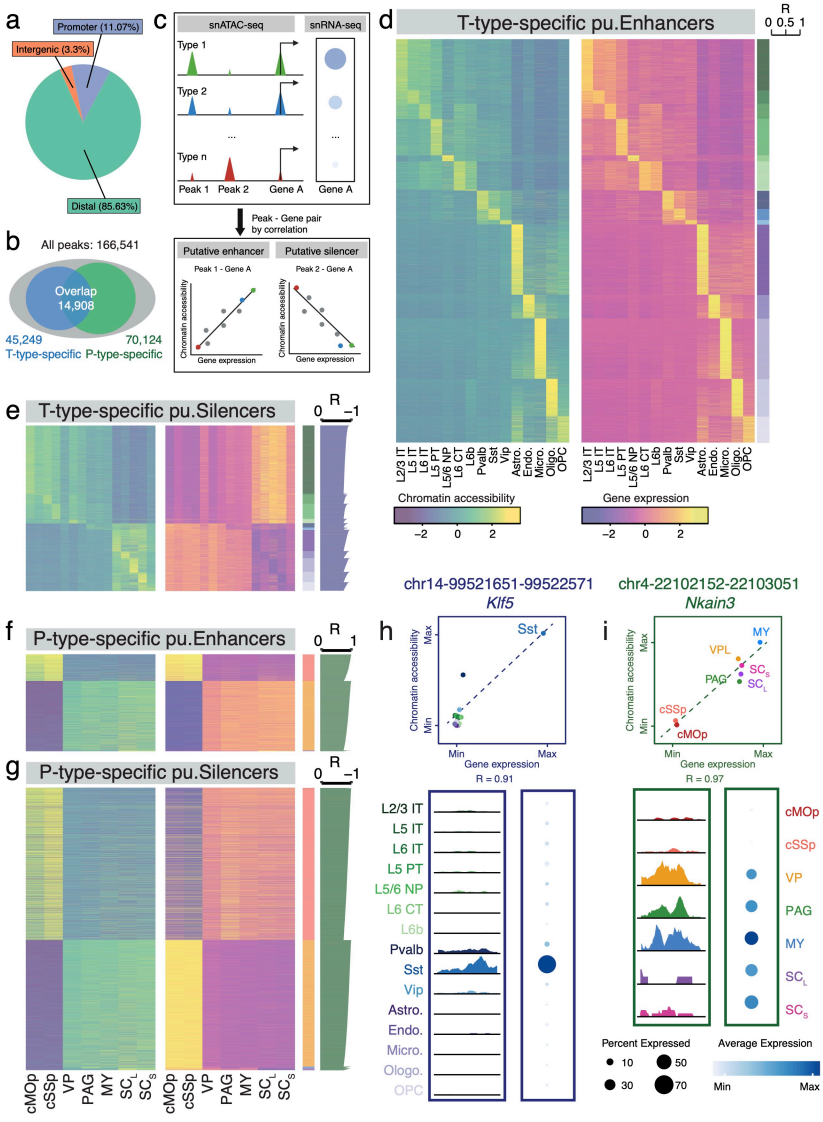


Figure 4. Genomic regulatory networks of cortical cell types using Projection-TAGs

a, Pie chart categorizing peaks in snATAC-seq based on their genomic locations. Peaks are mapped to one of the three categories: promoter regions (-1,000 bp to +100 bp of transcription start site [TSS]), distal regions (<200 kb upstream or downstream of TSS or within gene body, excluding promoter), and intergenic regions (>200kb upstream or downstream of TSS, excluding gene body) **b**, Distribution of DA peaks in any T-types and/or P-types across all peaks detected in snATAC-seq. **c**, Statistical workflow of identifying putative enhancers and silencers and their regulated genes. **d**, heatmaps of the accessibility of pu.Enhancer peaks in snATAC-seq data (left) and the expression of their regulated genes in snRNA-seq data (right) in individual T-types. On the right side next to the heatmaps, the color bar indicates the T-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual T-types. See Fig. 2a for the colors of T-types. **e**, Heatmaps of T-type-specific pu.Silencers and their regulated genes. **f** and **g**, heatmaps of P-type-specific puEnhancers and pu.Silencers and their regulated genes. On the right side next to the heatmaps, the color bar indicates the P-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual P-types. See Fig. 2b for the colors of P-types. **h**, Top: Scatter plot showing the correlation between the accessibility of peak chr14-99521651-99522571 and the expression of gene *Klf5* in individual T-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 chr14-99521651-99522571, displayed as the average fraction of transposase-sensitive fragments per nucleus in each T-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 *Klf5* in each T-type. Colors indicate T-types. **i**, Top: Scatter plot showing the correlation between the accessibility of peak chr4-22102152-22103051 and the expression of gene *Nkain3* in individual P-types (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr4-22102152-22103051, displayed as the average fraction of transposase-sensitive fragments per nucleus in each P-type (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Nkain3* in each P-type. Colors indicate P-types.

genes for any peak-gene pairs that are within 5M bp window of the same chromosome (Fig. 4c). A peak-gene pair with strong positive correlation (Pearson's r > 0.75) is identified as a putative enhancer (pu.Enhancer) and its putative regulated gene, whereas a pair with strong negative correlation (Pearson's r < -0.75) is categorized as a putative silencer (pu.Silencer) and its regulated gene. We identified 11,666 pu.Enhancers and 3,932 pu.Silencers for individual T-types (Fig. 4d,e), and 2,270 pu.Enahncers and 6,706 pu.Silencers for individual P-types (Fig. 4f,g). 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[78]. Second, these putative GREs show significant alignment with GREs previously annotated by the ENCODE consortium, particularly those identified in the mouse brain as opposed to other organs (Extended Data Fig. 9a)[79]. Furthermore, our analysis reveals four experimentally validated functional enhancers that coincide with peaks identified in our snATAC-seq data[73]. Notably, two of these validated enhancers overlap with the T-type-specific pu.Enhancers identified in our analysis, and the predicted T-types aligned with the cell types demonstrating the highest enhancer activity as verified experimentally (Extended Data Fig. 9b).

Among identified putative GREs, the peak 99,521,651-99,522,571 on chromosome 14 is ~14 kbp upstream of the TSS of *Klf5* gene (Fig. 4h). This peak, differentially accessible in SST GABAergic neurons, is positively correlated (Pearson's r = 0.91) to *Klf5*, which is differentially expressed in the same T- type, suggesting that this peak may contain a pu.Enhancer that might drive the expression of *Klf5* specifically in SST neurons. The accessibility of peak 22,102,152-22,103,051 on chromosome 4 is positively correlated to the expression of *Nkain3* in individual P-types (Fig. 4i), suggesting this peak may contain a pu.Enhancer that might drive the expression of *Nkain3* specifically in the ET-projecting neurons. Additionally, peaks chr19-14,870,895-14,871,826 and chr18-44,929,342-44,930,225 contain T-type-specific and P-type-specific pu.Silencers, respectively. While accessible in the genome, they may reduce the expression of *Fth1* in the IT-projecting neurons and *Kctd16* in ET-projecting neurons, respectively (Extended Data Fig. 9c,d). Thus, Projection-TAGs offer a powerful, high-throughput platform to perform systemic multiomic analyses to gain insight into the gene expression and chromatin accessibility profiles of diverse projection neurons.

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

Projection-TAGs allow us to elucidate the spatial, gene expression, and chromatin accessibility profiles of diverse projection neurons at single-cell resolution. As Projection-TAG AAVs mediate stable BC expression with low cellular toxicity, we examined if we could potentially leverage Projection-TAGs to obtain a snapshot of activity-dependent recruitment of distinct projection neurons and their molecular features in the context of a specific stimulus by combining Projection-TAGs with Act-seq[80-82]. This approach will link the activity of specific subsets of neuronal cell types and their projections following stimulus by surveying transcriptionally activated nuclei based on the immediate early genes (IEG) expression. The MOp and the SSp circuitry have been extensively implicated in the modulation of distinct pain modalities[83-88], we wanted to gain insight into the gene expression changes in the cell types and their projections neurons in response to visceral pain stimulus. To study the cell populations acutely activated by visceral pain, we first labeled the seven projection targets of the MOp and SSp as described earlier (Fig. 1a). On the day of experiment, we induced acute inflammatory visceral pain with cyclophosphamide (CYP) or inject saline as the control, followed by combinatorial snRNA-seq and snATAC-seq of the MOp and SSp 30 minutes after the stimulus, at which we observed significant spontaneous behaviors (Extended Data Fig. 10a).

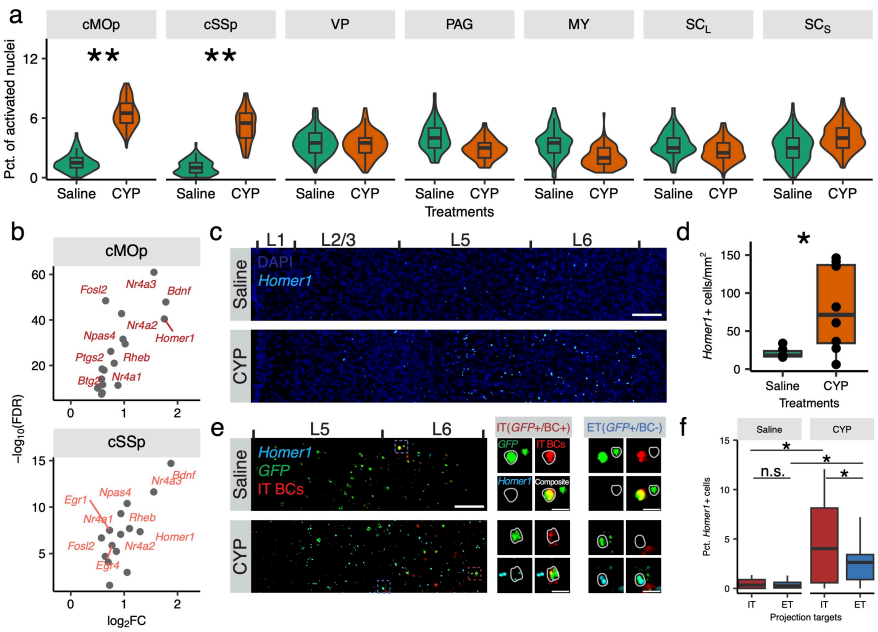


Figure 5. Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain **a**, Violin plots and box plots showing percentage of activate nuclei between treatments in individual P-types. For each P-type and treatment, 200 nuclei were randomly sampled to calculate the percentage and the process repeated for a total of 1,000 times. The *p* values are calculated as the fraction of repeats where percentage of activated nuclei from CYP-treated animals is lower than that of saline-treated animal. (***p* = 0.005 for cMOp, ***p* = 0.001 for cSSp, *p* = 0.4 for VP, *p* = 0.64 for PAG, *p* = 0.74 for MY, *p* = 0.62 for SC_L, *p* = 0.25 for SC_S). **b**, Volcano plots showing significantly induced IEGs (avg_log2FC > 0.5) in activated nuclei from CYP-treated mice compared to the same number of randomly sampled nuclei (with matched T-types) from saline-treated mice for the two IT P-types. **c**, Representative FISH images showing *Homer1* expression in MOp from brain sections (approximately -0.2 mm relative to bregma) of saline- and CYP-treated mice. Scale bar: 100 µm. **d**, Quantification of *Homer1*+ cells in MOp and SSp areas from brain sections from CYP- and saline-treated mice (six slices from 0 to -0.92 mm relative to bregma per group, **p* = 0.021, one-side t-test). **e**, Left: representative FISH images of layers 5 and 6 of MOp from the same slices in Fig. 4c showing the expression of *Homer1, GFP*, and two Projection-TAGs labeling neurons projecting to cMOp and cSSp. Scale bar: 100 µm. Right: IT-projecting neurons are defined as cells co-labeled with *GFP* and Projection-TAGs labelling cMOp and cSSp projections, whereas ET-projecting neurons are defined as cells labeled with *GFP* only. **f**, Quantification of *Homer1* expression in IT and ET-projecting neurons in MOp and SSp areas (Data collected from 6 cortical slices per group. **p* = 0.02, *F*(3,30) = 3.88, one-way ANOVA; *p=0.01 in IT between CYP and saline, *p=0.02 in ET between CYP and saline, *p* = 0.40 between IT and ET in saline, **p* = 0.04 between IT and ET in CYP, post-hoc one-sided t-tests with Bonferroni correction).

Nuclei of the same T-types from CYP- and saline-treated mice are transcriptionally and epigenetically similar, suggesting visceral pain did not significantly alter the gene expression or chromatin accessibility of MOp and SSp cell types at the acute time point (Extended Data Fig. 10b,c). To pinpoint the cell populations that are acutely activated by visceral pain, we compared the percentage of activated nuclei in individual T-types or P-types between CYP- and saline-treated mice. While the total percentage of activated nuclei in individual libraries did not significantly differ between treatments (Extended Data Fig. 10e), CYP significantly increased the percentage of activated nuclei in *Vip*-expressing GABAergic interneurons compared to saline (Extended Data Fig. 10f), suggesting preferential recruitment of interneuron populations following the acute visceral pain stimulus. Furthermore, CYP significantly activated nuclei in the cMOp- and cSSp-projecting neurons compared to saline, while none of the ET-projecting P-types exhibited significant activation (Fig. 5a). The data suggests that visceral pain selectively recruits neurons projecting to IT targets in the acute phase of the pain induction state.

We 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[88-92]. We identified 17 IEGs that are significantly induced in either cMOp- or cSSp-projecting activated neurons (Log2FC > 0.5, FDR < 0.05, comparing activated nuclei from CYP-treated mice to the same number of randomly sampled nuclei from saline-treated mice). Among them, *Nr4a3*, *Bdnf*, *Rheb*, and *Homer1* showed higher fold change (Log2FC > 1) in both IT-projecting P-types (Fig. 5b). Interestingly, CYP did not lead to significant *Fos* induction in snRNA-seq or FISH (Fig. 5b, Extended Data Fig. 10d). *Homer1* expression is directly linked to neuronal activity and synaptic plasticity[93-96]. 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 induced significant *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 injected Projection-TAG AAVs in the seven projection targets as described in Fig. 1a. We reasoned we could define IT projecting neurons by screening only for the BCs expressed from the neurons projecting to cMOp and cSSp, whereas GFP+/BC- as ET-projecting neurons (Fig. 5e).

401 Using this approach, we found that CYP induced *Homer1* expression in both IT- and ET-projecting
402 neurons. However, IT-projection neurons exhibited higher magnitude of *Homer1* induction, as percentage
403 of *Homer1*+ cells increased by 9-fold in IT-projecting neurons, compared to 4.6-fold in the ET-projecting
404 neurons (Fig. 5f). Notably, we did not notice any difference of *Homer1* expression in IT- and ET-projecting
405 neurons from saline-treated mice. In summary, our data shows that Projection-TAGs could be readily
406 implemented with transcriptional analysis such Act-seq to identify the molecular markers and recruitment
407 of distinct specific brain wide projections and cell types following the stimulus of interest.

408 Discussion

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

423 Projection-TAGs could potentially be compatible with additional commercial platforms, such as
424 high-throughput spatial transcriptomics platforms like Xenium and Visium, that would allow spatially-
425 resolved investigation of diverse cell types and distinct anatomical organization of the projection neurons.
426 Another potential application of Projection-TAGs includes connecting neuronal activity to the distinct
427 projections by integrating the oScarlet version with in vitro and in vivo calcium imaging experiments. The
428 cell type and projection information from imaging-based molecular assays can be integrated with the real-
429 time neuronal activity information from the calcium imaging experiments via post-hoc imaging alignment
430 and registration[89,97-99].

431 The number of projections that can be labeled with Projection-TAG AAVs is not inherently
432 constrained, and Projection-TAGs could be easily scalable as we have screened 50 BCs that could
433 potentially be packaged to increase multiplexing of projection tagging. Efficient labeling of Projection-
434 TAGs relies on successful targeting of desired brain regions and transduction efficacy of AAVrg viruses.
435 It has been well appreciated that AAV serotypes have distinct tropism and selective labeling of the brain
436 nuclei. With the discovery of capsid selection using directed evolution[100-102] and advances in sequencing
437 techniques and computational tools to optimize exogenous transcript detection[103,104], we anticipate that
438 the improved retrograde labeling efficiency of viruses will enhance wider applications of Projection-TAGs.
439 Given its flexibility, usability, and compatibility with commercial platforms, we envision that Projection-
440 TAGs can be readily applied to study diverse projection types in the central and peripheral nervous
441 system.

442

443 # Methods

444 **Generation of Projection-TAG BCs**

445 We used 100-bp BCs because they allow us to design HCR probes to detect the spatial distribution of
446 BC transcripts, and would improve their detection rate in RNA-seq compared to shorter BCs that are
447 commonly used. We first retrieved 60 previously reported BCs[54]. We then filtered out the BCs that
448 contain the sequence of restriction enzymes (BamHI, HpaI, and NotI). Next, we performed sequence
449 alignment using blastn suite searching in the "Nucleotide Collection (nr/nt)" database and
450 "refseq_representative_genomes" database against the genomes of human (taxid:9606), mouse
451 (taxid:10090), rat (taxid:10116), and Primates (taxid:9443) and filtered out the ones that showed
452 significant similarities[105]. We then calculated the Euclidean distance between any of the two BCs using
453 DistanceMatrix() function in R package DECIPHER. We then cataloged and reported the first 50 BCs
454 sorted based on the average Euclidean distances with other BCs.

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

456 To generate the backbone of pAAV-CAG-Sun1-GFP-WPRE-pA, we first linearized two previously
457 reported plasmids: pAAV-CAG-H2B-GFP (Addgene, Plasmid #116869)[27] using HindIII and pAAV-Ef1a-
458 DIO-Sun1GFP-WPRE-pA (Addgene, Plasmid #160141)[106] using AscI, and end filled. We then digested
459 them with SpeI and NheI, respectively, followed by gel extraction of the fragments at around 4,700 bp
460 (pAAV-CAG backbone) and 2,400 bp (Sun1-GFP), respectively. We next generated the plasmid pAAV-
461 CAG-Sun1-GFP-WPRE-pA by ligating the two fragments together. To insert BCs into pAAV-CAG-Sun1-
462 GFP-WPRE-pA, we first generated a gene block fragment for each BC with the following structure (5'
463 overhang-NotI-BC-SV40 polyA-HpaI-3' overhang) and sequence (AAGGAAAAAAgcggccgc-100bp
464 sequence of BC-
465 ttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt
466 gctttatttgtaaccattataagctgcaataaacaagttaacAACCGCTGCCG). We then digested both pAAV-CAG-
467 Sun1-GFP-WPRE-pA and BC gene blocks with NotI and HpaI, and generated pAAV-CAG-Sun1-GFP-
468 WPRE-BC-pA by ligating the fragments. We generated 12 plasmids each expressing a unique BC (1 -
469 12). We packaged the plasmids into AAV2retro (BrainVTA, China) and generated a set of AAVs each
470 expressing a unique BC. To generate a set of 12 plasmids pAAV-CAG-oScarlet-WPRE-BC-pA each
471 expressing a unique BC, we generated BamHI-oScarlet-WPRE-NotI gene fragment and replaced Sun1-
472 GFP-WPRE fragment in pAAV-CAG-Sun1-GFP-WPRE-BC-pA using BamHI and NotI. All restriction
473 enzymes were purchased from New England Biolabs, and all plasmids produced in this study are in the
474 process of being deposited to Addgene.

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

476 We acquired human embryo kidney cell line HEK 293T/17 (ATCC) and cultured it using DMEM
477 (Corning), supplemented with 10% Fetal Bovine Serum (Gibco), and 1% Penicillin/Streptomycin
478 (Sigma), on either 60 mm dishes for RNA-seq or 8-well chambered slides (Nunc™ Lab-Tek™) pre-
479 coated with PDL-collagen for multiplexed FISH. Once the cells reached 70 – 80% confluency, we
480 transfected the DNA of one BC plasmid into each of the HEK samples using lipofectamine 2000
481 (Invitrogen) according to the manual. Cells were incubated for 24 – 36 hours before analysis. For RNA-
482 seq, around 1 million cells were used for each sample and RNA was extracted and purified using
483 RNAqueous™ Total RNA Isolation Kit (Invitrogen) following the manual. Library preparation and
484 sequencing were conducted by the McDonnel Institute of Genomics at Washington University School of
485 Medicine. For FISH, cells were washed with DPBS and fixed using 4% formaldehyde for 10 minutes 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-hour light-dark cycle (6:00 am to 6:00 pm) 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 experiments.

Stereotaxic surgeries

Mice were given a single subcutaneous injection of 0.5ml 0.9% sterile sodium chloride and Buprenorphine-SR, 1 hour 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 BC), in single animals (Supplementary Table 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 minutes prior to needle withdrawal. needles are changed between each Projection-TAG injection to avoid contamination between Projection TAGs. Stereotaxic surgery injecting into SC_L and SC_S was performed at week 0. For SC_L 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_S 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 (550–580mm 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 MOp, SSp, VP, PAG, and MY using the following coordinates relative to Bregma: MOp (anterior-posterior [AP] -0.52 mm, medial-lateral [ML] +1.81 mm, dorsal-ventral [DV] -0.86 mm, left hemisphere), SSp (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 (approximately 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.

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

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

532 **Tissue processing**

533 In the experiments reported in this paper, we traced the seven projection targets of the MOp and SSp
534 by performing two stereotaxic surgeries using undiluted AAVs in 8-10 weeks-old C57BL/6 female mice.
535 Individual BC-expressing AAVs were injected into the SCL and SCS at week zero, followed by a second
536 surgery of AAV injections into the MOp, SSp, VP, PAG, and MY at week two. Animals were perfused,
537 cortical samples were collected for analysis at week five, and the rest of the brain was sectioned to
538 confirm the injection sites (Fig. 1a, Extended Data Fig. 4a,b).

539 To collect biological samples for FISH studies, animals were anesthetized with ketamine
540 cocktail, and perfused with DEPC-PBS followed by 4% DEPC-PFA. Brain and spinal cord tissues were
541 dissected and incubated in 4% DEPC-PFA at 4 C for 6 - 8 hours. Tissues were then incubated in 30%
542 sucrose in 1X DEPC-PBS at 4 C for 24 - 48 hours until they sank to the bottom of the tube. Tissues
543 were then embedded in OCT and stored at -80℃ before slicing. Tissues were sliced coronally into
544 sections with 30 μm thickness using a cryostat (Leica), and tissue slices were mounted on microscope
545 slides. Slides were stored at -20℃ for a least one hour before moving to -80℃ for long-term storage.
546 Slides containing the brain regions of injections were examined under a fluorescent microscope, and
547 the fluorescent signal from Sun1-GFP was used to confirm the injection sites. As an optional step, the
548 native fluorescence of Sun1-GFP can be photobleached from mouse brain sections by floating slices in
549 1×PBS with 24 mM NaOH and 4.5% H_2O_2 and exposing under the UV light (27 total watts, OPPSK) for
550 30 minutes at room temperature. Slices were immediately rinsed with DEPC-PBS twice before
551 mounting on a microscope slide. To collect biological samples for molecular experiments such as
552 qPCR or sequencing, animals were anesthetized with ketamine cocktail, and perfused with ice-cold
553 NMDG-based cutting solution (NMDG 93 mM, KCl 2.5 mM, NaH_2PO_4 1.25 mM, $NaHCO_3$ 30 mM,
554 HEPES 20 mM, Glucose 25 mM, Ascorbic acid 5 mM, Tiourea 2 mM, Sodium Pyruvate 3 mM, $MgSO_4$
555 10 mM, $CaCl_2$ 0.5 mM, N-acetylcysteine 12 mM; pH adjusted to 7.3 with 12N HCl, and bubbled with
556 95% O_2 and 5% CO_2). The spinal cord was dissected, fixed, and sectioned by following the sample
557 preparation process mentioned above in order to confirm the injection sites of SC_L and SC_S. The brain
558 was submerged in ice-cold NMDG-based cutting solution and sliced coronally into sections with 400 μm
559 thickness using a Compresstome (Precisionary, VF-210-0Z). MOp and SSp samples were prepared by
560 micro-dissecting the respective regions under a microscope and collected into a nucleus-free centrifuge
561 tube placed on dry ice. Samples were stored at -80℃ for long-term storage, and the rest of the brain
562 sections were placed on microscope slides and the fluorescent signal from Sun1-GFP was examined
563 under a fluorescent microscope to confirm injection sites.

564 **Multiplexed FISH with HCR**

565 HCR (hybridization chain reaction) v3.0 probes, amplifiers, and reagents were purchased from
566 Molecular Instruments[107]. Multiplexed FISH was performed according to the manual of HCR RNA-FISH
567 with minor modifications. For hybridization, samples were equilibrated in hybridization buffer (500 ml)
568 for 30 min at 37℃ and hybridized with probe sets (16 nM each probe) in hybridization buffer overnight
569 at 37℃. Samples were washed in probe wash buffer and gradually switched to 5×SSCT (5×SSC,
570 0.1%Tween-20) at 37℃. HCR was carried out at room temperature. Samples were equilibrated in
571 amplification buffer for 30 min and the amplifier hairpins (6 pmol each hairpin) (conjugated to Alexa-
572 488, Alexa-546, Alexa-647, and Cyanine 7) were heated to 95℃ and snap-cooled in a dark drawer for
573 30 min. Hairpins were mixed and diluted to 0.6 nM each hairpin in amplification buffer. Samples were

then incubated with hairpins 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 and 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℃. Samples were washed with 5×SSCT for 5 minutes 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, 535/50m), Alexa-546 (Chroma 49304, 546/10x, 572/23m), Alexa-647 (Chroma 49006, 620/60x, 700/75m), and Cyanine 7 (Chroma 49007, 710/75x, 810/80m). 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[108]. 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[109,110]. 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[111]. 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 illastik segmentation algorithms was validated against manual segmentation. Illastik accurately identified 96.8% of manually segmented cells with a 6.6% false positive rate (Extended Data Fig. 5b). 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[112]. To identify cells that 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 illastik. If the Euclidian distance of two objects from separate channels is less than the average radius of all objects, the two objects were identified as one cell that co-expresses both marker genes. For BC detection and specificity analysis, cSSp and cMOp were excluded from this analysis because they are known to contain neurons projecting to other injection sites and may be labeled by BCs injected into other regions which could contaminate the final distribution of the BC.

617 **Single-nuclei isolation and FACS**

618 Nuclear extraction was performed according to the protocol described previously with minor
619 modifications[113]. Mouse cortical tissues were transferred to a dounce homogenizer in homogenization
620 buffer (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl, pH 8.0, 5 ug/mL actinomycin, 1%
621 BSA, and 0.08 U/ul RNase inhibitor, 0.01% NP40) on ice. Samples were homogenized for 10 strokes
622 with the loose pestle in a total volume of 1 mL, followed by 10 additional strokes with the tight pestle.
623 The tissue homogenate was then passed through a 50 µm filter and diluted 1:1 with working solution
624 (50% iodixanol, 25 mM KCl, 5 mM MgCl2, and 10 mM Tris-HCl, pH 8.0). Nuclei were layered onto an
625 iodixanol gradient after homogenization and ultracentrifuged as described previously. After
626 ultracentrifugation, nuclei were collected between the 30 and 40% iodixanol layers and diluted with
627 resuspension buffer (1xPBS with 1% BSA, and 0.08 U/ul RNase inhibitor). Nuclei were centrifuged at
628 500 g for 10 min at 4°C and resuspended in resuspension buffer with 5 ng/ul of 7-AAD. For FACS,
629 gates on GFP and 7AAD were set using tissues collected from mice with no stereotaxic surgeries and
630 no 7AAD staining. GFP+/7-AAD+ events and GFP-/7-AAD+ events were sorted using 100 µm nozzle
631 on a BD FACSARIA II and collected separately into 1.5 mL microcentrifuge tubes containing 100 µl of
632 resuspension buffer.

633 **snRNA-seq and snATAC-seq**

634 For snRNA-seq, nuclei were further processed and sequenced according to the manufacturer's
635 manuals of Parse-biosciences Evercode WT mini Kit V2 (Parse) or 10X Genomics Chromium Single
636 Cell Gene Expression 3' V3.1 Assay (10X RNA). For combinatorial snRNA-seq and snATAC-seq, nuclei
637 were processed and sequenced according to the manufacturer's manual of 10X Genomics Chromium
638 Single Cell Multiome Assay (10X Multiome). Libraries were sequenced on a NovaSeq6000 with 150
639 cycles each for Read1 and Read2, targeting 100,000 paired reads/nucleus for snRNA-seq libraries and
640 50,000 paired reads/nucleus for snATAC-seq libraries. Raw sequencing data from individual libraries
641 were processed and mapped to the mouse genome GRCm38 (modified by adding the sequences of
642 Sun1-GFP, WPRE, and BC1-12 [Supplementary Table 1]) using 10X Genomics cellranger-6.1.2 (10X
643 RNA), 10X Genomics cellranger-arc-2.0.1 (10X Multiome), or Parse Computational Pipeline v1.1.2
644 (Parse). For snATAC-seq, accessible peaks were identified using cellranger-arc aggr by analyzing the
645 combined fragment signals across all snATAC-seq libraries in the dataset.

646 **Quality control, clustering, T-type and P-type annotation of snRNA-seq**

647 The gene-cell count matrices from all snRNA-seq libraries were concatenated using R package Seurat
648 (V4.3.0)[114]. To be included in the snRNA-seq analysis, nuclei were required to contain at least 500
649 unique genes, less than 15,000 total UMIs, and fewer than 2% of the counts deriving from
650 mitochondrial genes. There were 69,657 nuclei that met these criteria. Raw counts were scaled to
651 10,000 transcripts per nucleus using NormalizeData() function to control the sequencing depth between
652 nuclei. Counts were centered and scaled for each gene using ScaleData() function. Highly variable
653 genes were identified using FindVariableFeatures() and the top 20 principal components were retrieved
654 with RunPCA() using default parameters. For dimension reduction and visualization, Uniform Manifold
655 Approximation and Projection (UMAP) coordinates were calculated using RunUMAP(). Nuclei clustering
656 was performed using FindClusters() based on the variable features from the top 20 principal
657 components, with the resolution set at 0.6, and the marker genes for each cluster was identified using
658 FindAllMarkers() comparing nuclei in one cluster to all other nuclei. Doublet or low-quality nuclei were
659 identified if they meet any of the following criteria: 1). Assigned to a cluster with no significantly enriched
660 marker genes (FDR < 0.05, log2FC > 1); 2). Assigned to a cluster in which five or more mitochondrial
661 genes were identified among top 20 marker genes (sorted by avg_log2FC); 3). Identified as doublets

662 using R package DoubletFinder with doublet expectation rate at 5%[115]. In total 5,991 nuclei were
663 identified as doublet or low-quality and thus excluded from the dataset. The remaining 63,666 nuclei
664 were clustered as described above. Transcriptional cell types (T-types) were assigned to each cluster
665 based on the canonical marker genes previously reported (Extended Data Fig. 6f). For annotation of
666 projection targets (P-types), individual BCs expressing (> 0 UMIs) in each nucleus are first identified,
667 and the P-types are assigned to each nucleus based on the injection sites of its expressing BCs in a
668 specific experiment. A nucleus may not be assigned any p-types if no BC is expressed or multiple P-
669 types if more than one BC is expressed.

670 **Anchoring analysis of snRNA-seq**

671 To validate the T-type assignment, we directly compare our snRNA-seq data to a published scRNA-seq
672 of mouse anterior lateral motor cortex previously described[44]. We used Seurat to anchor our snRNA-
673 seq data to the published dataset. First, FindTransferAnchors() was used to identify anchors
674 (conserved features) between datasets. TransferData() was run to transfer cell type labels described in
675 the published dataset to each nucleus in the our snRNA-seq data.

676 **Quality control and clustering of snATAC-seq**

677 The peak-cell counts matrix was loaded and analyzed using R package Signac[116]. To be included for
678 snATAC-seq analysis, nuclei were required to be present in snRNA-seq data and contain at least 1,000
679 fragments overlapping with peaks. There were 41,553 nuclei that met these criteria. Term frequency-
680 inverse document frequency normalization was perfomed using RunTFIDF () function and variable
681 peaks were identified using FindTopFeatures (). Dimension reduction was performed with singular
682 value decomposition using RunSVD() function and Uniform Manifold Approximation and Projection
683 (UMAP) coordinates were calculated using RunUMAP(). Nuclei clustering was performed using
684 FindClusters() based on the top 20 dimensions with the resolution of 0.8.

685 **Differential analyses of gene expression and chromatin accessibility**

686 To identify marker genes and peaks that are enriched in distinct T-types or P-types, differential
687 expression/accessibility analysis was performed using findAllMarkers() in Seurat/Signac, comparing
688 nuclei from T-type or P-type to all other nuclei. Genes and peaks with FDR < 0.05 were reported. To
689 identify genes and peaks that are differentially expressed/accessible between any two T-types or P-
690 types, differential expression/accessibility analysis was performed using findMarkers() in Seurat/Signac,
691 comparing nuclei from one T-type or P-type to nuclei from another T-type or P-type. Genes and peaks
692 with FDR < 0.05 were reported.

693 To identify genes and peaks that are differentially regulated by CYP in individual T-types or P-
694 types, differential expression/accessibility analysis was performed using findMarkers() in Seurat/Signac,
695 comparing nuclei from CYP-treated animals to nuclei from Saline-treated animals of the same T-type or
696 P-type. Genes and peaks with avg_log2FC > 0.5 and FDR < 0.05 were reported. To identify IEGs that
697 are activated by CYP in P-types of IT targets, differential expression analysis was performed using
698 findMarkers() in Seurat, comparing activated nuclei defined by Act-seq from CYP-treated animals to the
699 same number of randomly sampled nuclei (with matched T-type distribution) from Saline-treated
700 animals in each P-type. IEGs with avg_log2FC > 0.5 and FDR < 0.05 were reported.

701 **Identification of putative genomic regulatory elements**

702 To identify snATAC-seq peaks that are correlated with gene expression in cis and may act as genomic
703 regulatory elements (GREs) to regulate the expression of target genes, we modified a computational
704 pipeline previously described[117]. we first calculated, using snATAC-seq data, the average accessibility

705 of each peak and, using snRNA-seq data, the average expression of each gene in individual T-types or
706 P-types. We then computed the Pearson correlation coefficient r between the accessibility of peaks and
707 the expression of genes across P-types or T-types for any peak-gene pair, in which the center of the
708 peak is located within 5 mbps of the center of the TSS of the gene on the same chromosome. To
709 identify putative GREs that may direct T-type-specific gene expression, we identified pairs of T-type-
710 specific snATAC-seq peaks (avg_log2FC > 0.5, FDR < 0.05, comparing peak accessibility in nuclei of
711 one T-type to all others) and T-type-specific snRNA-seq genes (avg_log2FC > 0.5, FDR < 0.05,
712 comparing gene expression in nuclei of one T- type to all others) in the same T- type. A peak-gene pair
713 with a strong positive correlation (Pearson's r > 0.75) is identified as a putative enhancer (pu.Enhancer)
714 and its putative regulated gene, whereas a pair with a strong negative correlation (Pearson's r < -0.75)
715 is categorized as a putative silencer (pu.Silencer) and its regulated gene. To identify putative GREs that
716 may direct P-type-specific gene expression, we identified pairs of P-type-specific snATAC-seq peaks
717 (avg_log2FC > 0.5, FDR < 0.05, comparing peak accessibility in nuclei of one P-type to all others) and
718 P-type-specific snRNA-seq genes (avg_log2FC > 0.5, FDR < 0.05, comparing gene expression in
719 nuclei of one P-type to all others) in the same P-type. P-type-specific pu.Enhancers and pu.Silencers
720 and their regulated genes are identified as described above.

721 **Act-seq analysis**

722 To identify nuclei that are activated by visceral pain, the activity score for each nucleus in snRNA-seq
723 data was calculated using AddModuleScore() in Seurat based on the expression of a set of 139
724 immediate early genes (IEG) previously described[80]. For T-type analysis, a nucleus was considered
725 transcriptionally ''activated'' if its activity score was 2 standard deviations higher than the average
726 activity scores across all nuclei from the saline-treated animals of the same T-type. To identify T-types
727 or P-types that are activated by visceral pain, 200 nuclei per P-type or T-type per treatment were
728 randomly sampled and the percentage of activated nuclei per P-type or T-type per treatment was
729 calculated. This comparison was repeated for a total of 1,000 times. The *p* values are calculated as the
730 fraction of repeats where the percentage of activated nuclei from CYP-treated animals is lower than
731 that of the saline-treated animals. A nucleus is considered transcriptionally 'activated' if its activity
732 score, calculated by aggregating the expression of IEGs, is greater than the threshold calculated based
733 on the scores of nuclei from saline-treated samples.

734 **Statistical analysis and visualization**

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

741 Data availability

742 Raw and processed data of snRNA-seq and snATAC-seq experiments included in this study are in the
743 process of being deposited to the NCBI Gene Expression (GEO) SRA with accession number XXX.

744 Acknowledgements

745 This work was funded by National Institute on Drug Abuse (R01DA056829) and National Institute of
746 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; F.L., H.H, T.O, Y.Z, collected the samples for the library preparation and *in situ* validation studies. 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 comments from all authors.

Competing interest

The authors declare that they have no conflict of interest.

References

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

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

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

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

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

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

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

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

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

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

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

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

13 Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030 e1016 (2018). https://doi.org:10.1016/j.cell.2018.07.028
14 Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat Neurosci* **22**, 1182-1195 (2019). https://doi.org:10.1038/s41593-019-0417-0
15 Gouwens, N. W. *et al.* Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* **183**, 935-953 e919 (2020). https://doi.org:10.1016/j.cell.2020.09.057
16 Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317-332 (2023). https://doi.org:10.1038/s41586-023-06812-z
17 Zhang, M. *et al.* Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343-354 (2023). https://doi.org:10.1038/s41586-023-06808-9
18 Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333-342 (2023). https://doi.org:10.1038/s41586-023-06818-7
19 Liu, H. *et al.* Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* **624**, 366-377 (2023). https://doi.org:10.1038/s41586-023-06805-y
20 Zu, S. *et al.* Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378-389 (2023). https://doi.org:10.1038/s41586-023-06824-9
21 Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552-561 (2023). https://doi.org:10.1038/s41586-023-06569-5
22 Winter, C. C. *et al.* A transcriptomic taxonomy of mouse brain-wide spinal projecting neurons. *Nature* **624**, 403-414 (2023). https://doi.org:10.1038/s41586-023-06817-8
23 Zhou, J. *et al.* Brain-wide correspondence of neuronal epigenomics and distant projections. *Nature* **624**, 355-365 (2023). https://doi.org:10.1038/s41586-023-06823-w
24 Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195-202 (2019). https://doi.org:10.1038/s41586-019-1716-z
25 Munoz-Castaneda, R. *et al.* Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159-166 (2021). https://doi.org:10.1038/s41586-021-03970-w
26 Gao, L. *et al.* Single-neuron projectome of mouse prefrontal cortex. *Nat Neurosci* **25**, 515-529 (2022). https://doi.org:10.1038/s41593-022-01041-5
27 Economo, M. N. *et al.* Distinct descending motor cortex pathways and their roles in movement. *Nature* **563**, 79-84 (2018). https://doi.org:10.1038/s41586-018-0642-9
28 Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137-143 (2021). https://doi.org:10.1038/s41586-021-03705-x
29 Kim, E. J. *et al.* Extraction of Distinct Neuronal Cell Types from within a Genetically Continuous Population. *Neuron* **107**, 274-282 e276 (2020). https://doi.org:10.1016/j.neuron.2020.04.018
30 Kaspar, B. K. *et al.* Targeted retrograde gene delivery for neuronal protection. *Mol Ther* **5**, 50-56 (2002). https://doi.org:10.1006/mthe.2001.0520
31 Tervo, D. G. *et al.* A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron* **92**, 372-382 (2016). https://doi.org:10.1016/j.neuron.2016.09.021
32 Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat Methods* **4**, 47-49 (2007). https://doi.org:10.1038/nmeth999
33 Reardon, T. R. *et al.* Rabies Virus CVS-N2c(DeltaG) Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron* **89**, 711-724 (2016). https://doi.org:10.1016/j.neuron.2016.01.004
34 Soudais, C., Laplace-Builhe, C., Kissa, K. & Kremer, E. J. Preferential transduction of neurons by canine adenovirus vectors and their efficient retrograde transport in vivo. *FASEB J* **15**, 2283-2285 (2001). https://doi.org:10.1096/fj.01-0321fje
35 Stoeckel, K., Schwab, M. & Thoenen, H. Role of gangliosides in the uptake and retrograde axonal transport of cholera and tetanus toxin as compared to nerve growth factor and wheat germ agglutinin. *Brain Res* **132**, 273-285 (1977). https://doi.org:10.1016/0006-8993(77)90421-8
36 Vercelli, A., Repici, M., Garbossa, D. & Grimaldi, A. Recent techniques for tracing pathways in the central nervous system of developing and adult mammals. *Brain Res Bull* **51**, 11-28 (2000). https://doi.org:10.1016/s0361-9230(99)00229-4

37 Ju, G., Han, Z. S. & Fan, L. Z. Fluorogold as a retrograde tracer used in combination with immunohistochemistry. *J Neurosci Methods* **29**, 69-72 (1989). https://doi.org:10.1016/0165-0270(89)90109-x
38 Lanciego, J. L. & Wouterlood, F. G. A half century of experimental neuroanatomical tracing. *J Chem Neuroanat* **42**, 157-183 (2011). https://doi.org:10.1016/j.jchemneu.2011.07.001
39 Schulte, S. J., Fornace, M. E., Hall, J. K. & Pierce, N. A. HCR spectral imaging: 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent samples. *bioRxiv* (2023). https://doi.org:10.1101/2023.08.30.555626
40 Wang, Y. *et al.* Multimodal mapping of cell types and projections in the central nucleus of the amygdala. *Elife* **12** (2023). https://doi.org:10.7554/eLife.84262
41 Bienkowski, M. S. *et al.* Integration of gene expression and brain-wide connectivity reveals the multiscale organization of mouse hippocampal networks. *Nat Neurosci* **21**, 1628-1643 (2018). https://doi.org:10.1038/s41593-018-0241-y
42 Baek, M., Menon, V., Jessell, T. M., Hantman, A. W. & Dasen, J. S. Molecular Logic of Spinocerebellar Tract Neuron Diversity and Connectivity. *Cell Rep* **27**, 2620-2635 e2624 (2019). https://doi.org:10.1016/j.celrep.2019.04.113
43 Phillips, J. W. *et al.* A repeated molecular architecture across thalamic pathways. *Nat Neurosci* **22**, 1925-1935 (2019). https://doi.org:10.1038/s41593-019-0483-3
44 Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78 (2018). https://doi.org:10.1038/s41586-018-0654-5
45 Zhang, Z. *et al.* Epigenomic diversity of cortical projection neurons in the mouse brain. *Nature* **598**, 167-173 (2021). https://doi.org:10.1038/s41586-021-03223-w
46 Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86-102 (2021). https://doi.org:10.1038/s41586-021-03950-0
47 Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975-987 (2016). https://doi.org:10.1016/j.neuron.2016.07.036
48 Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772-786 e719 (2019). https://doi.org:10.1016/j.cell.2019.09.023
49 Klingler, E. *et al.* Temporal controls over inter-areal cortical projection neuron fate diversity. *Nature* (2021). https://doi.org:10.1038/s41586-021-04048-3
50 Chen, Y. *et al.* High-throughput sequencing of single neuron projections reveals spatial organization in the olfactory cortex. *Cell* **185**, 4117-4134 e4128 (2022). https://doi.org:10.1016/j.cell.2022.09.038
51 Sun, Y. C. *et al.* Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24**, 873-885 (2021). https://doi.org:10.1038/s41593-021-00842-4
52 Kuramoto, E. Method for labeling and reconstruction of single neurons using Sindbis virus vectors. *J Chem Neuroanat* **100**, 101648 (2019). https://doi.org:10.1016/j.jchemneu.2019.05.002
53 Zhao, Q. *et al.* A multidimensional coding architecture of the vagal interoceptive system. *Nature* (2022). https://doi.org:10.1038/s41586-022-04515-5
54 Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat Biotechnol* **39**, 865-876 (2021). https://doi.org:10.1038/s41587-021-00837-3
55 Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat Neurosci* **18**, 170-181 (2015). https://doi.org:10.1038/nn.3917
56 Beloozerova, I. N., Sirota, M. G. & Swadlow, H. A. Activity of different classes of neurons of the motor cortex during locomotion. *J Neurosci* **23**, 1087-1097 (2003). https://doi.org:10.1523/JNEUROSCI.23-03-01087.2003
57 Pennartz, C. M. *et al.* Corticostriatal Interactions during Learning, Memory Processing, and Decision Making. *J Neurosci* **29**, 12831-12838 (2009). https://doi.org:10.1523/JNEUROSCI.3177-09.2009
58 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142 (2015). https://doi.org:10.1126/science.aaa1934

59 Yao, Z. *et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103-110 (2021). https://doi.org:10.1038/s41586-021-03500-8
60 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-346 (2016). https://doi.org:10.1038/nn.4216
61 Bishop, H. I. *et al.* Distinct Cell- and Layer-Specific Expression Patterns and Independent Regulation of Kv2 Channel Subtypes in Cortical Pyramidal Neurons. *J Neurosci* **35**, 14922-14942 (2015). https://doi.org:10.1523/JNEUROSCI.1897-15.2015
62 Hunnicutt, B. J. *et al.* A comprehensive thalamocortical projection map at the mesoscopic level. *Nat Neurosci* **17**, 1276-1285 (2014). https://doi.org:10.1038/nn.3780
63 J. L. MacDonald, R. M. F., E. Azim, S. J. Shnider, B. J. Molyneaux, P. Arlotta, J. D. Macklis. in *Patterning and Cell Type Specification in the Developing CNS and PNS* 475-502 (Elsevier, 2013).
64 Macias, M., Lopez-Virgen, V., Olivares-Moreno, R. & Rojas-Piloni, G. Corticospinal neurons from motor and somatosensory cortices exhibit different temporal activity dynamics during motor learning. *Front Hum Neurosci* **16**, 1043501 (2022). https://doi.org:10.3389/fnhum.2022.1043501
65 Kawaguchi, Y. Pyramidal Cell Subtypes and Their Synaptic Connections in Layer 5 of Rat Frontal Cortex. *Cereb Cortex* **27**, 5755-5771 (2017). https://doi.org:10.1093/cercor/bhx252
66 Buenrostro, J. D. *et al.* in *Nature* Vol. 523 486-490 (Nature Publishing Group, 2015).
67 Nelson, A., Abdelmesih, B. & Costa, R. M. Corticospinal populations broadcast complex motor signals to coordinated spinal and striatal circuits. *Nat Neurosci* **24**, 1721-1732 (2021). https://doi.org:10.1038/s41593-021-00939-w
68 Zemke, N. R. *et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390-402 (2023). https://doi.org:10.1038/s41586-023-06819-6
69 Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-154 (2015). https://doi.org:10.1038/nrm3949
70 Halfon, M. S. Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends Genet* **36**, 149-151 (2020). https://doi.org:10.1016/j.tig.2019.12.005
71 Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**, 1 (2012). https://doi.org:10.1186/1756-8935-5-1
72 Hrvatin, S. *et al.* in *eLife* Vol. 8 (eLife Sciences Publications Ltd, 2019).
73 Mich, J. K. *et al.* Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep* **34**, 108754 (2021). https://doi.org:10.1016/j.celrep.2021.108754
74 Graybuck, L. T. *et al.* Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449-1464 e1413 (2021). https://doi.org:10.1016/j.neuron.2021.03.011
75 Lambert, J. T. *et al.* Parallel functional testing identifies enhancers active in early postnatal mouse brain. *Elife* **10** (2021). https://doi.org:10.7554/eLife.69479
76 Salimando, G. J. *et al.* Human OPRM1 and murine Oprm1 promoter driven viral constructs for genetic access to mu-opioidergic cell types. *Nat Commun* **14**, 5632 (2023). https://doi.org:10.1038/s41467-023-41407-2
77 Lawler, A. J. *et al.* Machine learning sequence prioritization for cell type-specific enhancer design. *Elife* **11** (2022). https://doi.org:10.7554/eLife.69571
78 Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007-1012 (2014). https://doi.org:10.1126/science.1246426
79 Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**, D882-D889 (2020). https://doi.org:10.1093/nar/gkz1062
80 Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. in *Neuron* Vol. 96 313-329.e316 (Cell Press, 2017).
81 Hrvatin, S. *et al.* in *Nature Neuroscience* Vol. 21 120-129 (2018).
82 Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat Biotechnol* **39**, 30-34 (2021). https://doi.org:10.1038/s41587-020-0605-1
83 Zhu, X. *et al.* Somatosensory cortex and central amygdala regulate neuropathic pain-mediated peripheral immune response via vagal projections to the spleen. *Nat Neurosci* **27**, 471-483 (2024). https://doi.org:10.1038/s41593-023-01561-8

84 Ziegler, K. *et al.* Primary somatosensory cortex bidirectionally modulates sensory gain and nociceptive behavior in a layer-specific manner. *Nat Commun* **14**, 2999 (2023). https://doi.org:10.1038/s41467-023-38798-7
85 Chen, T. *et al.* Top-down descending facilitation of spinal sensory excitatory transmission from the anterior cingulate cortex. *Nat Commun* **9**, 1886 (2018). https://doi.org:10.1038/s41467-018-04309-2
86 Liu, Y. *et al.* Touch and tactile neuropathic pain sensitivity are set by corticospinal projections. *Nature* **561**, 547-550 (2018). https://doi.org:10.1038/s41586-018-0515-2
87 Gan, Z. *et al.* Layer-specific pain relief pathways originating from primary motor cortex. *Science* **378**, 1336-1343 (2022). https://doi.org:10.1126/science.add4391
88 Taub, D. G. *et al.* The secondary somatosensory cortex gates mechanical and heat sensitivity. *Nat Commun* **15**, 1289 (2024). https://doi.org:10.1038/s41467-024-45729-7
89 Condylis, C. *et al.* Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022). https://doi.org:10.1126/science.abl5981
90 Sorensen, A. T. *et al.* A robust activity marking system for exploring active neuronal ensembles. *Elife* **5** (2016). https://doi.org:10.7554/eLife.13918
91 Mukherjee, D. *et al.* Salient experiences are represented by unique transcriptional signatures in the mouse brain. *Elife* **7** (2018). https://doi.org:10.7554/eLife.31220
92 Gonzales, B. J., Mukherjee, D., Ashwal-Fluss, R., Loewenstein, Y. & Citri, A. Subregion-specific rules govern the distribution of neuronal immediate-early gene induction. *Proc Natl Acad Sci U S A* **117**, 23304-23310 (2020). https://doi.org:10.1073/pnas.1913658116
93 Bottai, D. *et al.* Synaptic activity-induced conversion of intronic to exonic sequence in Homer 1 immediate early gene expression. *J Neurosci* **22**, 167-175 (2002). https://doi.org:10.1523/JNEUROSCI.22-01-00167.2002
94 French, P. J. *et al.* Subfield-specific immediate early gene expression associated with hippocampal long-term potentiation in vivo. *Eur J Neurosci* **13**, 968-976 (2001). https://doi.org:10.1046/j.0953-816x.2001.01467.x
95 Brakeman, P. R. *et al.* Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 284-288 (1997). https://doi.org:10.1038/386284a0
96 Szumlinski, K. K. *et al.* Behavioral and neurochemical phenotyping of Homer1 mutant mice: possible relevance to schizophrenia. *Genes Brain Behav* **4**, 273-288 (2005). https://doi.org:10.1111/j.1601-183X.2005.00120.x
97 Bugeon, S. *et al.* A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330-338 (2022). https://doi.org:10.1038/s41586-022-04915-7
98 Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370** (2020). https://doi.org:10.1126/science.abb2494
99 von Buchholtz, L. J. *et al.* Decoding Cellular Mechanisms for Mechanosensory Discrimination. *Neuron* **109**, 285-298 e285 (2021). https://doi.org:10.1016/j.neuron.2020.10.028
100 Han, Z. *et al.* AAV11 enables efficient retrograde targeting of projection neurons and enhances astrocyte-directed transduction. *Nat Commun* **14**, 3792 (2023). https://doi.org:10.1038/s41467-023-39554-7
101 Deverman, B. E. *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat Biotechnol* **34**, 204-209 (2016). https://doi.org:10.1038/nbt.3440
102 Lin, K. *et al.* AAV9-Retro mediates efficient transduction with axon terminal absorption and blood-brain barrier transportation. *Mol Brain* **13**, 138 (2020). https://doi.org:10.1186/s13041-020-00679-1
103 Pool, A. H., Poldsam, H., Chen, S., Thomson, M. & Oka, Y. Recovery of missing single-cell RNA-sequencing data with optimized transcriptomic references. *Nat Methods* **20**, 1506-1515 (2023). https://doi.org:10.1038/s41592-023-02003-w
104 You, Y. *et al.* Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. *Genome Biol* **22**, 339 (2021). https://doi.org:10.1186/s13059-021-02552-3
105 Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**, 203-214 (2000). https://doi.org:10.1089/10665270050081478

1005 106 Lawler, A. J. *et al.* Cell Type-Specific Oxidative Stress Genomic Signatures in the Globus
1006 Pallidus of Dopamine-Depleted Mice. *J Neurosci* **40**, 9772-9783 (2020).
1007 https://doi.org:10.1523/JNEUROSCI.1634-20.2020
1008 107 Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed,
1009 quantitative, sensitive, versatile, robust. *Development* **145** (2018).
1010 https://doi.org:10.1242/dev.165753
1011 108 Yates, S. C. *et al.* QUINT: Workflow for Quantification and Spatial Analysis of Features in
1012 Histological Images From Rodent Brain. *Front Neuroinform* **13**, 75 (2019).
1013 https://doi.org:10.3389/fninf.2019.00075
1014 109 Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas.
1015 *Cell* **181**, 936-953 e920 (2020). https://doi.org:10.1016/j.cell.2020.04.007
1016 110 Puchades, M. A., Csucs, G., Ledergerber, D., Leergaard, T. B. & Bjaalie, J. G. Spatial
1017 registration of serial microscopic brain images to three-dimensional reference atlases with the
1018 QuickNII tool. *PLoS One* **14**, e0216796 (2019). https://doi.org:10.1371/journal.pone.0216796
1019 111 Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**,
1020 1226-1232 (2019). https://doi.org:10.1038/s41592-019-0582-9
1021 112 Groeneboom, N. E., Yates, S. C., Puchades, M. A. & Bjaalie, J. G. Nutil: A Pre- and Post-
1022 processing Toolbox for Histological Rodent Brain Section Images. *Front Neuroinform* **14**, 37
1023 (2020). https://doi.org:10.3389/fninf.2020.00037
1024 113 Yang, L., Tochitsky, I., Woolf, C. J. & Renthal, W. Isolation of Nuclei from Mouse Dorsal Root
1025 Ganglia for Single-nucleus Genomics. *Bio-protocol* **11**, e4102-e4102 (2021).
1026 114 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529
1027 (2021). https://doi.org:10.1016/j.cell.2021.04.048
1028 115 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell
1029 RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
1030 https://doi.org:10.1016/j.cels.2019.03.003
1031 116 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state
1032 analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021). https://doi.org:10.1038/s41592-021-
1033 01282-5
1034 117 Yang, L. *et al.* Human and mouse trigeminal ganglia cell atlas implicates multiple cell types in
1035 migraine. *Neuron* **110**, 1806-1821 e1808 (2022). https://doi.org:10.1016/j.neuron.2022.03.003
1036 118 Consortium, E. P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse
1037 genomes. *Nature* **583**, 699-710 (2020). https://doi.org:10.1038/s41586-020-2493-4

1038

Fig. 1 -- Multiplex projection tracing using Projection-TAGs

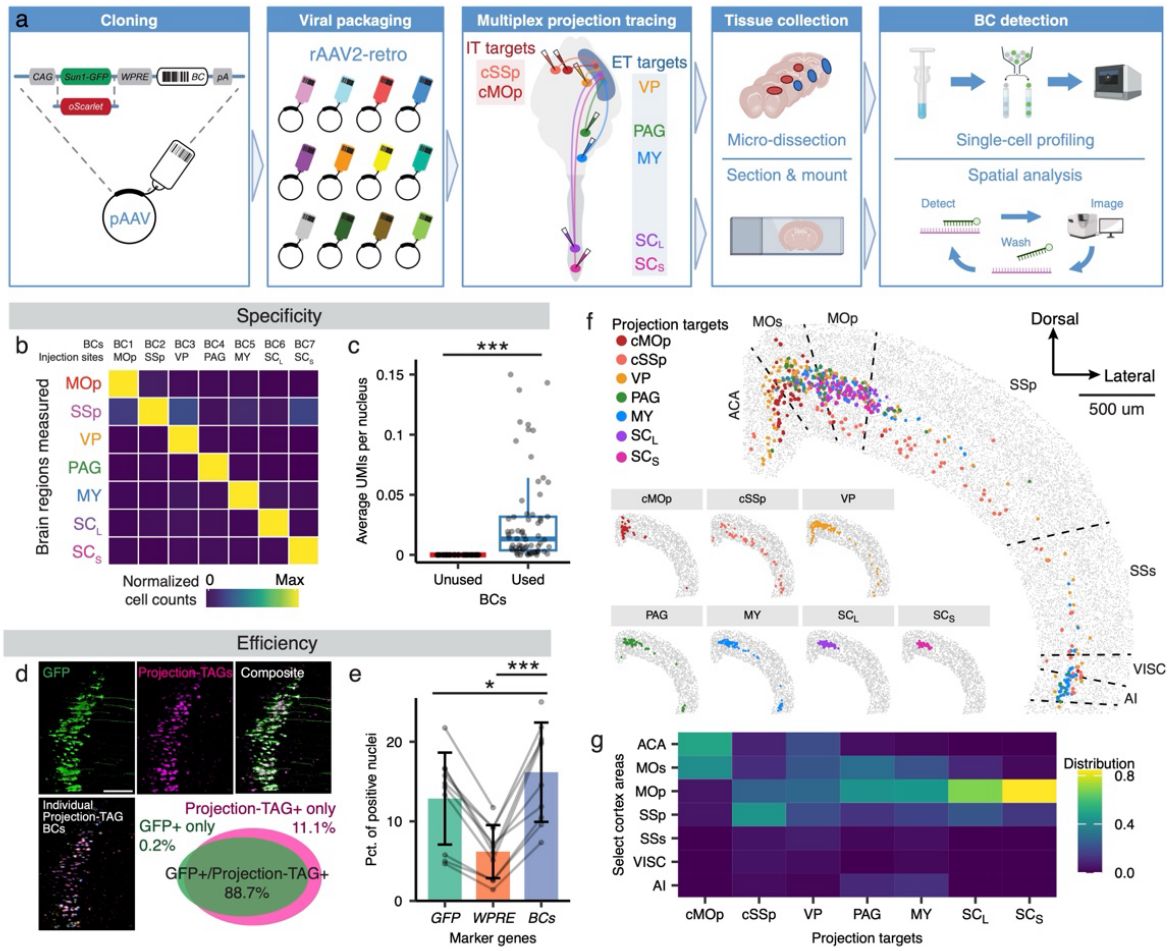


Figure 1. Multiplex projection tracing using Projection-TAGs
a, The experimental workflow of tagging neuronal projections of the MOp and SSp using Project-TAGs. **b**, Heatmap showing the enrichment of cells labeled with individual Projection-TAG BCs by multiplexed FISH in brain regions of injections. Cell counts are normalized by the max values in each brain region. Individual BCs and their injection sites are shown as on the column. **c**, Box plot showing the detection rate of Projection-TAG BCs in each snRNA-seq library prepared by FACS (Supplementary Table 2). Projection-TAG BCs are grouped by whether they were used in a given experiment. Detection rate is significantly higher for used BCs than unused ones (***p = 6.5e-08, one-sided student t-test). **d**, Representative MOp images showing cells labeled with the native fluorescence from Sun1-GFP protein and fluorescent signals from Projection-TAGs. Quantification shows the overlap between cells that are GFP+ and/or Projection-TAG+. FISH probes uniquely targeting Projection-TAGs (BCs 1-7) were visualized in separate imaging channels, shown as individual BCs, and the signals were aggregated pos-hoc, shown as Projection-TAGs. Scale bar = 100 μm. **e**, Dot plot showing the percentage of nuclei expressing select marker genes (>0 counts) in individual snRNA-seq libraries prepared by FACS. Bar and error bar indicate the average and standard deviation across libraries. There are significantly more Projection-TAG+ nuclei (nuclei expressing any of the BCs used in the experiment) compared to *GFP*+ and *WPRE*+ nuclei (***p = 7.98e-08, $F(2,18)$ = 46.32, one-way ANOVA; *p = 0.025 between BC and GFP, ***p = 5.13e-05 between BC and WPRE, post-hoc two-sided paired t-tests with Bonferroni correction). **f**, Spatial map of projection neurons in select cortex areas from a brain section (-0.65 mm relative to bregma). The projection targets are assigned to individual cells based on the Project-TAG BCs they express. Cells are downsampled to 200 per projection target and colored by their projection targets. Insets on the bottom left corner show the same cells separated by their individual projection targets. Grey dots are cells segmented based on DAPI signal. **g**, Heatmap showing the distribution of projection neurons in select cortex areas from Fig. 1f. Values in the heatmap are counts of neurons projecting to each target in each area normalized by total counts of Project-TAG+ cells for the given projection target. 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.

Fig 2 -- A multi-modal single-cell atlas of mouse cortex

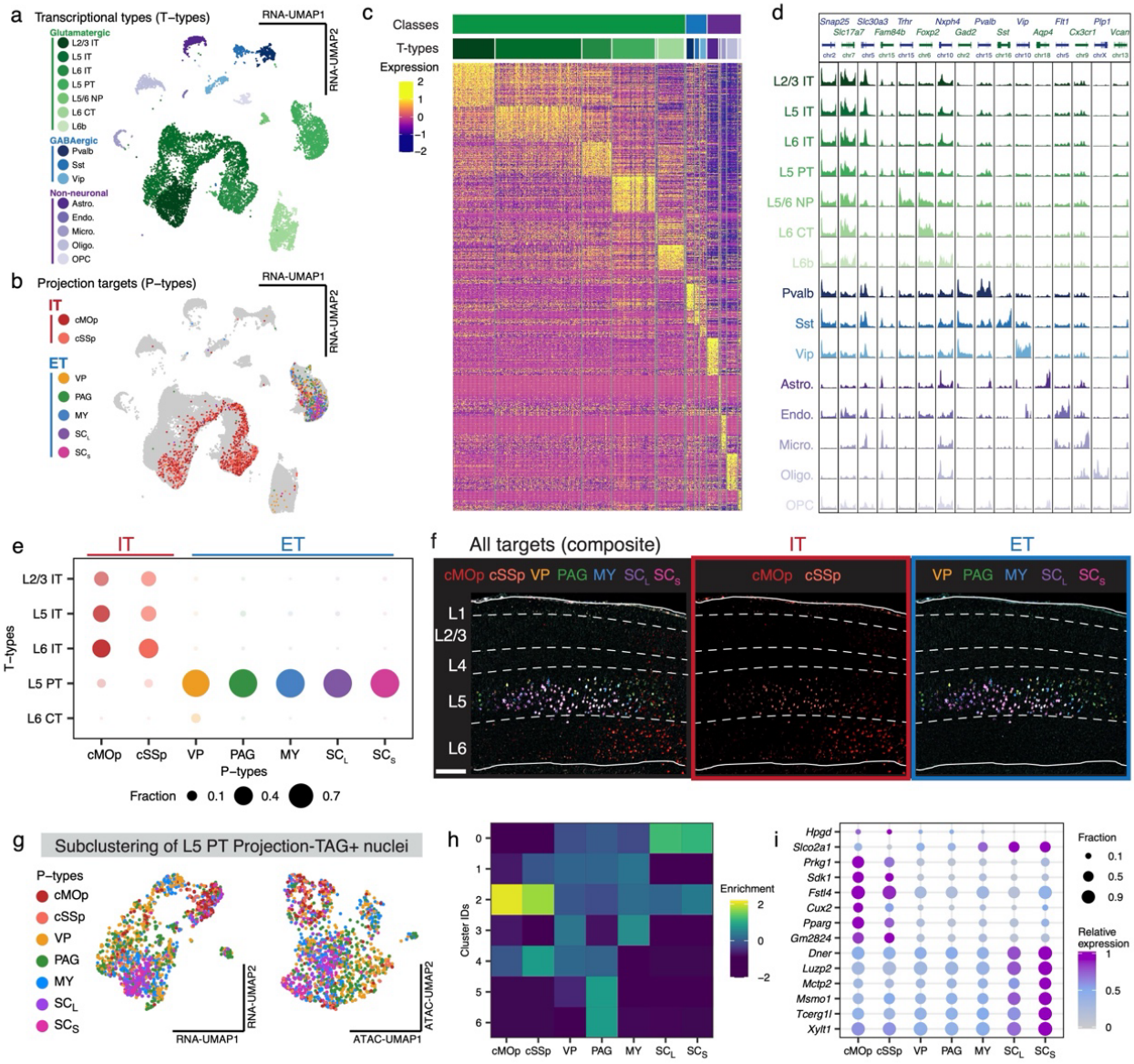


Figure 2. A multi-modal single-cell atlas of mouse cortex

a, Uniform manifold approximation and projection (UMAP)[1] visualization of snRNA-seq data showing 10,000 downsampled nuclei, colored by T-types. **b**, UMAP visualization of snRNA-seq data showing 17,000 nuclei (10,000 downsampled Projection-TAG- nuclei and 1,000 downsampled Projection-TAG+ nuclei for each Projection-TAG BC). Projection-TAG- nuclei are colored grey and Projection-TAG+ nuclei are colored by their projection targets (P-types). **c**, Heatmap showing expression of top 100 marker genes (sorted by Log2FC) per T-types in individual nuclei of snRNAs-seq. Nuclei are downsampled to 1,000 and ordered based on T-types. See Fig. 2a for the colors of classes and T-types. **d**, Coverage plot of average accessibility of canonical cell-type-specific marker genes in individual T-types of snATAC-seq. The chromatin accessibility is displayed as the average frequency of sequenced DNA fragments per nucleus for each T-type, grouped by 50 bins per displayed genomic region. **e**, Dot plot showing fraction of Projection-TAG+ nuclei annotated by individual T-types, normalized by the total counts of Projection-TAG+ nuclei per P-type. **f**, 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 BCs. Scale bar = 150 μm. Left: composite image of all FISH channels, middle and right: images of a subset of Projection-TAG channels labeling IT targets and ET targets, respectively. **g**, UMAP of L5 PT neurons that are Projection-TAG+. Nuclei are clustered based on snRNA-seq (left) and snATAC-seq (right), and colored by P-types. **h**, Heatmap showing the enrichment of P-types in each cluster. The enrichment indicates the Z-score for each P-type (by row) of the percentage of nuclei assigning to each cluster. **i**, Dot plot displaying the expression of select P-type-specific marker genes in nuclei of individual P-types. Dot size denotes the fraction of nuclei expressing a marker gene (>0 counts), and color denotes relative expression of a gene in each cell type (calculated as the mean expression of a gene relative to the highest mean expression of that gene across all P-types).

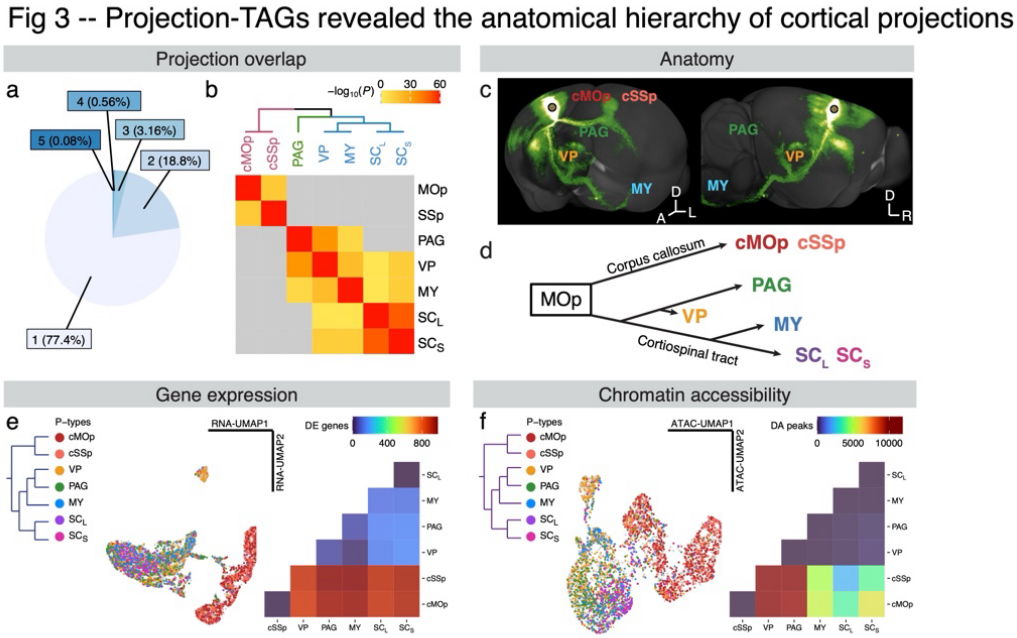


Figure 3. Projection-TAGs revealed the anatomical hierarchy of cortical projections
a, Pie chart showing the percentage of Projection-TAG+ nuclei, grouped by the number of unique Projection-TAG BCs detected in each nucleus. **b**, Heatmap showing the Bonferroni-corrected *p* values from pairwise hypergeometric tests of overlapping between nuclei expressing any of the two P-types labeled with Projection-TAGs. Hierarchical clustering was performed based on the Bonferroni-corrected *p* values of each P-type to each of the other P-types. **c**, Tracing the brain-wide axon collaterals of MOp neurons. Screenshots were acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/100141563)[2]. **d**, Summary diagram of the anatomical hierarchy of axonal projections from MOp to seven target regions. **e**, Gene expression distance of P-types Left: hierarchical clustering was performed based on the average expression in individual P-types of each gene detected in snRNA-seq; middle: UMAP of Projection-TAG+ nuclei from snRNA-seq, colored by P-types; right: number of DE genes between any two P-types. **f**, Chromatin accessibility distance of P-types. Left: hierarchical clustering was performed based on the average accessibility in individual P-types of each peak detected in snATAC-seq data; middle: UMAP of Projection-TAG+ nuclei from snATAC-seq, colored by P-types; right: number of DA peaks between any two P-types.

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

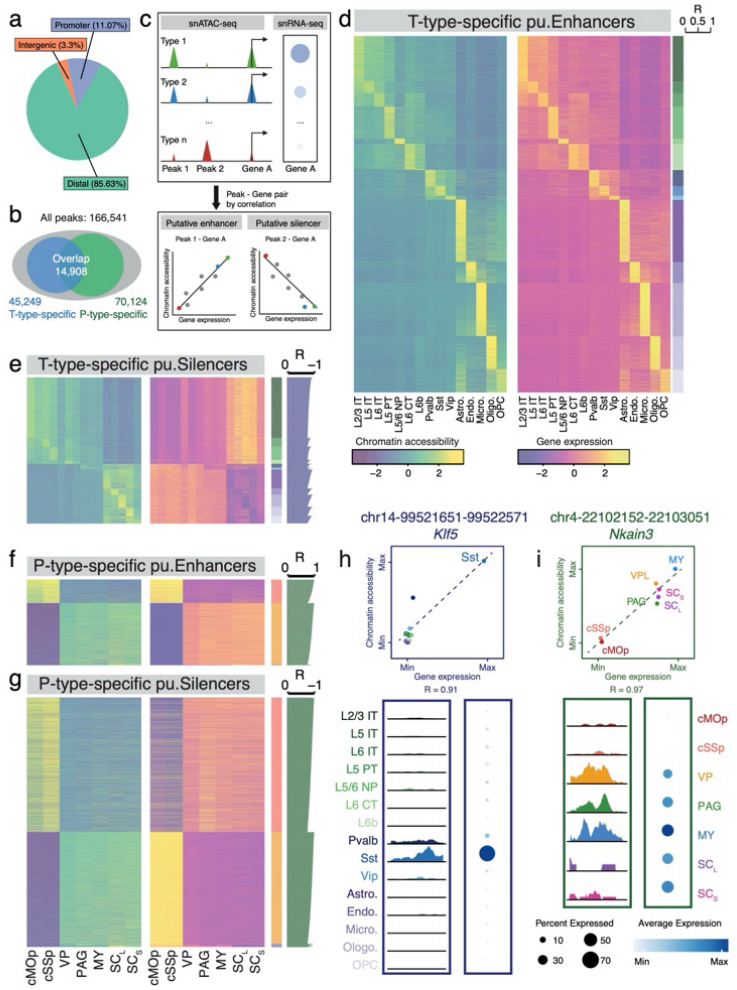


Figure 4. Genomic regulatory networks of cortical cell types using Projection-TAGs
a, Pie chart categorizing peaks in snATAC-seq based on their genomic locations. Peaks are mapped to one of the three categories: promoter regions (-1,000 bp to +100 bp of transcription start site [TSS]), distal regions (<200 kb upstream or downstream of TSS or within gene body, excluding promoter), and intergenic regions (>200kb upstream or downstream of TSS, excluding gene body) **b**, Distribution of DA peaks in any T-types and/or P-types across all peaks detected in snATAC-seq. **c**, Statistical workflow of identifying putative enhancers and silencers and their regulated genes. **d**, heatmaps of the accessibility of pu.Enhancer peaks in snATAC-seq data (left) and the expression of their regulated genes in snRNA-seq data (right) in individual T-types. On the right side next to the heatmaps, the color bar indicates the T-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual T-types. See Fig. 2a for the colors of T-types. **e**, Heatmaps of T-type-specific pu.Silencers and their regulated genes. **f** and **g**, heatmaps of P-type-specific puEnhancers and pu.Silencers and their regulated genes. On the right side next to the heatmaps, the color bar indicates the P-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual P-types. See Fig. 2b for the colors of P-types. **h**, Top: Scatter plot showing the correlation between the accessibility of peak chr14-99521651-99522571 and the expression of gene *Klf5* in individual T-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 chr14-99521651-99522571, displayed as the average fraction of transposase-sensitive fragments per nucleus in each T-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 *Klf5* in each T-type. Colors indicate T-types. **i**, Top: Scatter plot showing the correlation between the accessibility of peak chr4-22102152-22103051 and the expression of gene *Nkain3* in individual P-types (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr4-22102152-22103051, displayed as the average fraction of transposase-sensitive fragments per nucleus in each P-type (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Nkain3* in each P-type. Colors indicate P-types.

Fig 5 -- Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain

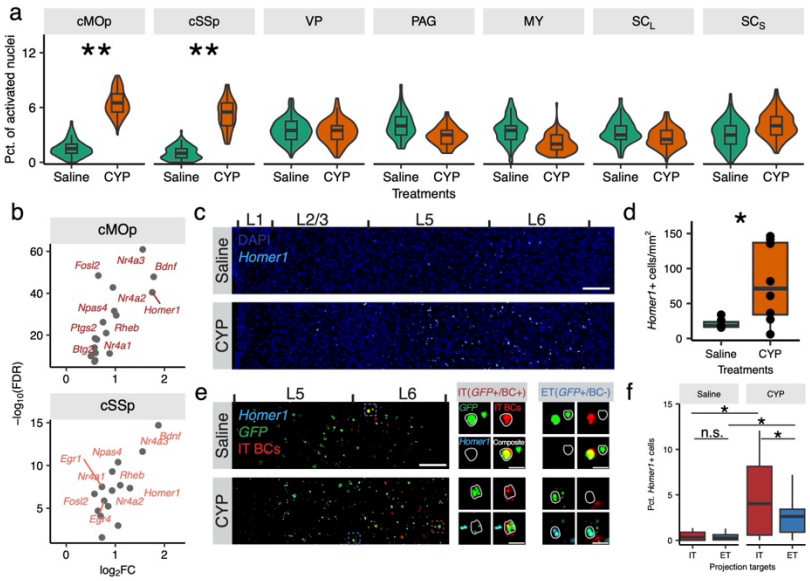
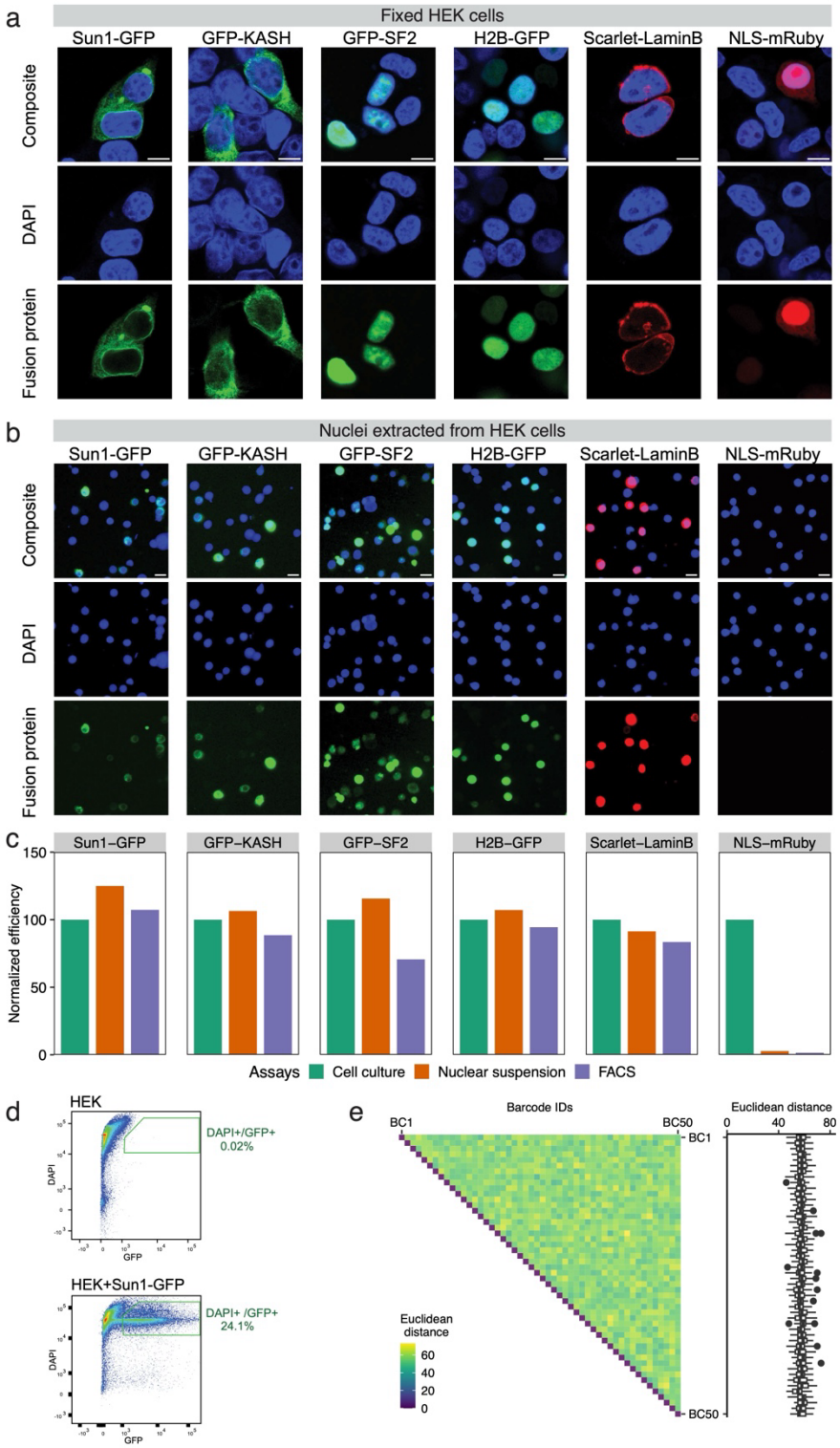


Figure 5. Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain
a, Violin plots and box plots showing percentage of activate nuclei between treatments in individual P-types. For each P-type and treatment, 200 nuclei were randomly sampled to calculate the percentage and the process repeated for a total of 1,000 times. The *p* values are calculated as the fraction of repeats where percentage of activated nuclei from CYP-treated animals is lower than that of saline-treated animal. (***p* = 0.005 for cMOp, ***p* = 0.001 for cSSp, *p* = 0.4 for VP, *p* = 0.64 for PAG, *p* = 0.74 for MY, *p* = 0.62 for SC_L, *p* = 0.25 for $SC_{S)}$. **b**, Volcano plots showing significantly induced IEGs (avg_log2FC > 0.5) in activated nuclei from CYP-treated mice compared to the same number of randomly sampled nuclei (with matched T-types) from saline-treated mice for the two IT P-types. **c**, Representative FISH images showing *Homer1* expression in MOp from brain sections (approximately -0.2 mm relative to bregma) of saline- and CYP-treated mice. Scale bar: 100 µm. **d**, Quantification of *Homer1*+ cells in MOp and SSp areas from brain sections from CYP- and saline-treated mice (six slices from 0 to -0.92 mm relative to bregma per group, **p* = 0.021, one-side t-test). **e**, Left: representative FISH images of layers 5 and 6 of MOp from the same slices in Fig. 4c showing the expression of *Homer1, GFP*, and two Projection-TAGs labeling neurons projecting to cMOp and cSSp. Scale bar: 100 µm. Right: IT-projecting neurons are defined as cells co-labeled with *GFP* and Projection-TAGs labelling cMOp and cSSp projections, whereas ET-projecting neurons are defined as cells labeled with *GFP* only. **f**, Quantification of *Homer1* expression in IT and ET-projecting neurons in MOp and SSp areas (Data collected from 6 cortical slices per group. **p* = 0.02, *F*(3,30) = 3.88, one-way ANOVA; *p=0.01 in IT between CYP and saline, *p=0.02 in ET between CYP and saline, *p* = 0.40 between IT and ET in saline, **p* = 0.04 between IT and ET in CYP, post-hoc one-sided t-tests with Bonferroni correction).

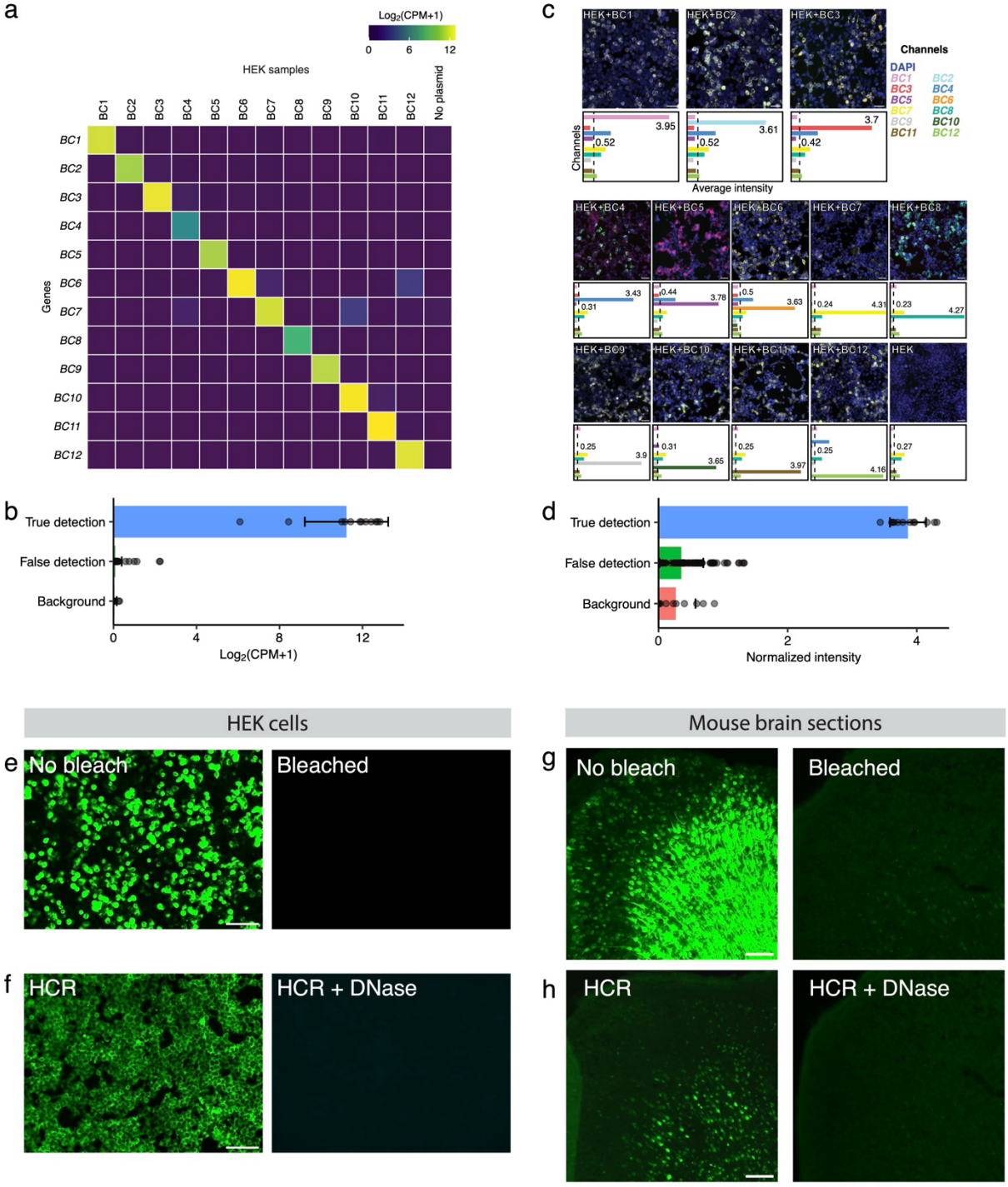
Extended Data Fig. 1



1143

1144 **Extended Data Fig. 1 – related to Fig. 1**
1145 **a**. Representative fluorescent images of the localization of fusion proteins in different cellular compartments in fixed HEK cells.
1146 Scale bar: 10 μm. **b**, Representative images of the fluorescent labeling using fusion proteins of nuclear resuspension extracted
1147 from HEK cells. Scale bar: 20 μm. **c**, Efficiency of fluorescent labeling of cells and nuclei, measured as percentage of HEK
1148 cells labeled with fusion proteins on fixed slices (cell culture), percentage of nuclei labeled with fusion proteins on nuclear
1149 suspension extracted from HEK cells, and percentage of nuclei labeled with fusion proteins during FACS. Percentages were
1150 normalized to the percentage of cell culture for each protein. **d**, Scatter plot showing the gate of sorting labeled HEK nuclei
1151 (DAPI+ and GFP+) extracted from HEK cells and HEK cells expressing Sun1-GFP. **e**, Left: pairwise Euclidean distance
1152 between any two of the 50 BCs cataloged in this study. Right: box plot showing the average Euclidean distance of individual
1153 BCs to other BCs.

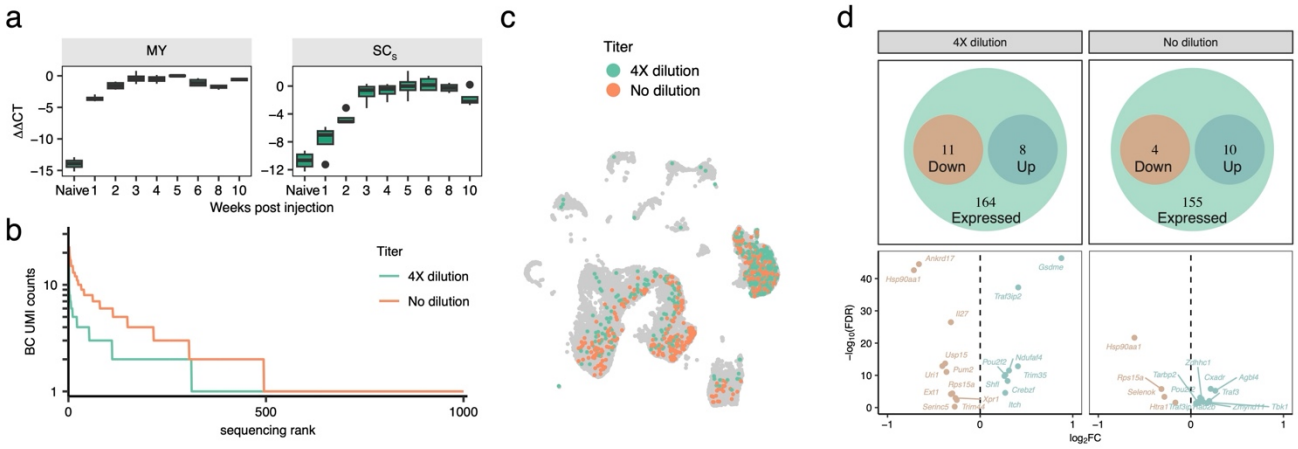
Extended Data Fig. 2



1154

1155 **Extended Data Fig. 2 – related to Fig. 1**
1156 **a**, Counts of sequencing reads that map to individual BCs (row) in individual HEK samples (column) in RNA-seq. Each HEK
1157 sample was either transfected with a plasmid expressing a unique Project-TAG BC or no transfection. **b**, Bar plot showing the
1158 average detection rate of Projection-TAGs in individual samples, error bars are standard deviation. True detection is the
1159 detection of the Projection-TAG BC expressed in a given sample transfected with one Projection-TAG BC-expressing plasmid,
1160 false detection is the detection of Projection-TAG BCs not expressed in the same samples, and the background is the
1161 detection of individual Projection-TAG BCs in the HEK sample with no transfection. **c**, Representative FISH images showing
1162 the detection of individual Projection-TAGs (BCs 1-12) in individual HEK samples each expressing a unique BC or the HEK
1163 sample with no transfection. The bar plots show the average signal intensity from each Projection-TAG BC channel across a
1164 field of view with 1 mm by 1 mm. The average intensity is normalized by the max values of each sample and each channel. In
1165 HEK samples transfected with Projection-TAG BC-expressing plasmids, the dotted line and the associated number represent
1166 the false detection rate (average intensity of Projection-TAG BCs not expressed in a given sample), whereas the number on
1167 the right indicates the true detection rate (intensity from the Projection-TAG BC expressed in a given sample). In the HEK
1168 sample with no transfection, the dotted line and the associated number represent the background (average intensity of all
1169 Projection-TAG BCs). Scale bar: 20 μm **d**, Bar plot showing average detection rate of Projection-TAGs across all samples,
1170 error bars are standard deviation. **e**, Representative images showing that photobleaching process removed the native Sun1-
1171 GFP fluorescence from HEK cells. Scale bar: 50 μm. **f**, Representative images showing that DNase incubation removed the
1172 fluorescent signal in the green channel from the FISH probe targeting *GAPDH* in HEK cells. Scale bar: 50 μm. **g**,
1173 Representative images showing that photobleaching process significantly reduced the native Sun1-GFP fluorescence from
1174 mouse brain sections. Scale bar: 100 μm. **h**, Representative images showing that DNase incubation removed the fluorescent
1175 signals in the green channel from the FISH probe targeting *GFP* in mouse brain sections. Scale bar: 100 μm.

Extended Data Fig. 3



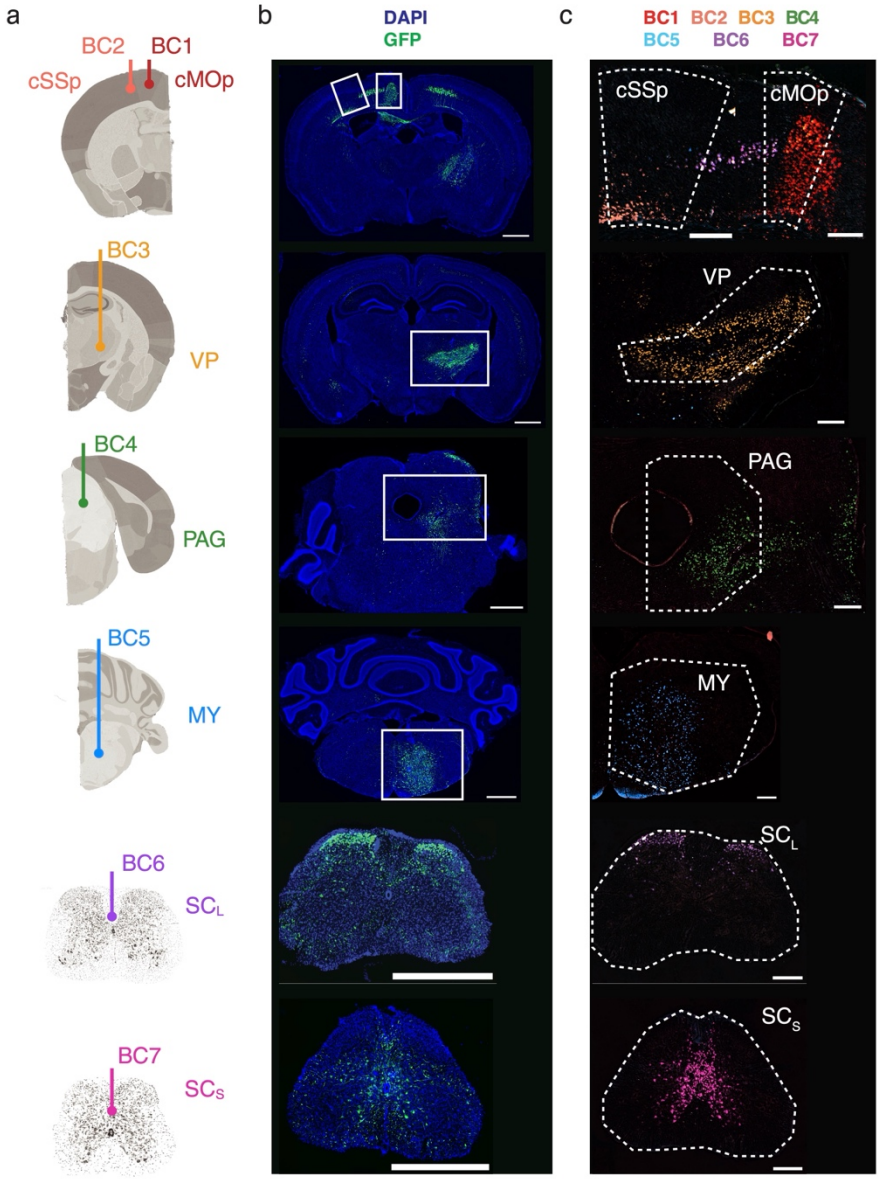
e

	Sequencing depth	BC detection	
	(UMIs/nucleus)	(Percentile by expression rank)	(Pct. BC+ nuclei)
10X Genomics	8.3 - 11.1k	57.5 - 70.1	0.2 - 0.58
Parse	19.9 - 20.6 k	49.7 - 58.5	0 - 0.06

Extended Data Fig. 3 – related to Fig. 1

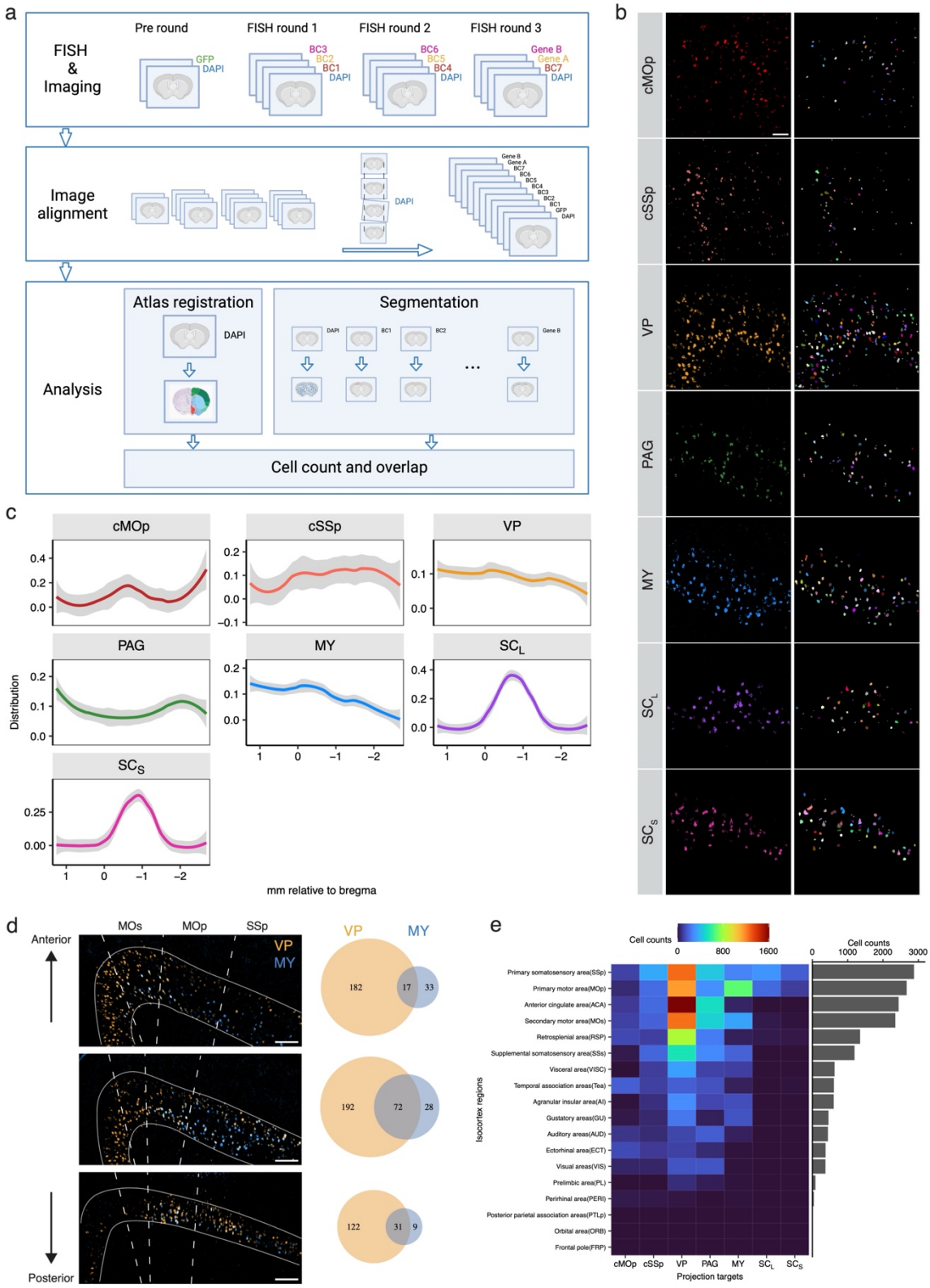
a, Box plot showing the time kinetics of Projection-TAG expression in the cortex. Animals received MY or SC_S injection with the Projection-TAG AAV (expressing BC2), cortical samples were collected at different time points post injection, and the Projection-TAG expression is measured using qPCR (n=4 for each time point and each injection). ΔCT is calculated as the CT value of BC2 minus that of GAPDH for each sample, ΔΔCT is calculated as the ΔCT at each time point minus the ΔCT at 4 weeks post injection. **b**, Plot showing the detection rate of Projection-TAGs (BCs) in 1,000 downsampled nuclei from each titer group in snRNA-seq. Nuclei were ranked based on their Projection-TAG detection rate. **c**, RNA-UMAP of 10,000 nuclei (1,000 downsampled BC+ nuclei from each titer group, colored by titer, and 8,000 downsampled BC- nuclei, colored grey). **d**, DE genes in response to viral infection. DE analysis was performed comparing the expression of genes annotated in the Gene Ontology list GO:0009615 (response to virus) between Projection-TAG+ nuclei and Projection-TAG- nuclei (downsampled to match the T-type distribution and nuclei counts to the Projection-TAG+ nuclei) in each tier group. Top panels show the number of upregulated (avg_log2FC > 0) and downregulated (avg_log2FC < 0) viral responding genes among the viral responding genes expressed (average expression > 0.5) in snRNA-seq nuclei from each titer group. Fisher's exact test was performed to test if viral responding genes are significantly enriched among DE genes (p = 0.81 and 0.77 for undiluted and diluted group, respectively). Bottom panels show the individual DE viral responding genes. **e**, Key sequencing and Projection-TAG detection metrics from snRNA-seq libraries prepared using different commercial kits from nuclei of the same biological samples (n = 2 for each kit).

Extended Data Fig. 4



Extended Data Fig. 4 – related to Fig. 1
a, Schematics showing the stereotaxic injection of individual BCs into each of the projection targets of the cortex. **b**, Representative images showing the native fluorescence of Sun1-GFP in the injection sites. Scale bar: 1 mm. **c**, Representative FISH images (zoom-in view of the region highlighted in b) showing the expression of Projection-TAGs in the injection sites. Multiplexed FISH was performed to visualize the Projection-TAGs (BCs 1-7) in the projection targets of the cortex (highlighted with dotted lines). Scale bar: 300 μm.

Extended Data Fig. 5

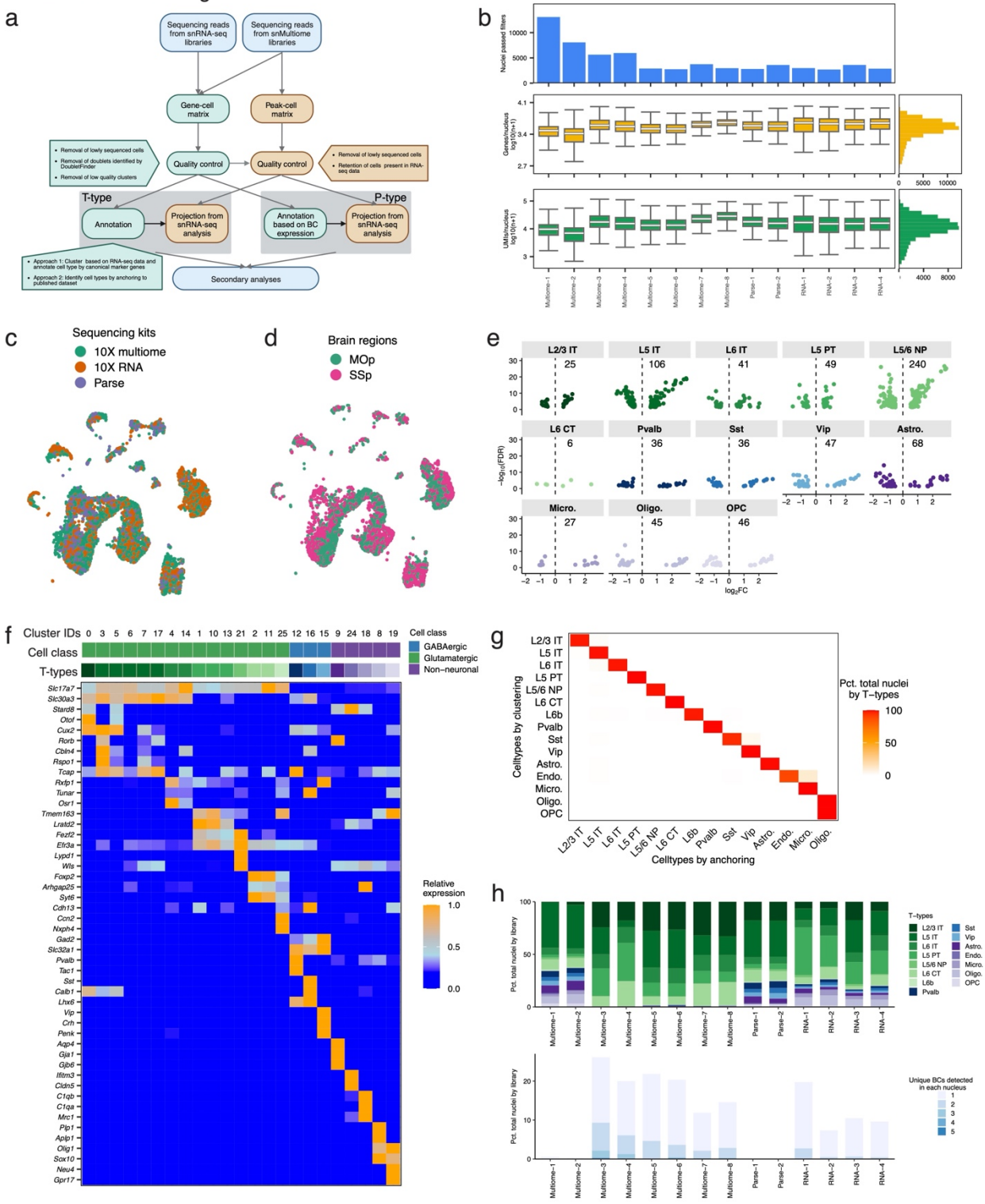


1202

Extended Data Fig. 5 – related to Fig. 1

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

Extended Data Fig. 6

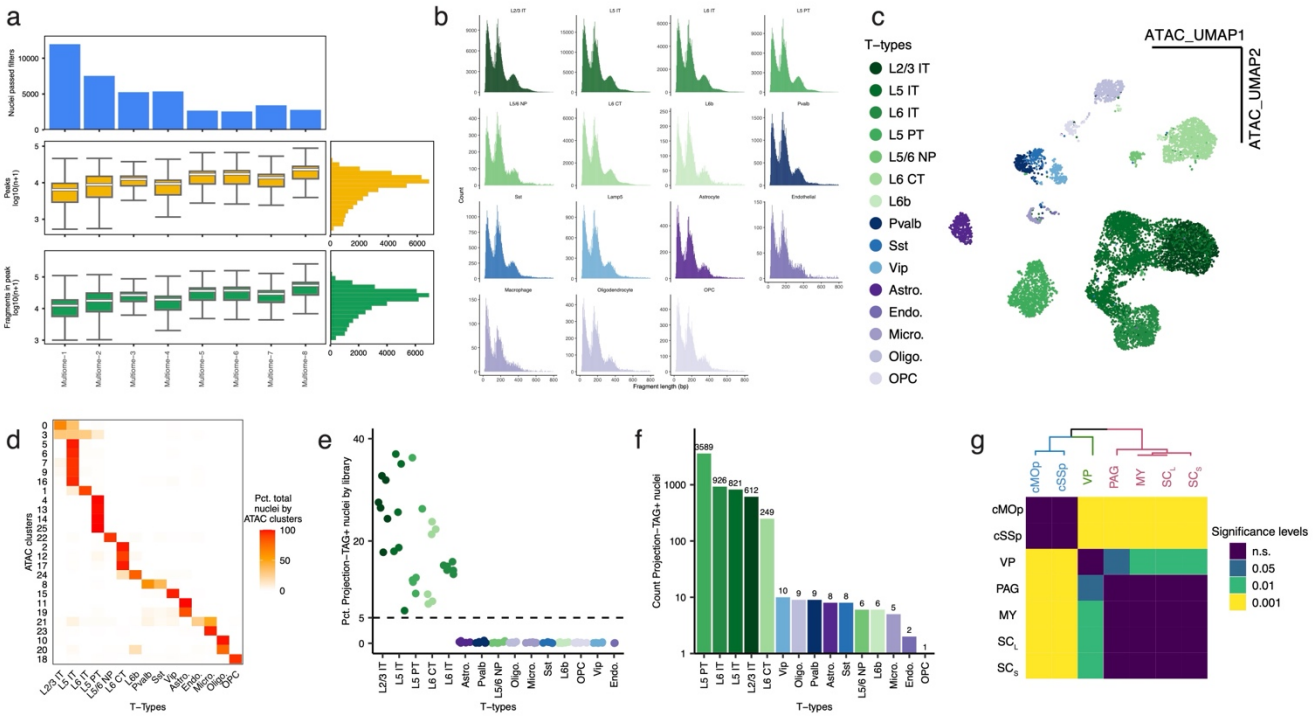


1214

Extended Data Fig. 6 – related to Fig. 2

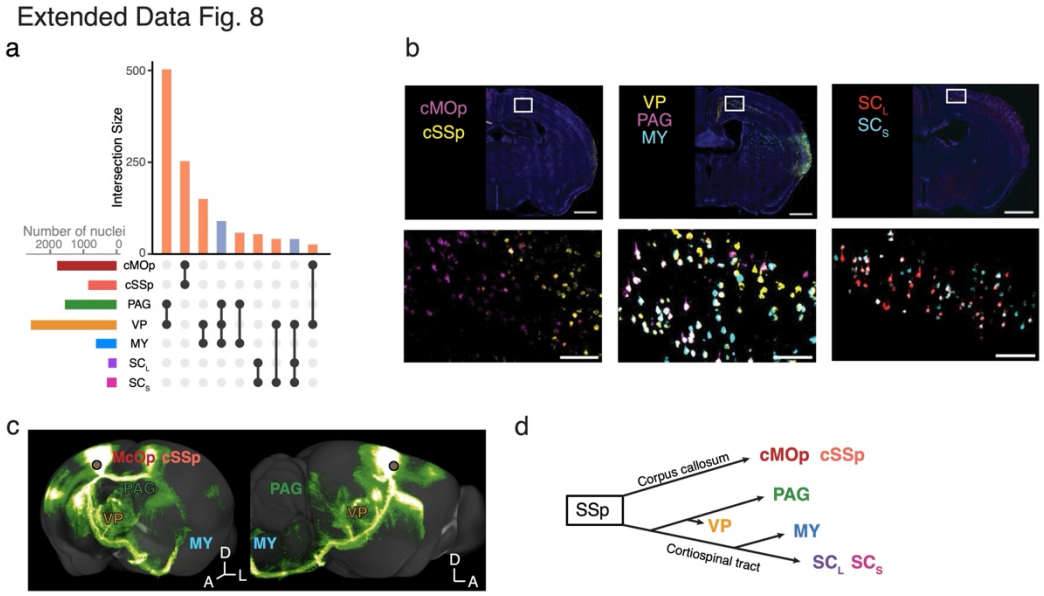
a, Workflow of the data analysis of the snRNA-seq and snATAC-seq and T-type and P-type annotation. **b**, snRNA-seq library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of genes per nucleus (log10 transformed) and the bottom row displays the number of UMIs per nucleus. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is aggregated across all samples and displayed on the horizontal histogram. **c** and **d**, RNA-UMAP of 1,000 downsampled nuclei per group. Nuclei were colored by sequencing kits (c) or brain regions (d). **e**, Volcano plots of DE genes between nuclei from MOp (avg_log2FC > 0.5) and SSp (avg_log2FC < -0.5) in individual T-types with at least 30 nuclei from each brain region. Numbers indicates total DE genes in each T-type. **f**, T-type assignment to each cluster identified in snRNA-seq data. Heatmap showing the distinct expression pattern of marker genes that are used to assign T-types. See h for the colors of T-types. **g**, Heatmap showing the correspondence of T-type assignment of nuclei in snRNA-seq data between the manual classifications (f) and the classifications assigned by anchoring to a published scRNA-seq dataset of mouse motor cortex[44]. Plot displays percentage of nuclei within the initial T-type assignment that is assigned to each cell type after anchoring to the published dataset. Overall, 98.2 ± 3.1% nuclei in each cluster were assigned the same T-types between the two approaches. **h**, Composition of T-types (top) and Projection-TAG+ nuclei grouped by number of unique BCs detected per nucleus (bottom) in each snRNA-seq library.

Extended Data Fig. 7



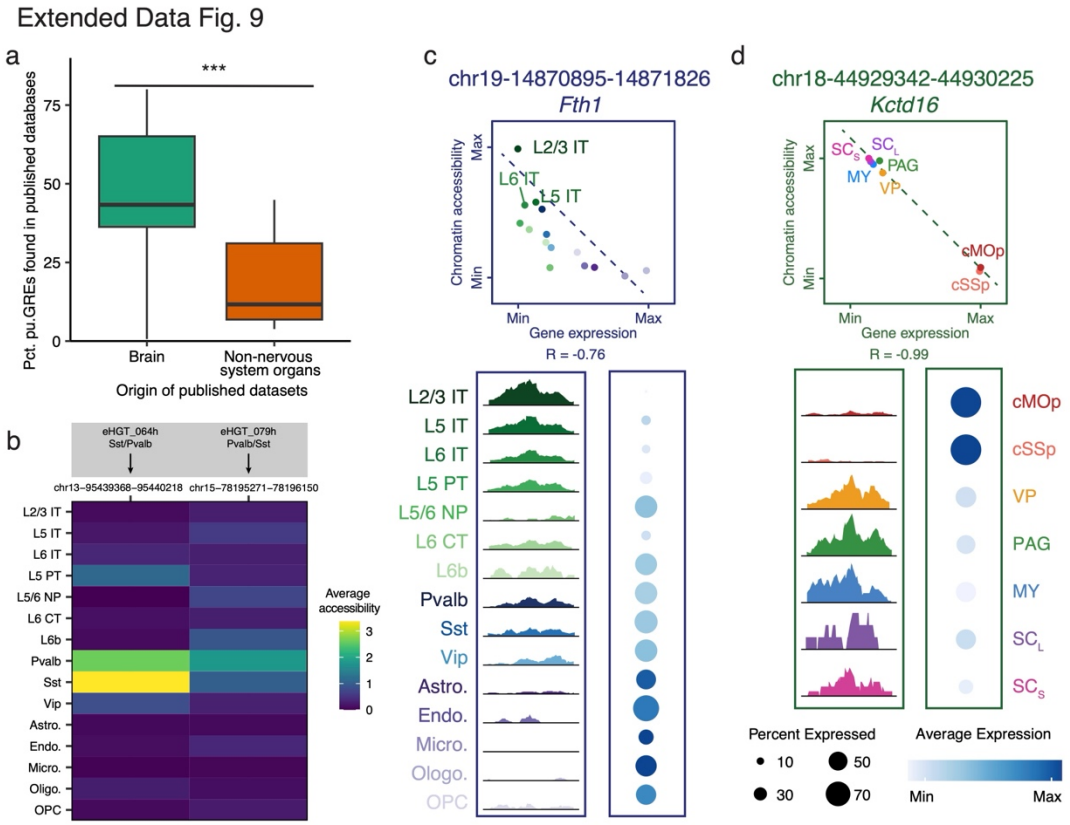
Extended Data Fig. 7 – related to Fig. 2

a, snATAC-seq library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of peaks per nucleus (log10 transformed) and the bottom row displays the number of fragments that overlap with peaks per nucleus. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is aggregated across all samples and displayed on the horizontal histogram. **b**, Distribution of snATAC-seq fragment lengths in individual T-types. **c**, ATAC-UMAP of downsampled 10,000 nuclei of snATAC-seq. Nuclei are colored by T-types. **d**, Correspondence of ATAC clusters (row) and T-types (column) of nuclei profiled combinatorially by snRNA-seq and snATAC-seq. Plot displays percentage of nuclei within each ATAC cluster that is assigned to each T-type. **e**, Dot plot showing the percentage of nuclei that are Projection-TAG+ in each T-type in each snRNA-seq library prepared by FACS. Dotted line indicates the cutoff for calling T-types enriched of projection neurons (5%). **f**, Bar plot showing the counts of Projection-TAG+ nuclei in each T-type in snRNA-seq data. **g**, Heatmap showing the significance levels of the Bonferroni-corrected *p* values of pairwise Fisher's exact tests of T-type distribution between any of the two P-types. Hierarchical clustering was performed based on the Bonferroni-corrected *p* values of each P-type to each of the other P-types.



Extended Data Fig. 8 – related to Fig. 3

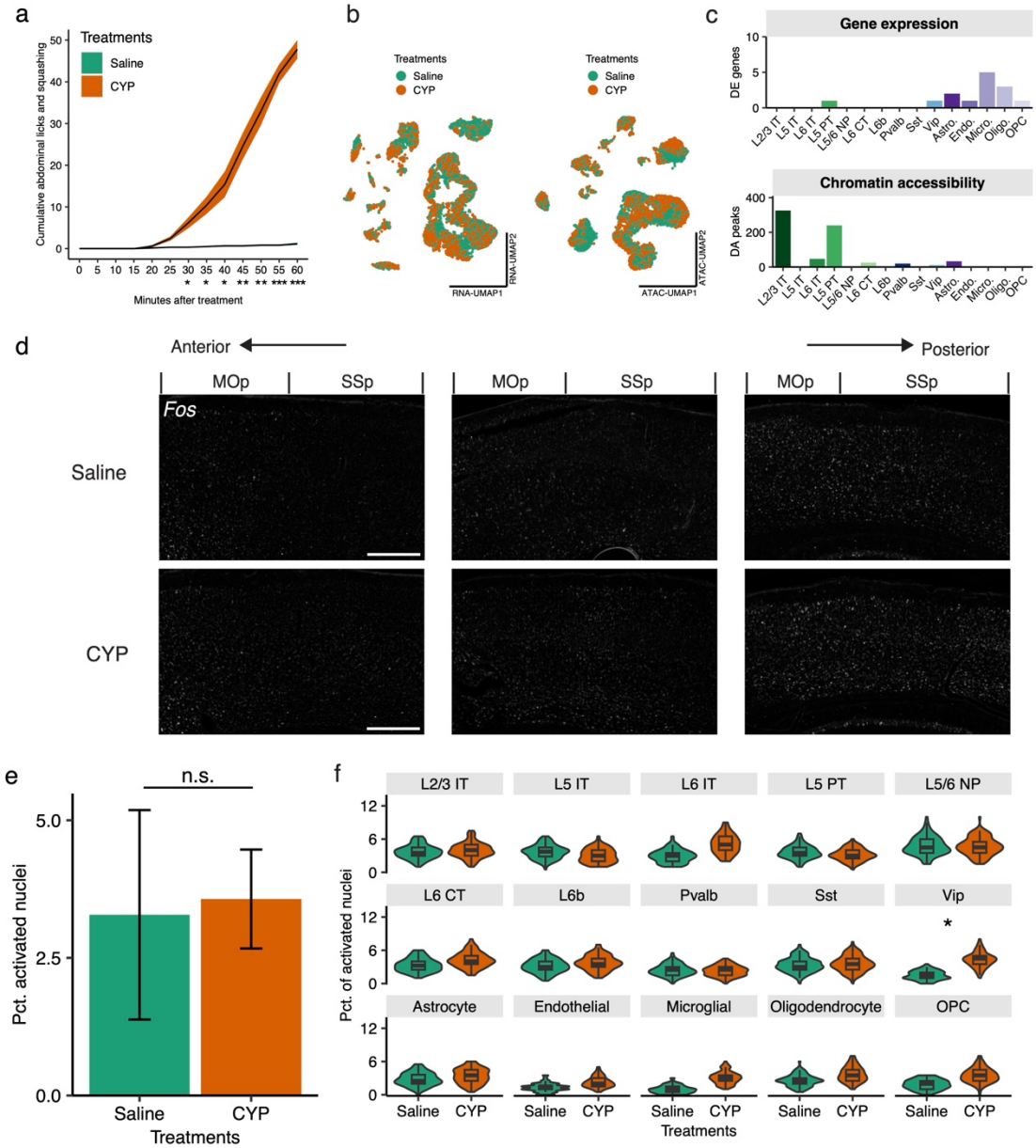
a, Upset plot showing the overlapping across P-types. Bar plot shows the counts of Projection-TAG+ nuclei in each P-type. Upset plot shows the combinations of overlapping across P-types with at least 10 nuclei. **b**, Representative FISH images showing projection neurons labelled by Projection-TAGs in the coronal section showing motor cortex from different anterior-posterior sections (top) and zoomed-in images (bottom) of the highlighted region demonstrate expression of two unique Projection-TAGs injected into downstream cMOp and cSSp; middle: demonstrate expression of three unique Projection-TAGs injected into the VP, PAG, and MY; right: demonstrate expression of two unique Projection-TAGs injected into SC_L and SC_S). Scale bar: 1 mm (top), 100 μm (bottom). **c** Tracing the brain-wide axon collaterals of SSp neurons. Images were acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/112935169)[2]. **d**, a summary diagram of the anatomical hierarchy of axonal projections from SSp to the seven target regions.



Extended Data Fig. 9 – related to Fig. 4

a, Box plot showing the overlapping of putative GREs identified from this study to the putative GREs identified by studies from the ENCODE consortium[79,118]. Overlap is significantly higher with studies of the brain compared to those of non-brain organs (***p = 8.26e-05, t-test). Brain studies: ENCSR422CGZ, ENCSR694ZBB, ENCFF426KUB, and ENCSR410XMU. Studies of organs that are not the nervous system: ENCSR052AZT (stomach), ENCSR100TMZ (lung), ENCSR296UQL (liver), and ENCSR180OIF (limb). **b**, Heatmap showing the chromatin accessibility of two T-type-specific pu.Enhancers in individual T-types. The peaks of the two pu.Enhancers overlap with the genomic coordinates liftover from two functional enhancers previously identified based on human snATAC-seq data[73]. Labels in the grey boxes show the functional enhancers and the cell types in which the enhancers exhibited highest activity in experimental validation from the previous study. **c**, Top: Scatter plot showing the correlation between the accessibility of peak chr19-14870895-14871826 and the expression of gene *Fth1* in individual T-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 chr19-14870895-14871826, displayed as the average fraction of transposase-sensitive fragments per nucleus in each T-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 *Fth1* in each T-type. Colors indicate T-types. **d**, Top: Scatter plot showing the correlation between the accessibility of peak chr18-44929342-44930225 and the expression of the gene *Kctd16* in individual P-types (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr18-44929342-44930225, displayed as the average fraction of transposase-sensitive fragments per nucleus in each P-type (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Kctd16* in each P-type. Colors indicate P-types.

Extended Data Fig. 10



Extended Data Fig. 10 – related to Fig. 5

a, Cumulative scores of quantified behaviors between mice treated with saline and CYP. CYP treatment resulted in significant increases in the spontaneous behaviors like abdominal licking and squashing behaviors compared to saline treatment. Ribbons show SEM. ** p = 0.0034, one-way ANOVA. p = NaN (0 min), NaN (5 min), NaN (10 min), NaN (15 min), 0.25 (20 min), 0.07 (25 min), *0.05 (30 min), *0.04 (35 min), *0.04 (40 min), **0.001 (45 min), **0.006 (50 min), ***0.0005 (55 min), ***0.0003 (60 min), respectively, post-hoc one-sided t-test, n=5-6 mice/group. **b**, RNA-UMAP (left) and ATAC-UMAP (right) of 5,000 downsampled nuclei from each treatment group, colored by treatments. **c**, Bar plots showing number of DE genes (top) and DA peaks (bottom) between nuclei from CYP (avg_log2FC > 0.5 or < -0.5, FDR < 0.05) and saline in individual T-types. **d**, Representative FISH images showing *Fos* expression in select cortex areas (top: 1.33 mm, middle: -0.95 mm, bottom -1.67 mm relative to bregma). Scale bar: 400 µm. **e**, Bar plot showing the average percentage of activate nuclei by Act-seq in individual snRNA-seq libraries of different treatments. Error bars show standard deviation. (p = 0.74, t-test). **f**, Violin plots and box plots showing the percentage of activate nuclei by Act-seq in individual T-types between treatments. Two hundred nuclei per T-type and treatment were randomly sampled to calculate the percentage and the process repeat for a total of 1,000 times. The p values are calculated as the fraction of repeats where percentage of activated nuclei from CYP-treated animals is lower than that of saline-treated (p = 0.42 for L2/3 IT, p = 0.61 for L5IT, p = 0.11 for L6 IT, p = 0.52 for L5 PT, p = 0.47 for L5/6 NP, p = 0.28 for L6 CT, p = 0.34 for L6b, p = 0.47 for Pvalb, p = 0.40 for Sst, *p = 0.02 for Vip, p = 0.31 for Astrocyte, p = 0.16 for Endothelial cells, p = 0.05 for Microglial cells, p = 0.24 for Oligodendrocytes, and p = 0.10 for OPC).

Fig. 1 -- Multiplex projection tracing using Projection-TAGs

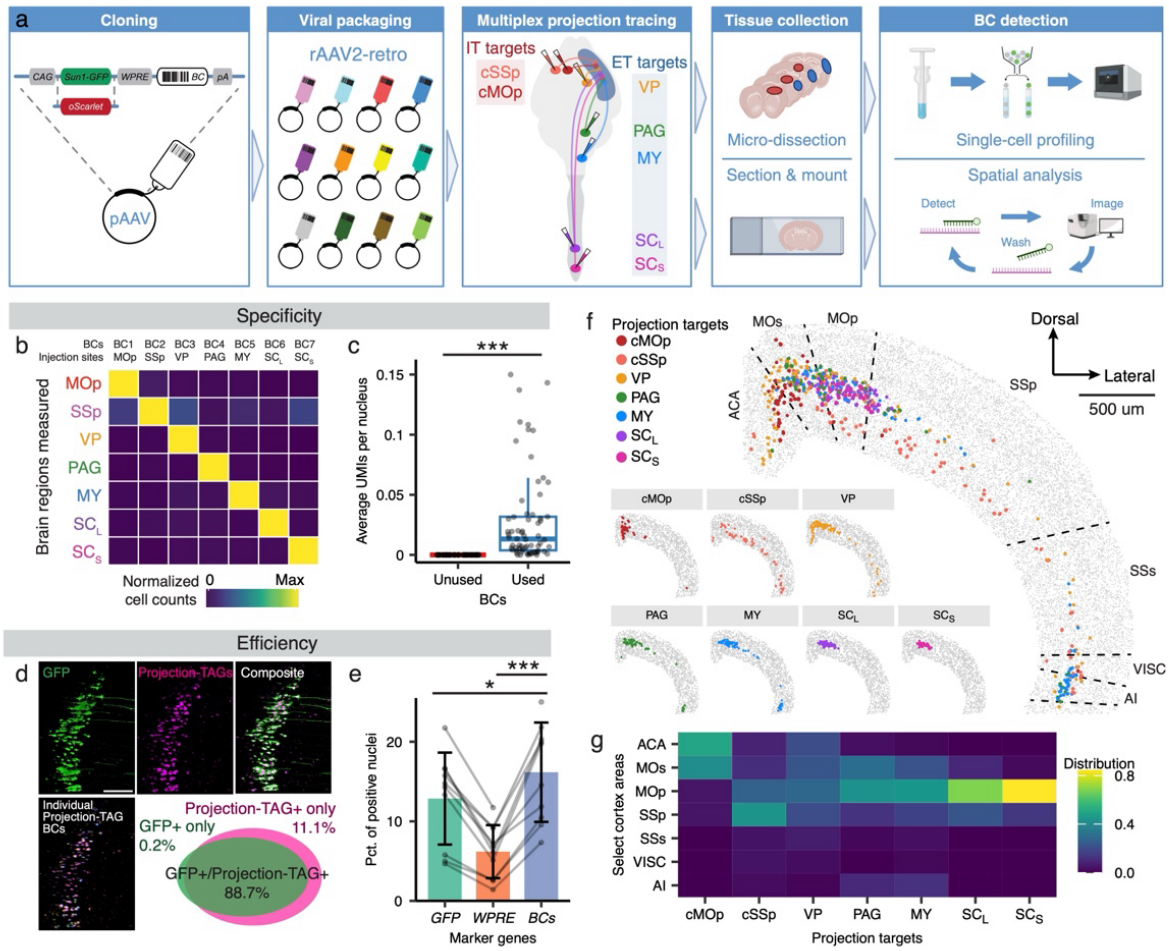


Figure 1. Multiplex projection tracing using Projection-TAGs

a, The experimental workflow of tagging seven neuronal projections of the MOp and SSp using Project-TAGs. **b**, Heatmap showing the enrichment of cells labeled with individual Projection-TAG BCs by multiplexed FISH in brain regions of injections. Cell counts are normalized by the max values in each brain region. Individual BCs and their injection sites are shown as on the column. **c**, Box plot showing the detection rate of Projection-TAG BCs in each snRNA-seq library prepared by FACS (Supplementary Table 2). Projection-TAG BCs are grouped by whether they were used in a given experiment. Detection rate is significantly higher for used BCs than unused ones (***p = 6.5e-08, one-sided student t-test). **d**, Representative MOp images showing cells labeled with the native fluorescence from Sun1-GFP protein and fluorescent signals from Projection-TAGs. Quantification shows the overlap between cells that are GFP+ and/or Projection-TAG+. FISH probes uniquely targeting Projection-TAGs (BCs 1-7) were visualized in separate imaging channels, shown as individual BCs, and the signals were aggregated pos-hoc, shown as Projection-TAGs. Scale bar = 100 µm. **e**, Dot plot showing the percentage of nuclei expressing select marker genes (>0 counts) in individual snRNA-seq libraries prepared by FACS. Bar and error bar indicate the average and standard deviation across libraries. There are significantly more Projection-TAG+ nuclei (nuclei expressing any of the BCs used in the experiment) compared to *GFP*+ and *WPRE*+ nuclei (***p = 7.98e-08, $F(2,18)$ = 46.32, one-way ANOVA; *p = 0.025 between BC and GFP, ***p = 5.13e-05 between BC and WPRE, post-hoc two-sided paired t-tests with Bonferroni correction). **f**, Spatial map of projection neurons in select cortex areas from a brain section (-0.65 mm relative to bregma). The projection targets are assigned to individual cells based on the Project-TAG BCs they express. Cells are downsampled to 200 per projection target and colored by their projection targets. Insets on the bottom left corner show the same cells separated by their individual projection targets. Grey dots are cells segmented based on DAPI signal. **g**, Heatmap showing the distribution of projection neurons in select cortex areas from Fig. 1f. Values in the heatmap are counts of neurons projecting to each target in each area normalized by total counts of Project-TAG+ cells for the given projection target. ACA: anterior cingulate area, MOs: secondary motor area, MOp: primary motor area, SSp: primary somatosensory area, SSs: supplemental somatosensory area, VISC: visceral area, AI: agranular insular area.

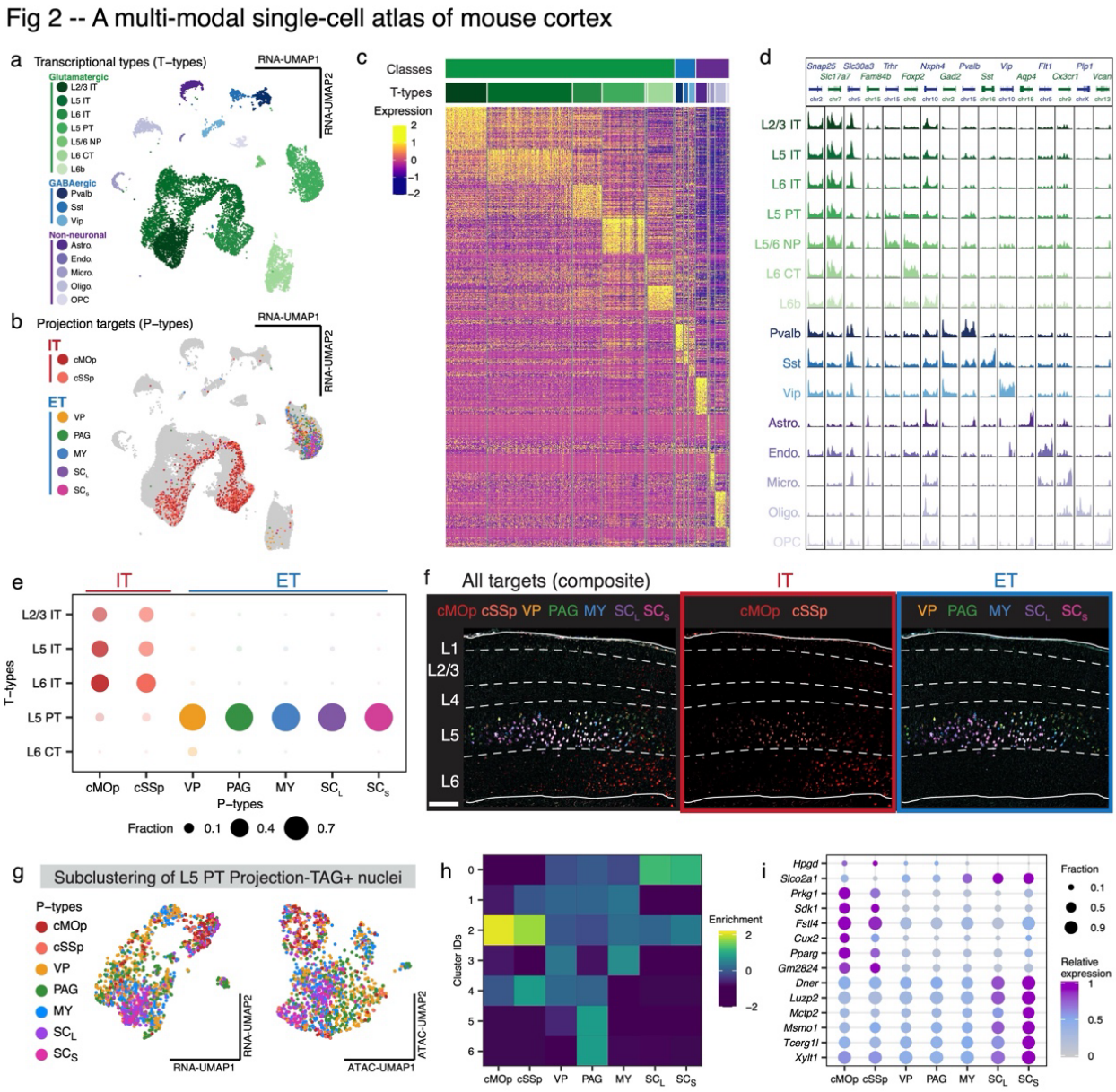


Figure 2. A multi-modal single-cell atlas of mouse cortex

a, Uniform manifold approximation and projection (UMAP)[1] visualization of snRNA-seq data showing 10,000 downsampled nuclei, colored by T-types. **b**, UMAP visualization of snRNA-seq data showing 17,000 nuclei (10,000 downsampled Projection-TAG- nuclei and 1,000 downsampled Projection-TAG+ nuclei for each Projection-TAG BC). Projection-TAG- nuclei are colored grey and Projection-TAG+ nuclei are colored by their projection targets (P-types). **c**, Heatmap showing expression of top 100 marker genes (sorted by Log2FC) per T-types in individual nuclei of snRNAs-seq. Nuclei are downsampled to 1,000 and ordered based on T-types. See Fig. 2a for the colors of classes and T-types. **d**, Coverage plot of average accessibility of canonical cell-type-specific marker genes in individual T-types of snATAC-seq. The chromatin accessibility is displayed as the average frequency of sequenced DNA fragments per nucleus for each T-type, grouped by 50 bins per displayed genomic region. **e**, Dot plot showing fraction of Projection-TAG+ nuclei annotated by individual T-types, normalized by the total counts of Projection-TAG+ nuclei per P-type. **f**, 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 BCs. Scale bar = 150 μm. Left: composite image of all FISH channels, middle and right: images of a subset of Projection-TAG channels labeling IT targets and ET targets, respectively. **g**, UMAP of L5 PT neurons that are Projection-TAG+. Nuclei are clustered based on snRNA-seq (left) and snATAC-seq (right), and colored by P-types. **h**, Heatmap showing the enrichment of P-types in each cluster. The enrichment indicates the Z-score for each P-type (by row) of the percentage of nuclei assigning to each cluster. **i**, Dot plot displaying the expression of select P-type-specific marker genes in nuclei of individual P-types. Dot size denotes the fraction of nuclei expressing a marker gene (>0 counts), and color denotes relative expression of a gene in each cell type (calculated as the mean expression of a gene relative to the highest mean expression of that gene across all P-types).

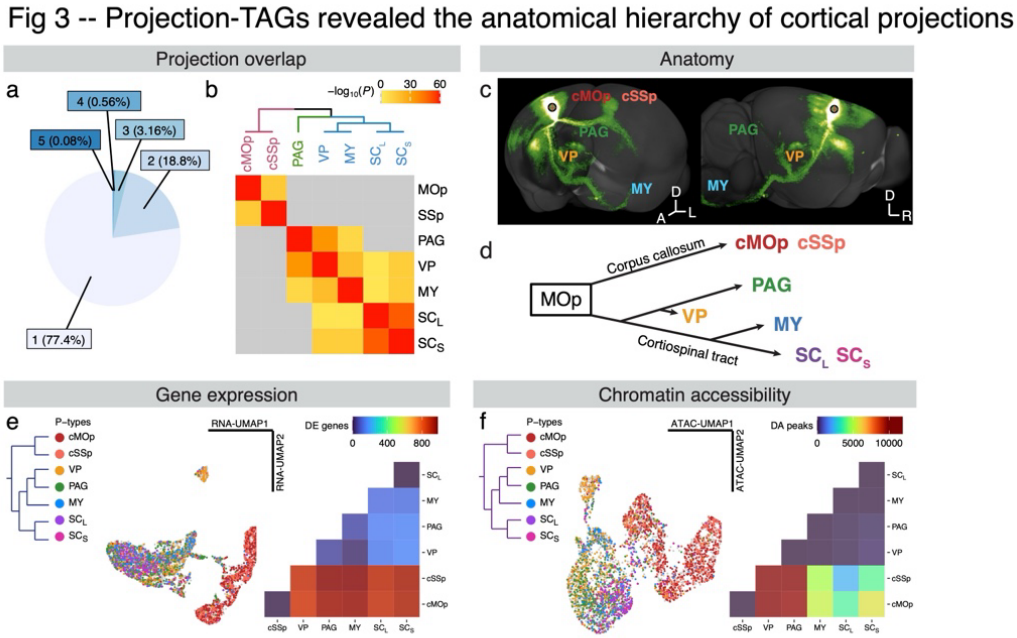


Figure 3. Projection-TAGs revealed the anatomical hierarchy of cortical projections

a, Pie chart showing the percentage of Projection-TAG+ nuclei, grouped by the number of unique Projection-TAG BCs detected in each nucleus. **b**, Heatmap showing the Bonferroni-corrected *p* values from pairwise hypergeometric tests of overlapping between nuclei expressing any of the two P-types labeled with Projection-TAGs. Hierarchical clustering was performed based on the Bonferroni-corrected *p* values of each P-type to each of the other P-types. **c**, Tracing the brain-wide axon collaterals of MOp neurons. Screenshots were acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/100141563)[2]. **d**, Summary diagram of the anatomical hierarchy of axonal projections from MOp to seven target regions. **e**, Gene expression distance of P-types Left: hierarchical clustering was performed based on the average expression in individual P-types of each gene detected in snRNA-seq; middle: UMAP of Projection-TAG+ nuclei from snRNA-seq, colored by P-types; right: number of DE genes between any two P-types. **f**, Chromatin accessibility distance of P-types. Left: hierarchical clustering was performed based on the average accessibility in individual P-types of each peak detected in snATAC-seq data; middle: UMAP of Projection-TAG+ nuclei from snATAC-seq, colored by P-types; right: number of DA peaks between any two P-types.

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

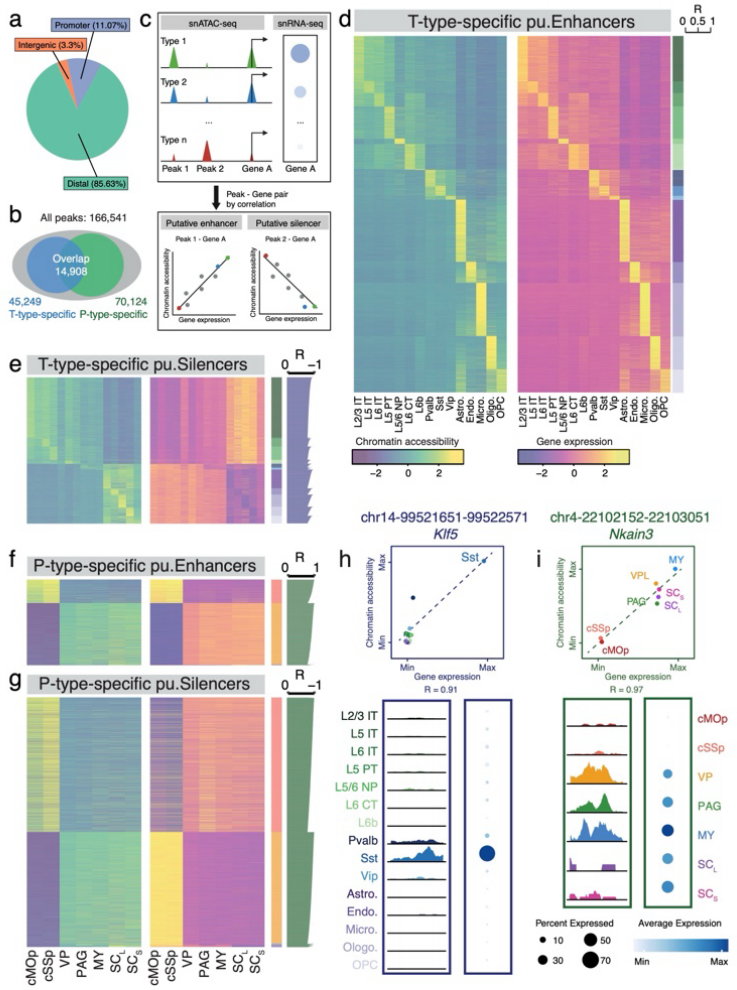


Figure 4. Genomic regulatory networks of cortical cell types using Projection-TAGs

a, Pie chart categorizing peaks in snATAC-seq based on their genomic locations. Peaks are mapped to one of the three categories: promoter regions (-1,000 bp to +100 bp of transcription start site [TSS]), distal regions (<200 kb upstream or downstream of TSS or within gene body, excluding promoter), and intergenic regions (>200kb upstream or downstream of TSS, excluding gene body) **b**, Distribution of DA peaks in any T-types and/or P-types across all peaks detected in snATAC-seq. **c**, Statistical workflow of identifying putative enhancers and silencers and their regulated genes. **d**, heatmaps of the accessibility of pu.Enhancer peaks in snATAC-seq data (left) and the expression of their regulated genes in snRNA-seq data (right) in individual T-types. On the right side next to the heatmaps, the color bar indicates the T-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual T-types. See Fig. 2a for the colors of T-types. **e**, Heatmaps of T-type-specific pu.Silencers and their regulated genes. **f** and **g**, heatmaps of P-type-specific puEnhancers and pu.Silencers and their regulated genes. On the right side next to the heatmaps, the color bar indicates the P-types in which the peak and gene have highest accessibility and expression, and the bar plot shows the coefficient (R) of Pearson correlation between the accessibility of the peak and the expression of the gene in individual P-types. See Fig. 2b for the colors of P-types. **h**, Top: Scatter plot showing the correlation between the accessibility of peak chr14-99521651-99522571 and the expression of gene *Klf5* in individual T-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 chr14-99521651-99522571, displayed as the average fraction of transposase-sensitive fragments per nucleus in each T-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 *Klf5* in each T-type. Colors indicate T-types. **i**, Top: Scatter plot showing the correlation between the accessibility of peak chr4-22102152-22103051 and the expression of gene *Nkain3* in individual P-types (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr4-22102152-22103051, displayed as the average fraction of transposase-sensitive fragments per nucleus in each P-type (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Nkain3* in each P-type. Colors indicate P-types.

Fig 5 -- Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain

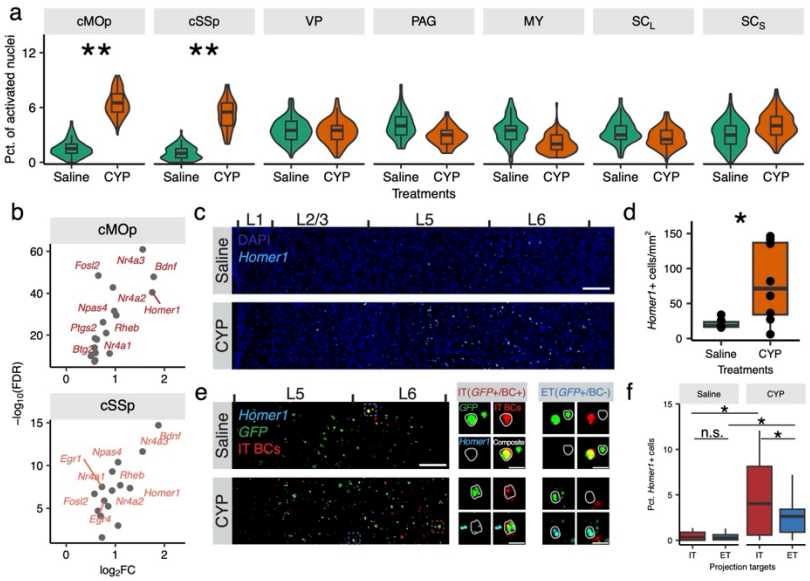
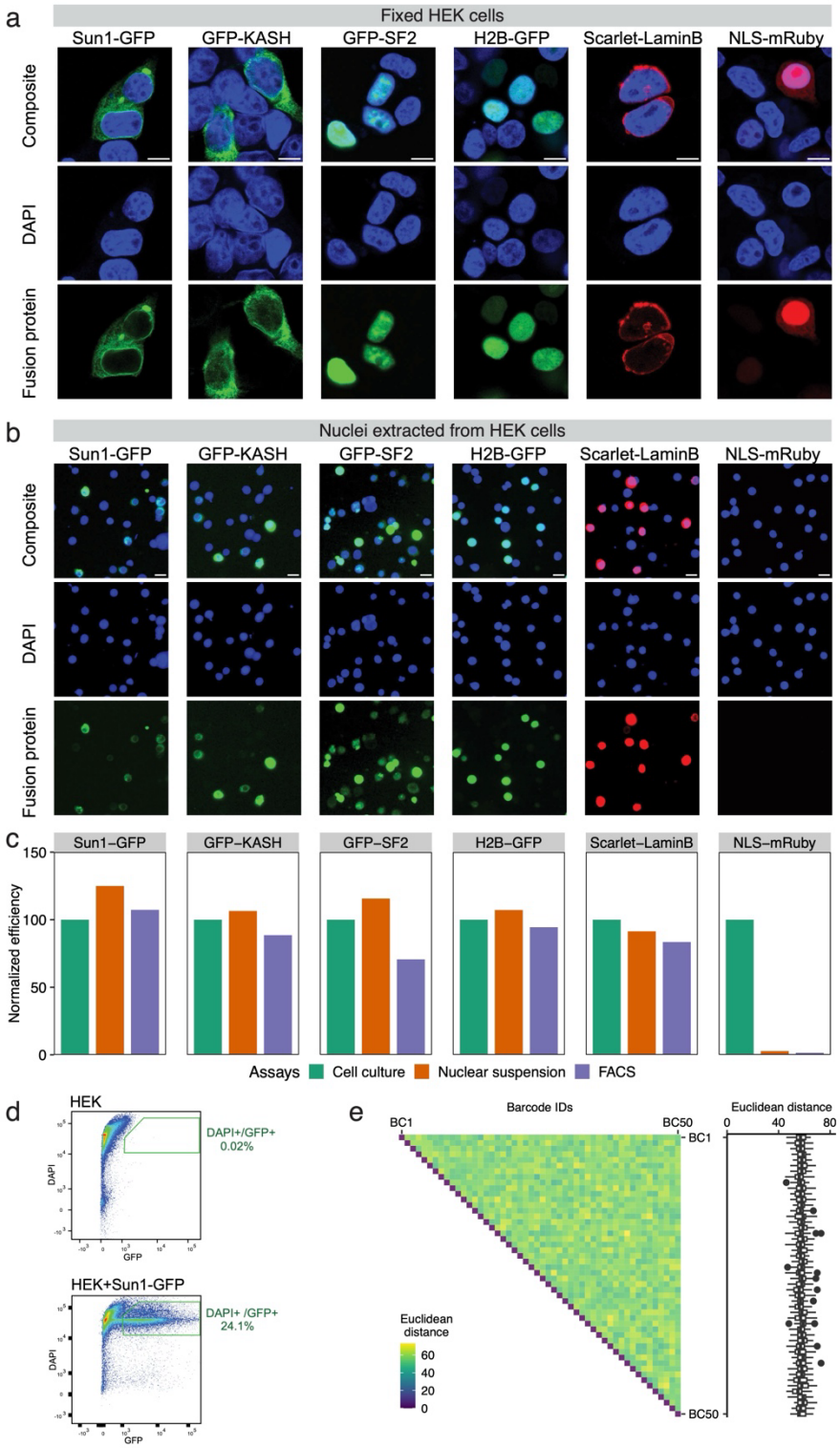


Figure 5. Projection-TAGs enable the detection of projection neurons acutely activated by visceral pain

a, Violin plots and box plots showing percentage of activate nuclei between treatments in individual P-types. For each P-type and treatment, 200 nuclei were randomly sampled to calculate the percentage and the process repeated for a total of 1,000 times. The *p* values are calculated as the fraction of repeats where percentage of activated nuclei from CYP-treated animals is lower than that of saline-treated animal. (***p* = 0.005 for cMOp, ***p* = 0.001 for cSSp, *p* = 0.4 for VP, *p* = 0.64 for PAG, *p* = 0.74 for MY, *p* = 0.62 for SC_L, *p* = 0.25 for $SC_{S)}$. **b**, Volcano plots showing significantly induced IEGs (avg_log2FC > 0.5) in activated nuclei from CYP-treated mice compared to the same number of randomly sampled nuclei (with matched T-types) from saline-treated mice for the two IT P-types. **c**, Representative FISH images showing *Homer1* expression in MOp from brain sections (approximately -0.2 mm relative to bregma) of saline- and CYP-treated mice. Scale bar: 100 µm. **d**, Quantification of *Homer1*+ cells in MOp and SSp areas from brain sections from CYP- and saline-treated mice (six slices from 0 to -0.92 mm relative to bregma per group, **p* = 0.021, one-side t-test). **e**, Left: representative FISH images of layers 5 and 6 of MOp from the same slices in Fig. 4c showing the expression of *Homer1, GFP*, and two Projection-TAGs labeling neurons projecting to cMOp and cSSp. Scale bar: 100 µm. Right: IT-projecting neurons are defined as cells co-labeled with *GFP* and Projection-TAGs labelling cMOp and cSSp projections, whereas ET-projecting neurons are defined as cells labeled with *GFP* only. **f**, Quantification of *Homer1* expression in IT and ET-projecting neurons in MOp and SSp areas (Data collected from 6 cortical slices per group. **p* = 0.02, *F*(3,30) = 3.88, one-way ANOVA; *p=0.01 in IT between CYP and saline, *p=0.02 in ET between CYP and saline, *p* = 0.40 between IT and ET in saline, **p* = 0.04 between IT and ET in CYP, post-hoc one-sided t-tests with Bonferroni correction).

Extended Data Fig. 1

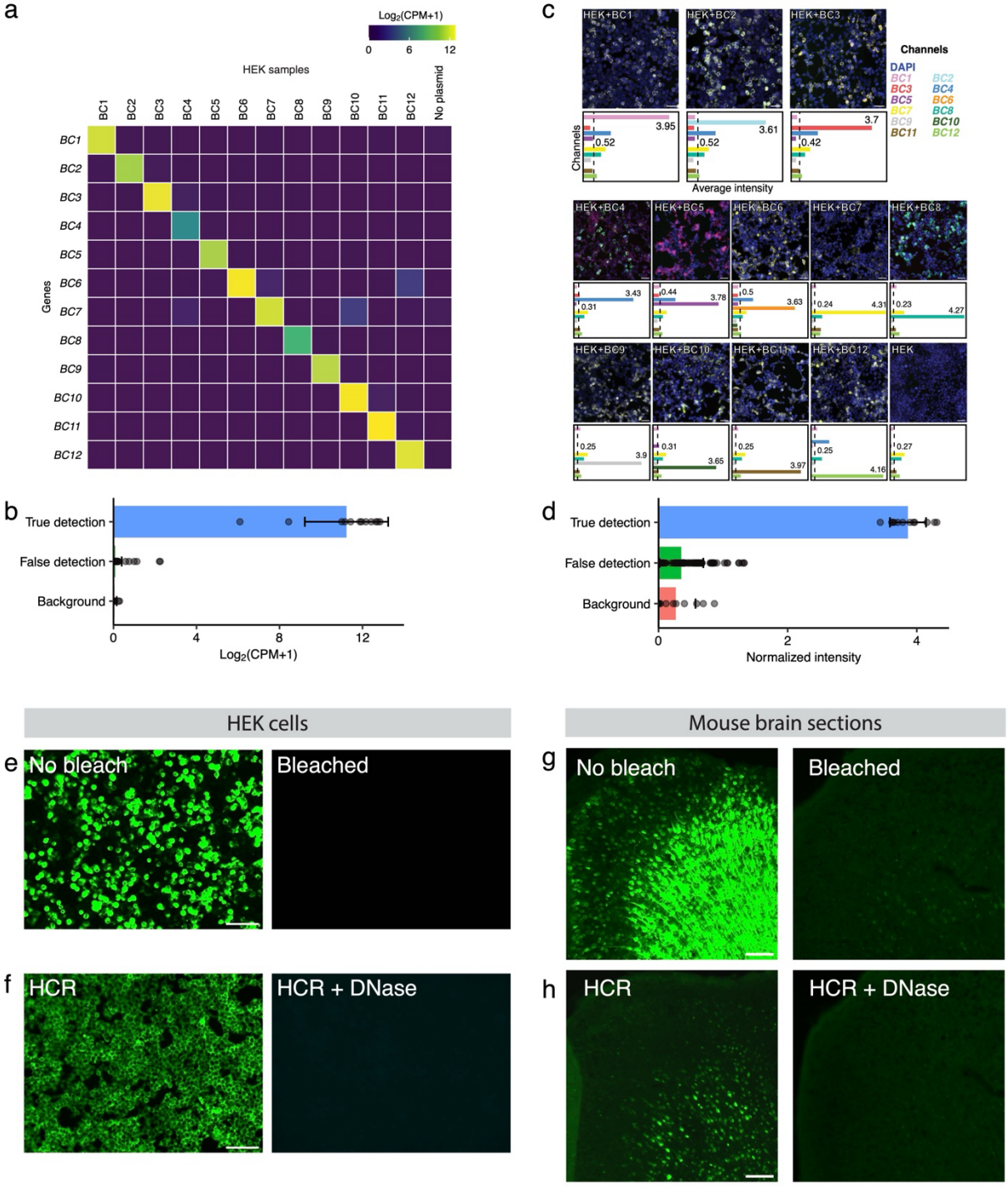


105

106 Extended Data Fig. 1 – related to Fig. 1

107 **a**. Representative fluorescent images of the localization of fusion proteins in different cellular compartments in fixed HEK cells.
108 Scale bar: 10 μm. **b**, Representative images of the fluorescent labeling using fusion proteins of nuclear resuspension extracted
109 from HEK cells. Scale bar: 20 μm. **c**, Efficiency of fluorescent labeling of cells and nuclei, measured as percentage of HEK
110 cells labeled with fusion proteins on fixed slices (cell culture), percentage of nuclei labeled with fusion proteins on nuclear
111 suspension extracted from HEK cells, and percentage of nuclei labeled with fusion proteins during FACS. Percentages were
112 normalized to the percentage of cell culture for each protein. **d**, Scatter plot showing the gate of sorting labeled HEK nuclei
113 (DAPI+ and GFP+) extracted from HEK cells and HEK cells expressing Sun1-GFP. **e**, Left: pairwise Euclidean distance
114 between any two of the 50 BCs cataloged in this study. Right: box plot showing the average Euclidean distance of individual
115 BCs to other BCs.

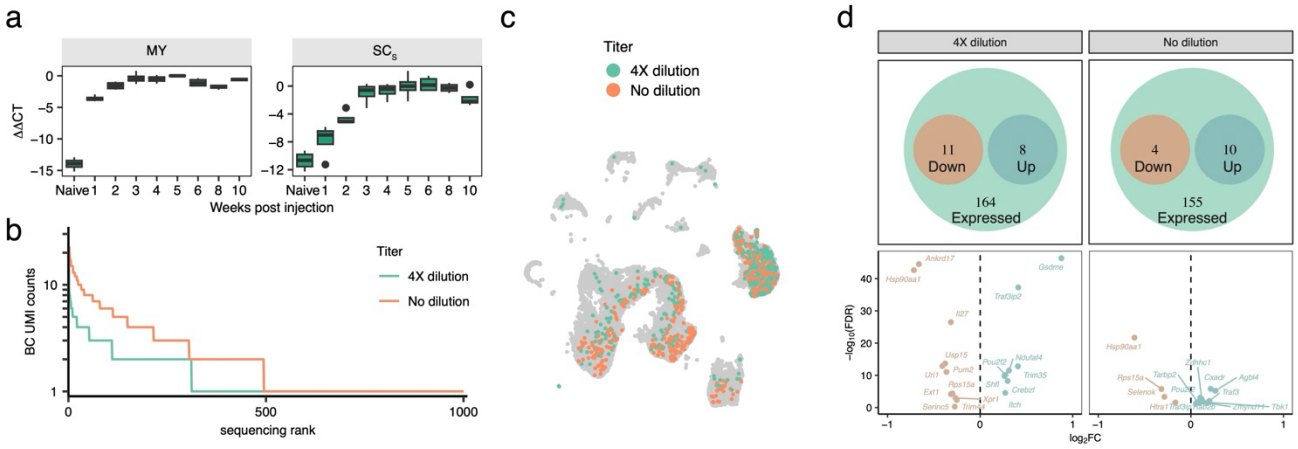
Extended Data Fig. 2



Extended Data Fig. 2 – related to Fig. 1

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

Extended Data Fig. 3



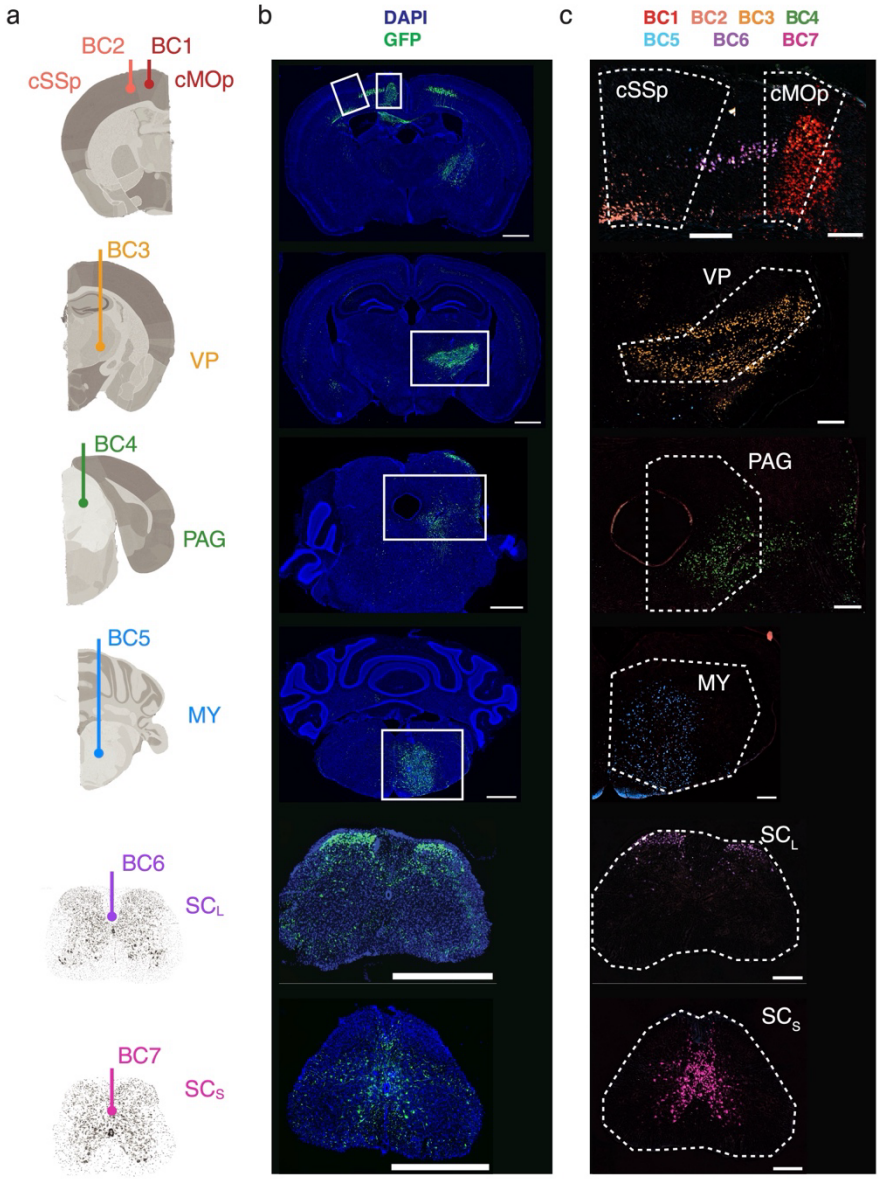
e

	Sequencing depth	BC detection	
	(UMIs/nucleus)	(Percentile by expression rank)	(Pct. BC+ nuclei)
10X Genomics	8.3 - 11.1k	57.5 - 70.1	0.2 - 0.58
Parse	19.9 - 20.6 k	49.7 - 58.5	0 - 0.06

Extended Data Fig. 3 – related to Fig. 1

a, Box plot showing the time kinetics of Projection-TAG expression in the cortex. Animals received MY or SC_S injection with the Projection-TAG AAV (expressing BC2), cortical samples were collected at different time points post injection, and the Projection-TAG expression is measured using qPCR (n=4 for each time point and each injection). ΔCT is calculated as the CT value of BC2 minus that of GAPDH for each sample, ΔΔCT is calculated as the ΔCT at each time point minus the ΔCT at 4 weeks post injection. **b**, Plot showing the detection rate of Projection-TAGs (BCs) in 1,000 downsampled nuclei from each titer group in snRNA-seq. Nuclei were ranked based on their Projection-TAG detection rate. **c**, RNA-UMAP of 10,000 nuclei (1,000 downsampled BC+ nuclei from each titer group, colored by titer, and 8,000 downsampled BC- nuclei, colored grey). **d**, DE genes in response to viral infection. DE analysis was performed comparing the expression of genes annotated in the Gene Ontology list GO:0009615 (response to virus) between Projection-TAG+ nuclei and Projection-TAG- nuclei (downsampled to match the T-type distribution and nuclei counts to the Projection-TAG+ nuclei) in each tier group. Top panels show the number of upregulated (avg_log2FC > 0) and downregulated (avg_log2FC < 0) viral responding genes among the viral responding genes expressed (average expression > 0.5) in snRNA-seq nuclei from each titer group. Fisher's exact test was performed to test if viral responding genes are significantly enriched among DE genes (p = 0.81 and 0.77 for undiluted and diluted group, respectively). Bottom panels show the individual DE viral responding genes. **e**, Key sequencing and Projection-TAG detection metrics from snRNA-seq libraries prepared using different commercial kits from nuclei of the same biological samples (n = 2 for each kit).

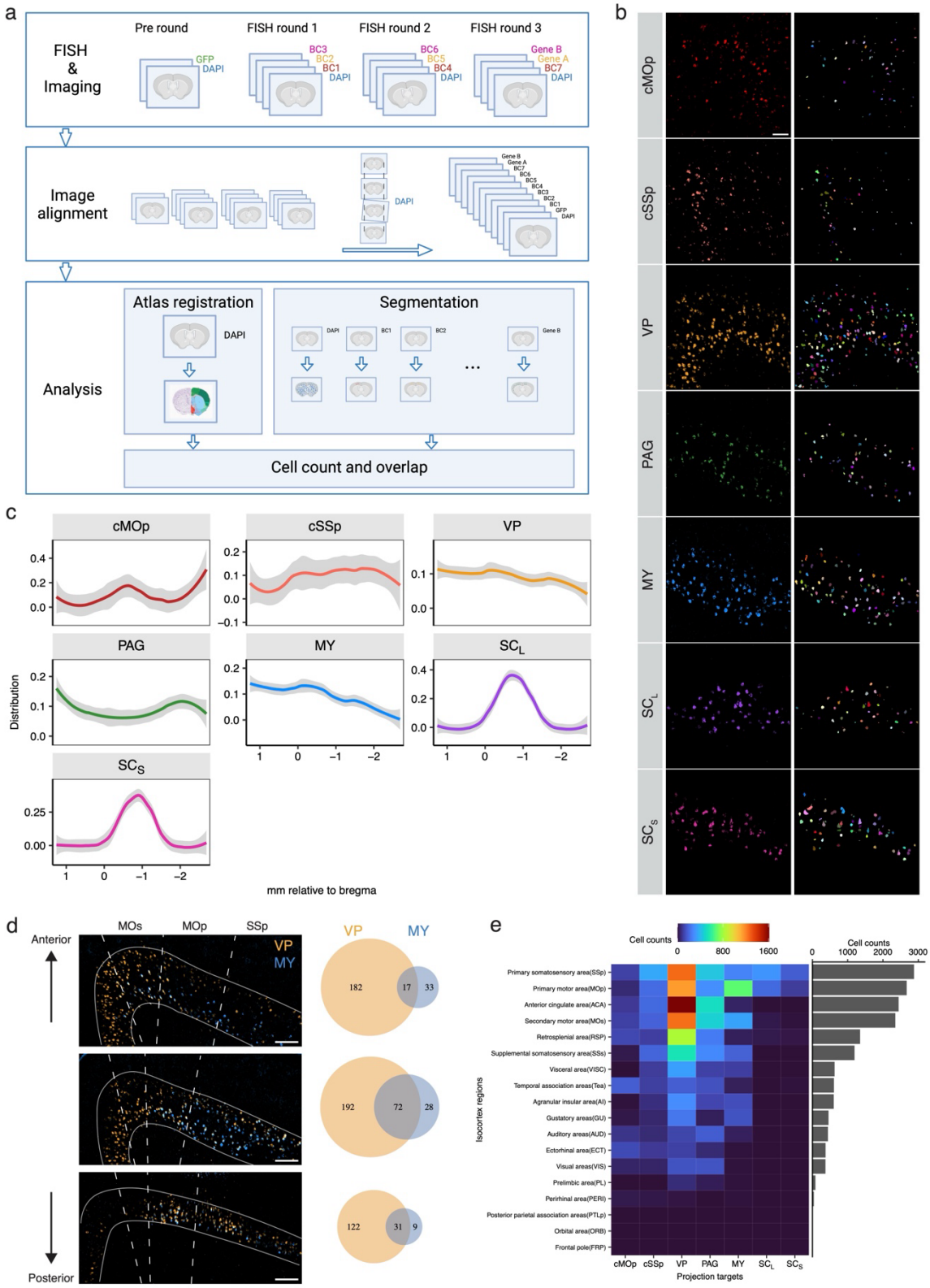
Extended Data Fig. 4



Extended Data Fig. 4 – related to Fig. 1

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

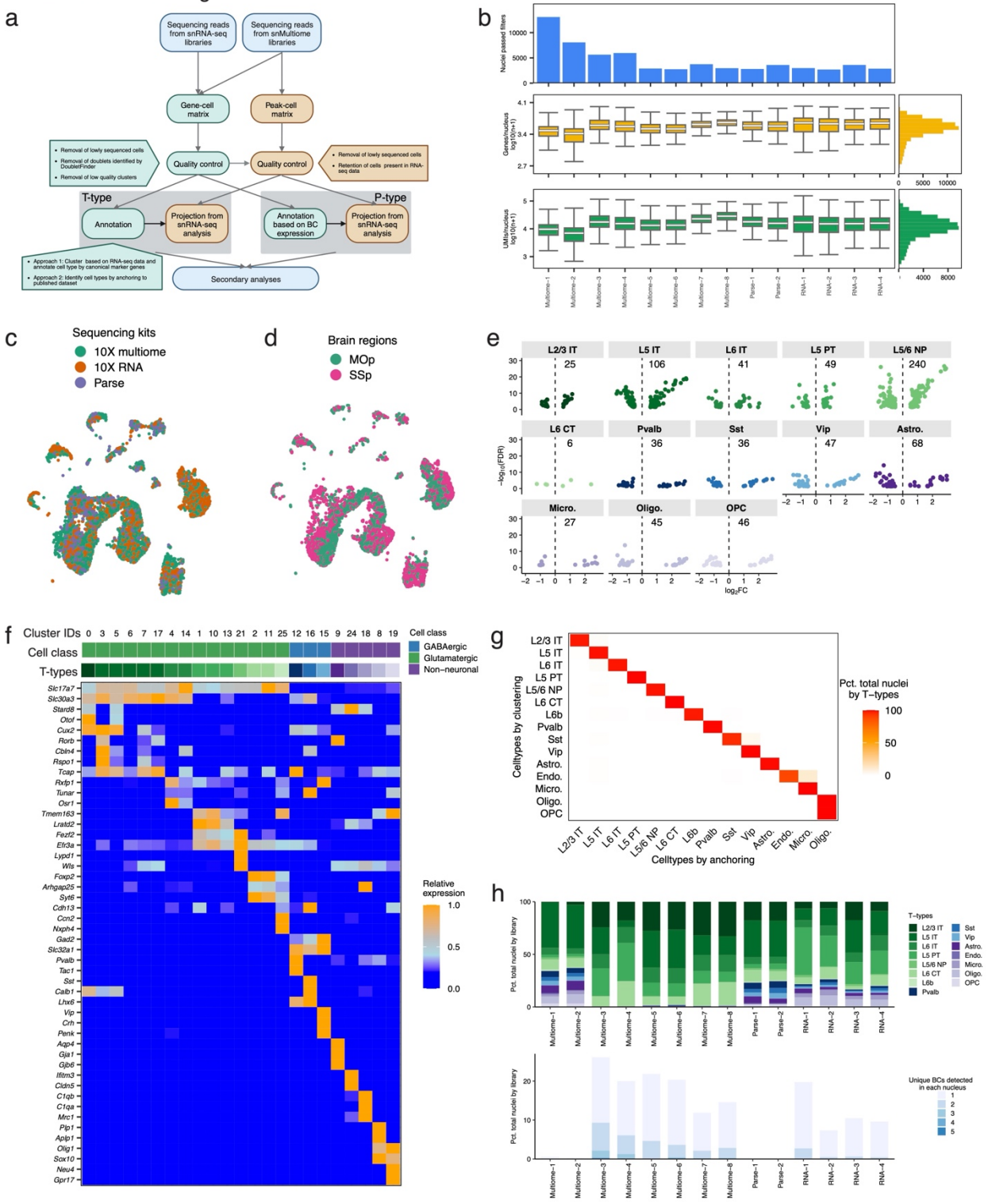
Extended Data Fig. 5



Extended Data Fig. 5 – related to Fig. 1

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

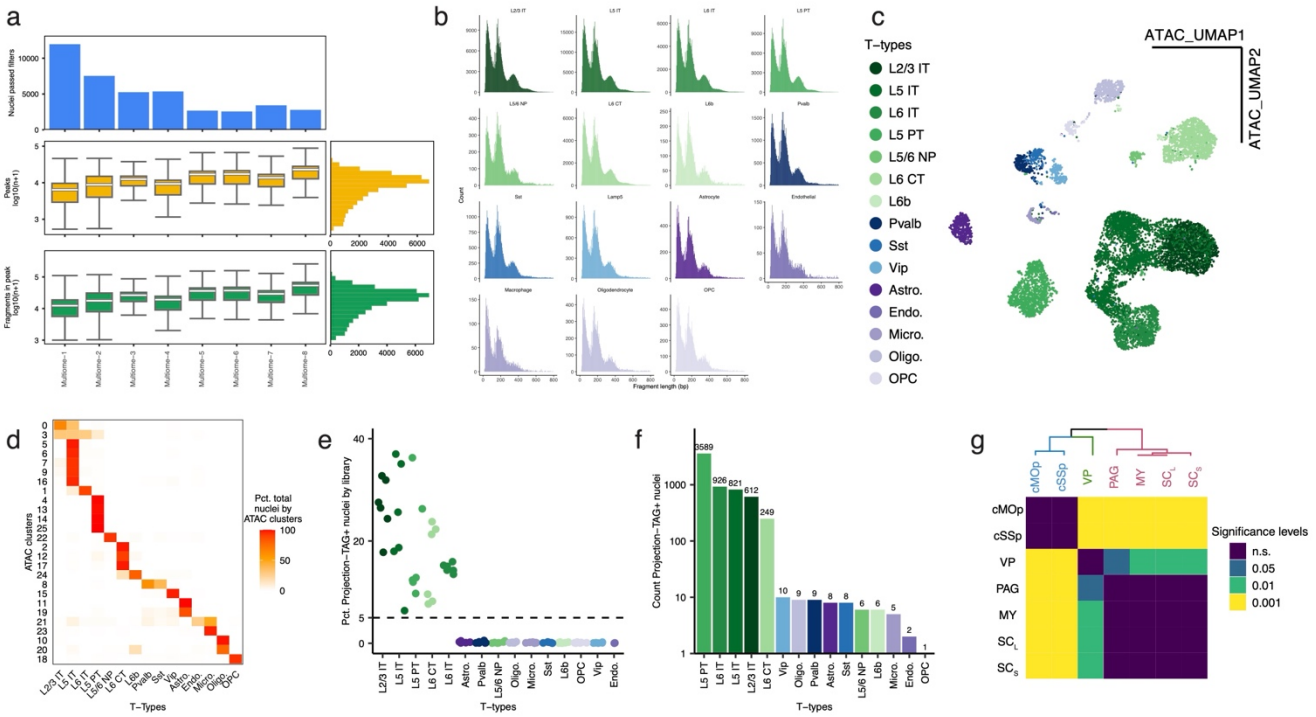
Extended Data Fig. 6



Extended Data Fig. 6 – related to Fig. 2

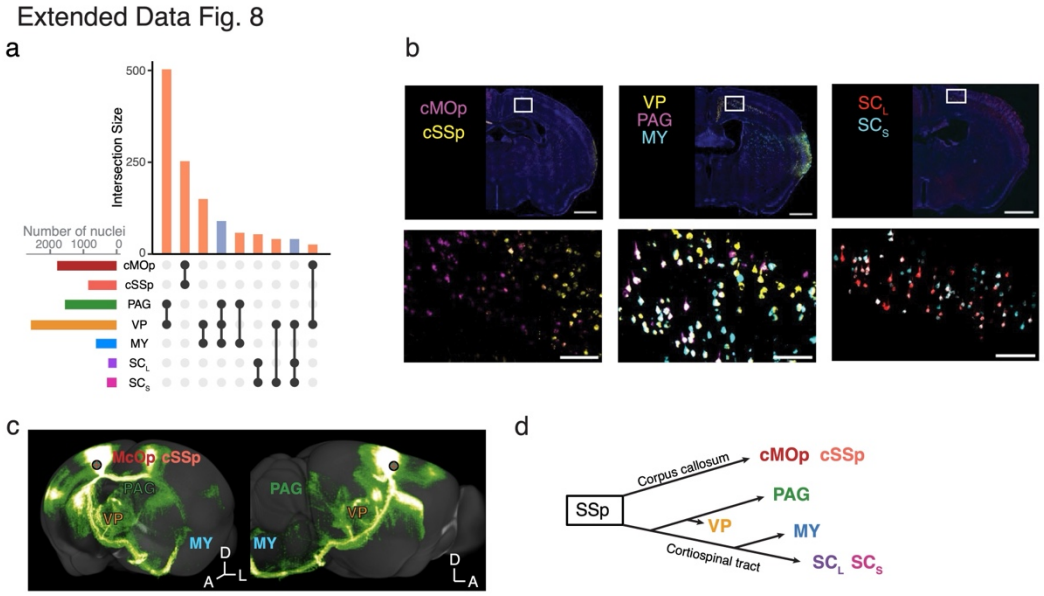
a, Workflow of the data analysis of the snRNA-seq and snATAC-seq and T-type and P-type annotation. **b**, snRNA-seq library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of genes per nucleus (log10 transformed) and the bottom row displays the number of UMIs per nucleus. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is aggregated across all samples and displayed on the horizontal histogram. **c** and **d**, RNA-UMAP of 1,000 downsampled nuclei per group. Nuclei were colored by sequencing kits (c) or brain regions (d). **e**, Volcano plots of DE genes between nuclei from MOp (avg_log2FC > 0.5) and SSp (avg_log2FC < -0.5) in individual T-types with at least 30 nuclei from each brain region. Numbers indicates total DE genes in each T-type. **f**, T-type assignment to each cluster identified in snRNA-seq data. Heatmap showing the distinct expression pattern of marker genes that are used to assign T-types. See h for the colors of T-types. **g**, Heatmap showing the correspondence of T-type assignment of nuclei in snRNA-seq data between the manual classifications (f) and the classifications assigned by anchoring to a published scRNA-seq dataset of mouse motor cortex[44]. Plot displays percentage of nuclei within the initial T-type assignment that is assigned to each cell type after anchoring to the published dataset. Overall, 98.2 ± 3.1% nuclei in each cluster were assigned the same T-types between the two approaches. **h**, Composition of T-types (top) and Projection-TAG+ nuclei grouped by number of unique BCs detected per nucleus (bottom) in each snRNA-seq library.

Extended Data Fig. 7



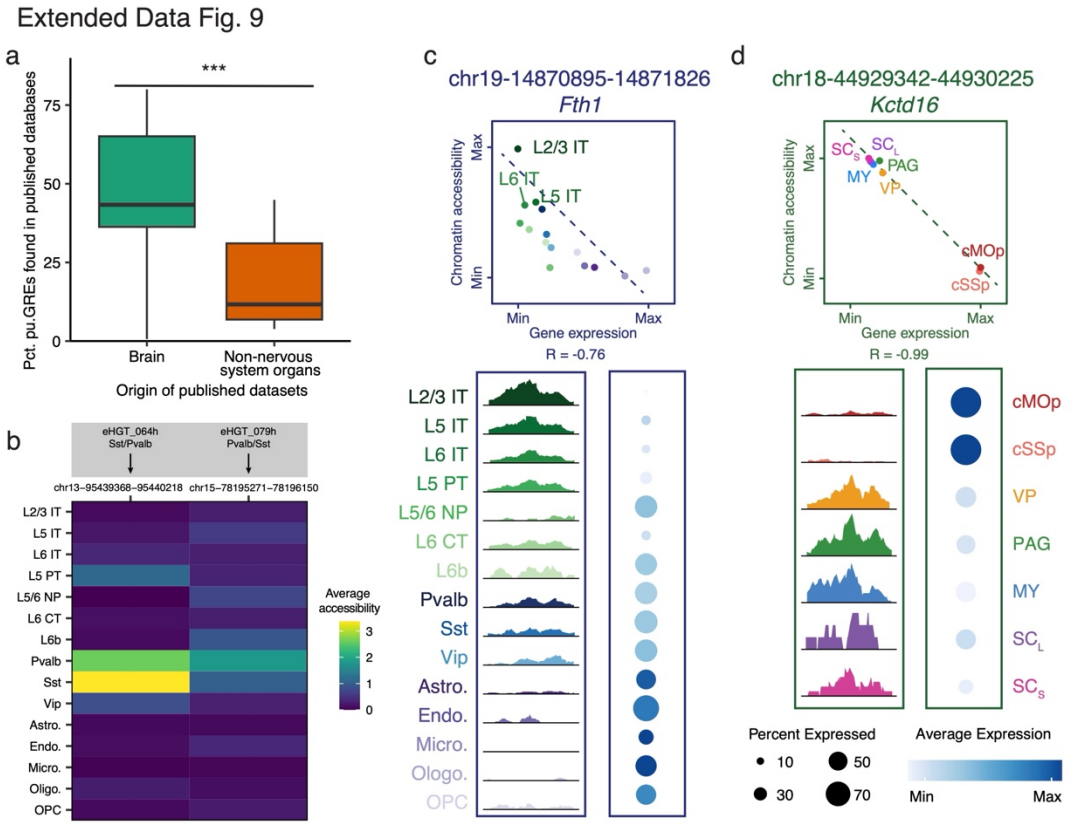
Extended Data Fig. 7 – related to Fig. 2

a, snATAC-seq library metrics. Top row displays number of nuclei passed quality control. Middle row displays box plots of number of peaks per nucleus (log10 transformed) and the bottom row displays the number of fragments that overlap with peaks per nucleus. Boxes indicate quartiles and whiskers are 1.5-times the interquartile range (Q1-Q3). The median is a white line inside each box. The distribution is aggregated across all samples and displayed on the horizontal histogram. **b**, Distribution of snATAC-seq fragment lengths in individual T-types. **c**, ATAC-UMAP of downsampled 10,000 nuclei of snATAC-seq. Nuclei are colored by T-types. **d**, Correspondence of ATAC clusters (row) and T-types (column) of nuclei profiled combinatorially by snRNA-seq and snATAC-seq. Plot displays percentage of nuclei within each ATAC cluster that is assigned to each T-type. **e**, Dot plot showing the percentage of nuclei that are Projection-TAG+ in each T-type in each snRNA-seq library prepared by FACS. Dotted line indicates the cutoff for calling T-types enriched of projection neurons (5%). **f**, Bar plot showing the counts of Projection-TAG+ nuclei in each T-type in snRNA-seq data. **g**, Heatmap showing the significance levels of the Bonferroni-corrected *p* values of pairwise Fisher's exact tests of T-type distribution between any of the two P-types. Hierarchical clustering was performed based on the Bonferroni-corrected *p* values of each P-type to each of the other P-types.



Extended Data Fig. 8 – related to Fig. 3

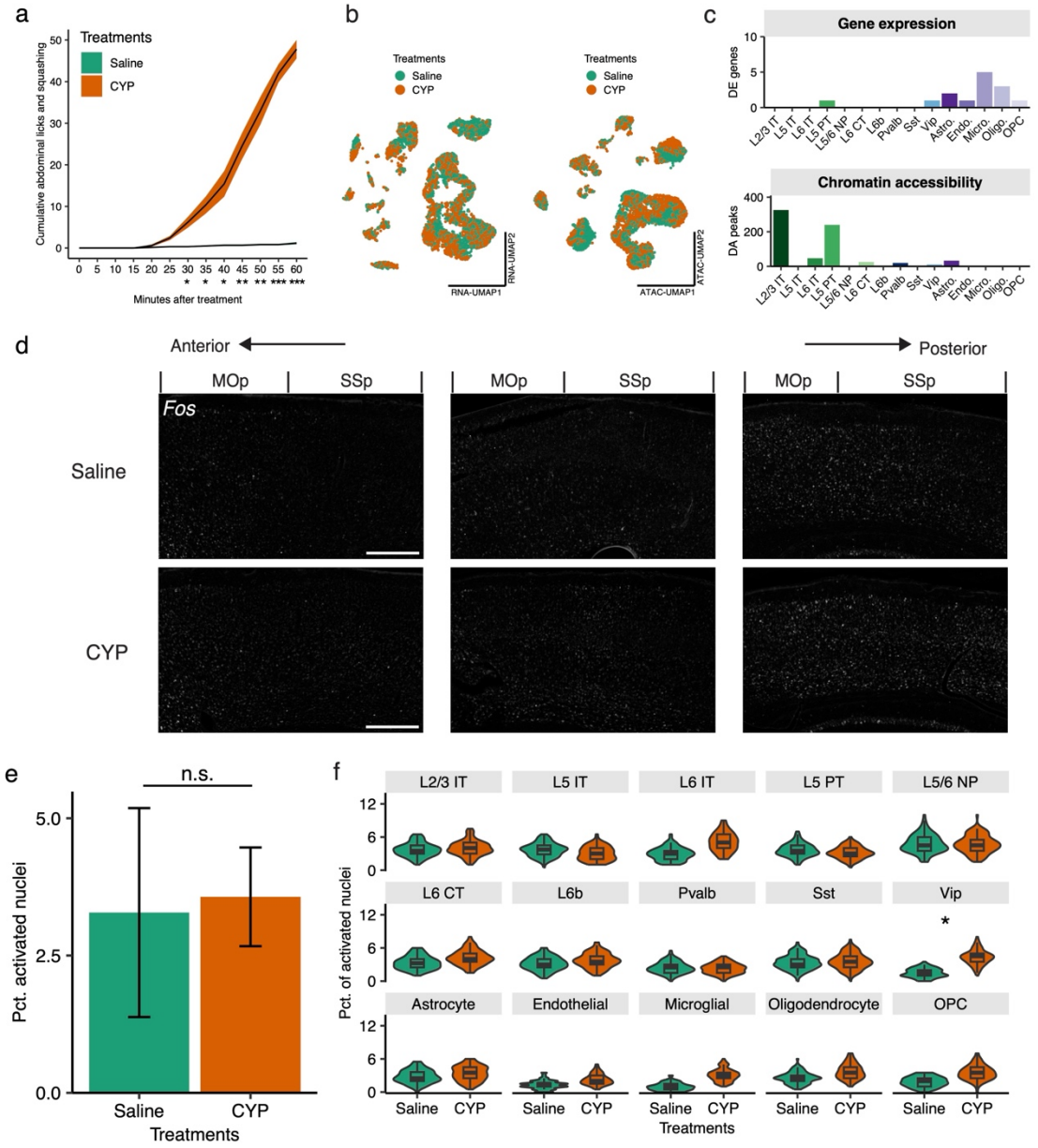
a, Upset plot showing the overlapping across P-types. Bar plot shows the counts of Projection-TAG+ nuclei in each P-type. Upset plot shows the combinations of overlapping across P-types with at least 10 nuclei. **b**, Representative FISH images showing projection neurons labelled by Projection-TAGs in the coronal section showing motor cortex from different anterior-posterior sections (top) and zoomed-in images (bottom) of the highlighted region demonstrate expression of two unique Projection-TAGs injected into downstream cMOp and cSSp; middle: demonstrate expression of three unique Projection-TAGs injected into the VP, PAG, and MY; right: demonstrate expression of two unique Projection-TAGs injected into SC_L and SC_S). Scale bar: 1 mm (top), 100 μm (bottom). **c** Tracing the brain-wide axon collaterals of SSp neurons. Images were acquired from Allen Brain Atlas data portal (https://connectivity.brain-map.org/projection/experiment/112935169)[2]. **d**, a summary diagram of the anatomical hierarchy of axonal projections from SSp to the seven target regions.



Extended Data Fig. 9 – related to Fig. 4

a, Box plot showing the overlapping of putative GREs identified from this study to the putative GREs identified by studies from the ENCODE consortium[79,118]. Overlap is significantly higher with studies of the brain compared to those of non-brain organs (***p = 8.26e-05, t-test). Brain studies: ENCSR422CGZ, ENCSR694ZBB, ENCFF426KUB, and ENCSR410XMU. Studies of organs that are not the nervous system: ENCSR052AZT (stomach), ENCSR100TMZ (lung), ENCSR296UQL (liver), and ENCSR180OIF (limb). **b**, Heatmap showing the chromatin accessibility of two T-type-specific pu.Enhancers in individual T-types. The peaks of the two pu.Enhancers overlap with the genomic coordinates liftover from two functional enhancers previously identified based on human snATAC-seq data[73]. Labels in the grey boxes show the functional enhancers and the cell types in which the enhancers exhibited highest activity in experimental validation from the previous study. **c**, Top: Scatter plot showing the correlation between the accessibility of peak chr19-14870895-14871826 and the expression of gene *Fth1* in individual T-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 chr19-14870895-14871826, displayed as the average fraction of transposase-sensitive fragments per nucleus in each T-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 *Fth1* in each T-type. Colors indicate T-types. **d**, Top: Scatter plot showing the correlation between the accessibility of peak chr18-44929342-44930225 and the expression of the gene *Kctd16* in individual P-types (chromatin accessibility and gene expression are normalized to their max values). bottom left: chromatin accessibility at the genomic locus of chr18-44929342-44930225, displayed as the average fraction of transposase-sensitive fragments per nucleus in each P-type (grouped by 50 bins per displayed genomic region). Bottom right: expression of *Kctd16* in each P-type. Colors indicate P-types.

Extended Data Fig. 10



Extended Data Fig. 10 – related to Fig. 5

a, Cumulative scores of quantified behaviors between mice treated with saline and CYP. CYP treatment resulted in significant increases in the spontaneous behaviors like abdominal licking and squashing behaviors compared to saline treatment. Ribbons show SEM. ** p = 0.0034, one-way ANOVA. p = NaN (0 min), NaN (5 min), NaN (10 min), NaN (15 min), 0.25 (20 min), 0.07 (25 min), *0.05 (30 min), *0.04 (35 min), *0.04 (40 min), **0.001 (45 min), **0.006 (50 min), ***0.0005 (55 min), ***0.0003 (60 min), respectively, post-hoc one-sided t-test, n=5-6 mice/group. **b**, RNA-UMAP (left) and ATAC-UMAP (right) of 5,000 downsampled nuclei from each treatment group, colored by treatments. **c**, Bar plots showing number of DE genes (top) and DA peaks (bottom) between nuclei from CYP (avg_log2FC > 0.5 or < -0.5, FDR < 0.05) and saline in individual T-types. **d**, Representative FISH images showing *Fos* expression in select cortex areas (top: 1.33 mm, middle: -0.95 mm, bottom -1.67 mm relative to bregma). Scale bar: 400 µm. **e**, Bar plot showing the average percentage of activate nuclei by Act-seq in individual snRNA-seq libraries of different treatments. Error bars show standard deviation. (p = 0.74, t-test). **f**, Violin plots and box plots showing the percentage of activate nuclei by Act-seq in individual T-types between treatments. Two hundred nuclei per T-type and treatment were randomly sampled to calculate the percentage and the process repeat for a total of 1,000 times. The p values are calculated as the fraction of repeats where percentage of activated nuclei from CYP-treated animals is lower than that of saline-treated (p = 0.42 for L2/3 IT, p = 0.61 for L5IT, p = 0.11 for L6 IT, p = 0.52 for L5 PT, p = 0.47 for L5/6 NP, p = 0.28 for L6 CT, p = 0.34 for L6b, p = 0.47 for Pvalb, p = 0.40 for Sst, *p = 0.02 for Vip, p = 0.31 for Astrocyte, p = 0.16 for Endothelial cells, p = 0.05 for Microglial cells, p = 0.24 for Oligodendrocytes, and p = 0.10 for OPC).