Connectome-seq: High-throughput Mapping of Neuronal Connectivity at Single-Synapse Resolution via Barcode Sequencing

Danping Chen[1,4], Alina Isakova[2,4], Wan Joe Zhou[1], Yunming Wu[3], Boxuan Simen Zhao[1,5]

1 Department of Cell and Developmental Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA.
2 Department of Bioengineering, Stanford University, Stanford, CA, USA.
3 Department of Biology and Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA.
4 These authors contributed equally
5 Lead contact

Abstract:

Understanding neuronal connectivity at single-cell resolution remains a fundamental challenge in neuroscience, with current methods particularly limited in mapping long-distance circuits and preserving cell type information. Here we present Connectome-seq, a high-throughput method that combines engineered synaptic proteins, RNA barcoding, and parallel single-nucleus and single-synaptosome sequencing to map neuronal connectivity at single-synapse resolution. This AAV-based approach enables simultaneous capture of both synaptic connections and molecular identities of connected neurons. We validated this approach in the mouse pontocerebellar circuit, identifying both established projections and potentially novel synaptic partnerships. Through integrated analysis of connectivity and gene expression, we identified molecular markers enriched in connected neurons, suggesting potential determinants of circuit organization. By enabling systematic mapping of neuronal connectivity across brain regions with single-cell precision and gene expression information, Connectome-seq provides a scalable platform for comprehensive circuit analysis across different experimental conditions and biological states. This advance in connectivity mapping methodology opens new possibilities for understanding circuit organization in complex mammalian brains.

Introduction

Understanding synaptic wiring diagrams of the brain, known as the connectomes, is a fundamental goal in neuroscience. This complex network of neuronal connections underlies all brain functions, from simple reflexes to complex cognitive processes. Breakthroughs in connectomics have provided unprecedented insights into brain organization. An early breakthrough was the complete reconstruction of the *C. elegans* connectome[1], representing the first comprehensive map of a nervous system with its approximately 300 neurons and their synaptic connections. Recent achievements include the complete reconstruction of the adult *Drosophila* brain connectome, which mapped approximately 140,000 neurons and over 50 million connections, representing the largest connectome obtained to date[2–5]. The *Drosophila* connectome already has a profound effect on neuroscience, enabling researchers to trace neural circuits, generate hypotheses about their function, and develop circuit models rooted in actual connectivity. This resource exemplifies the transformative potential of comprehensive connectome data.

Serial-section electron microscopy (ssEM) has been the foundational method enabling these comprehensive connectome reconstructions in invertebrates. Adaptations of ssEM techniques for the more complex vertebrate nervous system have also driven major progress in mapping neuronal connectivity at single-cell resolution in the vertebrate brain. Early and highly influential work using ssEM in the retina has provided critical insights into the neural circuitry underlying specific functions, such as direction selectivity in retinal ganglion cells[6–8]. More recently, dense reconstructions of various brain regions have revealed fundamental principles of circuit organization. For example, study in the medial entorhinal cortex demonstrated precise axonal synapse sorting that supports temporal processing of information[9]. Dense reconstructions of adult mouse somatosensory cortex uncovered novel distance-dependent axonal targeting principles of inhibitory and excitatory connectivity[10], while developmental studies in the same region revealed how inhibitory synaptic specificity emerges through distinct postnatal trajectories[11]. Comprehensive analysis of mouse visual cortex demonstrated structured functional connectivity rules governing inter-areal projections[12], subsequent work showed widespread like-to-like connectivity principles across cortical layers and areas[13], and detailed studies of thick-tufted layer 5 pyramidal cells revealed cell-type specific organization of proximal and extended connectivity patterns[14]. Cerebellar circuit mapping revealed how non-random redundant connectivity motifs help balance pattern separation with noise resilience[15]. While these studies have provided unprecedented insights into local circuit architecture, they remain fundamentally limited to volumes of a few hundred micrometers due to the technical challenges of imaging and reconstructing larger tissue volumes. Moreover, while invertebrate circuits like those in *C. elegans* and *Drosophila* exhibit a higher degree of stereotypy compared to mammals, it is now recognized that these circuits also possess individual variability[16,17]. Mammalian circuits exhibit a greater magnitude of individual variation in connectivity at the single-cell level. This biological variability means that studying mammalian circuits requires methods that are both practical and scalable enough to characterize connectivity in individual subjects, especially when aiming to correlate circuit structure with neural dynamics or behavior.

Meeting this challenge requires new approaches that can achieve both high throughput mapping and single-cell resolution in larger mammalian brains. Various techniques exist for mapping mammalian circuit connectivity at non-single-cell scales[18,19] or single cells at low throughput[20–22]. Light microscopy-based methods like mGRASP[23] and SynView[24] enable visualization of synaptic contacts between specific neuronal populations, while genetically modified viral tracers allow circuit component identification without requiring direct synaptic visualization. However, these approaches face important limitations - light microscopy often lacks the resolution to definitively identify synapses, viral tracing can be constrained by tropism and potential toxicity, and most importantly, these methods are inherently designed to examine a specific subset of connections based on genetic markers or viral targeting, making comprehensive, unbiased circuit mapping challenging.

Recent advances in sequencing-based approaches have transformed our ability to map neural connectivity at scale. MAPseq pioneered this revolution by using barcoded Sindbis virus to trace thousands of neuronal projections simultaneously[25], followed by BARseq which preserved spatial context through in situ sequencing[26], and BRICseq which enabled multiplexed tracing from multiple sites[27]. Building on these advances, researchers developed approaches using barcoded rabies virus to map synaptic connectivity. Notable developments include RABID-seq's

demonstration of the concept in glial cells[28], SBARRO's proof-of-principle in controlled sparse neuronal cultures[29], and recent work by Zhang et al. and Shin et al. that revealed cell type-specific connectivity patterns in cortical circuits[30,31]. However, these methods face significant technical limitations: Sindbis virus-based approaches cannot reveal post-synaptic partners and exhibit severe toxicity due to high-level viral payload expression, typically killing cells within days; whereas rabies-based methods struggle with barcode mixing in densely connected neurons and viral toxicity that may compromise circuit integrity.

To address these challenges, we developed Connectome-seq, an approach that attempts to overcome key limitations of existing methods while offering unique advantages for circuit mapping. Leveraging minimally toxic adeno-associated viruses (AAV) vectors and combining molecular engineering with high-throughput sequencing, Connectome-seq enables unbiased mapping of neuronal connections at single-synapse resolution across large brain volumes. This approach simultaneously captures neuronal identity, gene expression, and synaptic connectivity without the need for pre-selection of specific cell types or limitations based on cellular density.

We demonstrated the utility of Connectome-seq in the mouse pontocerebellar circuit, a universally conserved circuit in mammals that exemplifies the challenges of mapping long-distance connections in the mammalian brain. Connectome-seq identified thousands of connections across this circuit, including both known projections and potentially novel synaptic partnerships. The high-throughput, unbiased nature of our method enables comprehensive analysis of connectivity patterns while preserving critical information about cellular identity and molecular characteristics. This combination of features makes Connectome-seq particularly well-suited for systematic circuit mapping across different experimental conditions and biological states.

Results

The Connectome-seq workflow consists of three key steps: (1) molecular engineering of synaptic barcoding constructs and viral delivery of transgenes into target brain regions, (2) development of a parallel single-nucleus and single-synaptosome sequencing protocol, and (3) computational reconstruction of the connectome. Here, we present the development process and detailed results in each step, highlighting the advances that enable high-throughput synaptic mapping, and the biological insights gained from applying this method to map the mouse pontocerebellar circuit.

1. Molecular engineering to translate a synapse into a pair of RNA barcodes

Connectome-seq relies on the labeling of synaptic connections using engineered protein-RNA complexes to translate a synapse into a pair of unique RNA barcodes. The intended outcome is that one barcode will be associated with the pre-synaptic neuron and the other with the post-synaptic neuron, allowing for later profiling via high-throughput sequencing (Fig. 1A). Two protein anchors, termed SynBar (synaptic barcoding), were designed: PreSynBar to target presynaptic terminals and PostSynBar to target postsynaptic sites (Fig. 1B and Extended Data Fig. 1A). PreSynBar consists of a modified neurexin 1β protein fused to the large fragment of split-GFP (GFP1-10) and an RNA-binding domain, λN22 peptide. PostSynBar is based on neuroligin 1, also fused to the complementary small fragment of split-GFP (GFP11) and the same RNA-binding domain. These trans-membrane protein constructs are designed to interact at synapses,

reconstituting GFP and bringing together their associated RNA barcodes to anchor at the synaptic membranes. The protein scaffolds were largely adopted from the SynView constructs, where split-GFP fragments are embedded inside the synaptic proteins rather than at their N-terminals. This design choice aims to ensure GFP reconstitution occurs specifically upon proper synaptic protein interaction. An epitope tag (V5 for PreSynBar and HA for PostSynBar) was added to the extracellular domain of each construct to visualize tagged synapses via antibody staining.

An important design consideration for PreSynBar was the positioning of the λN22 RNA-binding domain. While PostSynBar tolerated the addition of λN22 at its extreme C-terminus, PreSynBar required internal placement within its C-terminal region. In trans-interaction assays, PreSynBar with λN22 at the extreme C-terminus failed to show proper interaction with PostSynBar, while PreSynBar with internal C-terminal λN22 showed evidence of trans-interaction (Extended Data Fig. 1B). This finding is consistent with previous studies showing the sensitivity of neurexin's C-terminal region to modifications, suggesting the importance of preserving its native structure for proper membrane targeting.

Two corresponding RNA barcodes, named PreRNA and PostRNA, were designed with the aim of efficient expression, synaptic localization, and detection by sequencing (Fig. 1C). Each barcode contains several domains designed to perform specific functions: PCR handles for amplification, a 30-nucleotide random region (N30) with sufficient complexity to yield unique sequences for each neuron, two copies of BoxB sequence that can be tightly bound by the λN22 peptide on SynBar proteins, and a capture sequence (CS1 or CS2) matching the 10x Genomics feature barcoding technology. We used large-scale Golden Gate assembly to insert the random N30 region into the plasmid to create a highly complex plasmid library that was later packaged into AAV as a complex viral library. To ensure efficient expression and transport, the barcodes include a poly-adenylated region for RNA Polymerase II-driven transcription, which appeared to improve nuclear export and synaptic abundance compared to Pol III-driven expression in our experiments (Extended Data Fig. 1C). To facilitate the necessary long-distance transport of PreRNA to distal presynaptic terminals, we explored redesigning the PreRNA barcodes to naturally localize to neurites. Drawing upon previous research identifying specific 5' and 3' untranslated regions (UTRs) that enhance mRNA localization[32,33], we tested several UTR modifications for their ability to promote PreRNA abundance in distal regions. Specifically, the addition of the 5' UTR of RPL7 led to the greatest increase in the abundance of PreRNA in both cytosolic and synaptosomal fractions of the postsynaptic region among all designs tested in our experiments, and was therefore incorporated into the final PreRNA design (Extended Data Fig. 1D).

To assess the stability of BoxB-λN22 interaction in our system, we performed RNA immunoprecipitation (RIP) assays using anti-GFP antibodies. The RIP experiments showed >500-fold enrichment of barcodes when using λN22-tagged SynBar proteins compared to untagged controls (SynView), suggesting that this protein-RNA interaction is sufficiently stable to withstand biochemical isolation (Extended Data Fig. 1E).

We further assessed the interaction of PreSynBar and PostSynBar through an established cultured-cell-based strategy used for other synaptic labeling approaches[24]. In cis validation experiments (where both constructs are expressed in the same cell), we observed strong reconstituted split-GFP signals when PreSynBar and PostSynBar were co-expressed within individual HEK 293 cells and

cultured neurons (Fig. 1D). For trans validation (where the constructs are expressed in separate cells that are then co-cultured), we examined cells expressing each construct separately: when PreSynBar-expressing cells (detected by V5 staining in cyan) were co-plated with PostSynBar-expressing cells (detected by HA staining in magenta), we observed specific split-GFP reconstitution (green) at points of contact between the two cell populations, in both HEK 293 cells (Fig. 1E) and HEK 293-neuron mixed cultures (Extended Data Fig. 1F). The merged images showed the colocalization of all three signals at these cellular junctions, consistent with protein-protein interaction and GFP reconstitution both within single cells (cis) and between adjacent cells (trans).

2. Development of Parallel Nuclear and Synaptosome Isolation Strategy

A key feature of Connectome-seq is the ability to simultaneously capture both cellular identity and synaptic connectivity information. This is achieved through parallel profiling of neuronal nuclei and synaptosomes from the same tissue, where co-existing barcodes in these two structures link each cell to a synapse (Fig. 2A). Nuclei contain both cellular transcriptomes and copies of either PreRNA or PostRNA, allowing identification of cell types and their associated barcodes. During gentle brain tissue homogenization, the cell membranes connecting pre-synaptic terminals to axons and post-synaptic sites to dendrites are physically separated. The exposed membrane surfaces at both the pre- and post-synaptic compartments then reseal, forming structures called synaptosomes[34,35]. Since PreSynBar and PostSynBar proteins are connected across the synaptic cleft, tightly bound to the associated RNA barcodes, these protein-RNA complexes should remain intact as long as the synaptic interface is preserved. This allows synaptosomes to maintain the paired PreRNA and PostRNA information from connected neurons, providing direct evidence of synaptic connections. By matching barcodes between nuclei and synaptosomes, we can link the identity of connected neurons and map their connectivity patterns.

To implement this strategy, we designed AAV constructs carrying Cre recombinase and SynBar components (Extended Data Fig. 2A). The Cre-expressing virus induces the expression of a Cre-dependent nuclear-localized fluorescent reporter (Sun1-Tag[36]), labeling nuclei of infected neurons with membrane-anchored GFP to enable their isolation. An important consideration in our virus design was to ensure PreSynBar expression and targeting. Initial versions of PreSynBar virus showed inconsistent protein localization to neuronal processes. We discovered that including two transcriptional cassettes – one driving PreSynBar protein and another expressing PreRNA – improved targeting efficiency (Extended Data Fig. 2B). This two-cassette design consistently achieved better axonal localization compared to single-cassette versions.

We next developed a protocol to purify and enrich both nuclei and synaptosomes from the same brain regions (Fig. 2B): Using gentle homogenization in a sucrose-based buffer (SET buffer), we performed sequential centrifugation to fractionate the homogenate into crude nuclei and synaptosomes and used flow cytometry for fluorescence-based selection to enrich the population containing barcodes.

For nuclear isolation, we used Sun1-Tag GFP signal for specific enrichment. A key consideration was the choice of DNA stain – while violet dye like DAPI and Hoechst are commonly used for nuclear sorting, their spectral overlap with GFP confounds the selection of GFP-positive events.

We therefore used SiR-DNA, a far-red DNA stain, to enable clear separation of DNA content signal from GFP fluorescence. Our optimized sorting strategy employed sequential gating (Fig. 2C): First, we selected events based on SiR-DNA signal intensity to enrich for nuclei containing approximately single-genome DNA content, which helped discriminate singlet nuclei from aggregates and debris. Within this singlet population, we then identified GFP-positive nuclei derived from infected neurons. This approach consistently yielded high-quality nuclear preparations, with GFP expression and nuclear singlet status confirmed by confocal microscopy post collection (Fig. 2D). RNA quality is also critical for single-nucleus RNA sequencing. To preserve RNA integrity, we minimized post-sorting handling of nuclei by sorting GFP-positive singlet nuclei directly into a collection buffer and immediately loading the sample onto the 10x Genomics microfluidic chip. This streamlined approach ensured consistent, high-quality nuclear RNA for efficient cell typing.

The next critical component of Connectome-seq is the ability to isolate and enrich double-barcoded synaptosomes. We found that crude synaptosomes (P2 fraction) provided a suitable starting material for our flow cytometry-based enrichment workflow. This approach effectively excluded non-synaptosomal fractions and debris while preserving intact synaptosomes, achieving similar goals to traditional sucrose gradient purification but with reduced processing time. We developed a systematic sequential gating strategy using flow cytometry to enrich synaptosomes based on their size, internal complexity, membrane integrity, and the presence of PreSynBar and PostSynBar-associated fluorescent signals (Fig. 2E). This flow cytometry-based approach proved crucial for reliable synaptosome detection and sorting.

Our initial plan to identify double-barcoded synaptosomes relied on flow cytometry detection of reconstituted split-GFP signal from interacting PreSynBar and PostSynBar proteins. However, we discovered that the native fluorescent signal fell below the detection sensitivity of standard flow cytometers. This limitation led us to explore several alternative approaches. First, we used axon- and dendrite-targeted fluorescent proteins (PreTag-GFP and PostTag-RFP, Extended Data Fig. 2C, D) for synaptosome labeling. This approach provided bright fluorescent signals suitable for flow cytometry and enabled the enrichment of connected synaptosomes via co-existing green and red fluorescence (Extended Data Fig. 2E). Confocal imaging of sorted double-positive synaptosomes revealed that most structures were true singlets containing both fluorescent proteins, verifying the robustness of our sequential gating strategy (Extended Data Fig. 2F). However, subsequent analysis of ambient barcode RNA showed high levels of contaminating PreRNA and PostRNA among sorted structures, specifically, a high level of PreRNA in postsynaptic nuclei (Extended Data Fig. 2G), indicating that fluorescent proteins not strictly localized to synaptic terminals cannot be used as synaptic barcoding complexes. This finding highlighted the importance of using synapse-specific protein anchors like our SynBar system.

We next explored immunostaining approaches for SynBar detection. We first fluorophore-conjugated a commercially available antibody specific only to reconstituted GFP but not the large split GFP fragment. This custom-conjugated antibody effectively detected GFP+ synaptosomes by flow cytometry (Extended Data Fig. 3A), and the enriched population was confirmed as true singlet by confocal imaging (Extended Data Fig. 3B), but batch variation made it impractical for routine use. We therefore engineered V5 and HA epitope tags into the extracellular domains of PreSynBar and PostSynBar proteins (Fig. 2B), enabling detection with commercially available

anti-V5-AlexaFluor 647 and anti-HA-PE antibodies. A critical optimization revealed that traditional sucrose-based SET buffer severely reduced antibody staining efficiency; switching to a PBS-based protocol provided optimal balance between staining efficiency and synaptosome preservation (Extended Data Fig. 3C).

Our final optimized sorting protocol employed sequential gating based on stringent criteria (Fig. 2E, F): First, we established the noise floor threshold using a buffer-only control sample. Second, we identified singlet synaptosomes based on their characteristic vertical distribution in the scatter plot, consistent with previous literature reports[34,37]. Third, we selected membrane-intact particles by gating on the top ~90% of CellBrite 405 signal intensity to exclude dimly stained outliers. Finally, we identified double-positive (V5+HA+) synaptosomes using gates calibrated against wildtype control samples stained in parallel - specifically, we set thresholds that yielded less than 0.5% false positives in the V5 channel and less than 1% false positives in the HA channel. Confocal imaging of sorted positive populations verified their singlet morphology and fluorescent signals (Fig. 2G), with quantitative analysis showing ~90% HA signal retention in HA+ sorted populations and ~65% V5 signal retention in V5+ sorted populations (Extended Data Fig. 3D). Furthermore, analysis of double-positive structures revealed a higher proportion in V5+ versus HA+ sorted populations (Extended Data Fig. 3E), consistent with the relative frequencies observed in flow cytometry data (Fig. 2F). Using this approach, we could routinely isolate approximately 20,000 double-positive synaptosomes per hour. They were similarly sorted into a small amount of collection buffer in a pre-wet PCR strip tube and directly subjected to 10x Genomics workflow like sorted GFP+ nuclei.

3. Development of parallel single-nucleus and single-synaptosome sequencing

To capture both cellular identity and synaptic connectivity information, we developed parallel single-nucleus and single-synaptosome sequencing. This method combines single-nucleus RNA sequencing with single-synaptosome barcode sequencing to link neuronal transcriptomes with their synaptic connections through an optimized library construction protocol. 10x Genomics Single Cell 3' protocol with featuring barcoding was modified to enable simultaneous capture of transcriptomes and barcode RNAs. Within each loaded microfluidics-generated oil emulsion droplet, single nucleus or synaptosome was encapsulated with a single gel bead containing three types of reverse transcription oligos: poly(dT) for mRNAs, and two specific capture sequences for PreRNA and PostRNA, respectively (Fig. 3A). This design enables concurrent reverse transcription of cellular transcripts and barcode RNAs within the same droplet without potential competition.

For nuclear libraries, we developed a custom library construction protocol optimized for both transcriptome and barcode capture. Our workflow involves: Initial pre-amplification (PreAmp1) targeting both mRNA and barcodes; size-based separation of barcode and mRNA libraries using optimized sample-to-purification-beads ratios; and separate amplification of transcriptome and barcode components (Fig. 3B). Taqman qPCR quantification of PreRNA and PostRNA abundance at both PCR#2 and #3 verified that expected barcode types were enriched in each nuclei sample (Fig. 3C). This final protocol achieved high-quality nuclear transcriptomes and efficient barcode co-detection, with close to 99% of nuclei containing expected barcodes (Data shown in the next section).

For synaptosome libraries, our initial attempts to capture both mRNA and barcodes were hampered by the dominance of mitochondrial RNA, which reduced barcode detection efficiency while not offering additional insights into transcriptome heterogeneity among synaptosomes (Extended Data Fig. 4A-D). We therefore chose to omit mRNA capture from synaptosomes and developed a library construction protocol for synaptosome-specific barcode recovery with five key modifications to the standard 10x Genomics workflow (Fig. 3D): First, we omitted template-switching oligonucleotides (TSO) during reverse transcription, as we already knew the complete sequences of PreRNA and PostRNA. Second, we used barcode-specific primers with short extension times during pre-amplification (PCR#1) to selectively amplify short barcodes rather than long mRNAs. Third, we separated PreRNA and PostRNA into two reactions during second amplification (PCR#2) using specific primer pairs for each. Fourth, we optimized all bead purification steps for the smaller size of barcode molecules. Fifth, we noted that standard primers for barcode amplification tend to misalign and favor the high abundance target when the difference in abundance was large between PreRNA and PostRNA, resulting in under-amplification of the less abundant target. Hence, we utilized locked nuclei acid (LNA) modified primers, replacing a few regular adenosines with LNA-modified ones at specific positions within the primer annealing region, which achieved more balanced amplification of both PreRNA and PostRNA (Extended Data Fig. 5A-D). Similarly, we implemented Taqman qPCR quantification of PreRNA and PostRNA abundance at each step, which verified that this final protocol achieved high specificity and minimal cross-contamination between barcode types (Fig. 3E).

For a typical parallel single-nucleus and single-synaptosome sequencing experiment, we processed approximately 20,000 to 30,000 nuclei from each brain region and around 30,000 to 140,000 double-positive synaptosomes. Our optimized protocol achieved robust detection of both cellular transcriptomes as well as nuclear and synaptic barcodes, enabling comprehensive mapping of neural circuits with cellular gene expression profile and single cell identity.

4. Application of Connectome-seq to the pontocerebellar circuit

To demonstrate the capability of Connectome-seq in mapping neuronal connectivity in vivo, we applied our method to map the mouse pontocerebellar circuit. This circuit was chosen for its well-characterized anatomy and the presence of diverse cell types, providing an excellent testbed for our technology.

We injected AAV vectors encoding PreSynBar and PreRNA constructs bilaterally into the pons, and PostSynBar and PostRNA constructs unilaterally into the cerebellar cortex of adult mice (Fig. 4A). The injections were performed using titers and volumes to achieve widespread expression while minimizing potential overexpression artifacts. After a two-week expression period, we observed robust and specific expression of PreSynBar (detected by anti-V5 antibody) in pontine neurons (Fig. 4B) and their axonal projections to the cerebellum (Fig. 4C), as well as strong expression of PostSynBar (detected by anti-HA antibody) in cerebellar neurons (Fig. 4D). The specificity of PreSynBar localization was evidenced by the colocalization with Vglut1 and Vglut2 markers expressed in the presynaptic mossy fiber terminals projected from pons (Extended Data Fig. 6A). Barcoded synapses were detected via abundant specific split GFP signals at the interface between PreSynBar- and PostSynBar-expressing neurons (Fig. 4E). RNA in situ hybridization showed colocalization of PreRNA and PostRNA with SynBar proteins at mossy fiber terminals,

confirming efficient barcode transport and localization at synapses *in vivo* (Fig. 4F, Extended Data Fig. 6B-D). Notably, PreRNA showed a clustered distribution pattern consistent with its localization in presynaptic glomeruli structures, while PostRNA exhibited a more dispersed pattern throughout postsynaptic neurons with partial enrichment at PostSynBar-positive synapses.

To capture the full diversity of cell types and connectivity in both regions while minimizing individual differences, we performed Connectome-seq on six biological replicates. For each replicate, we isolated and sequenced all GFP-positive singlet nuclei from both infected regions in pons and cerebellum, and tens of thousands synaptosomes from the same cerebellar region. For each sequencing experiment, we aimed to sort and sequence all GFP-positive nuclei from the dissected brain region to ensure capturing the nuclear barcode pool as completely as possible. For synaptosomes, we stained approximately one-third of the total synaptosomes isolated from each cerebellum. Under our sorting conditions, it took about 1.5 hours to collect sufficient double-positive synaptosomes (~30,000) for 10X sequencing, which typically consumed one-third of the stained sample - effectively sampling about 10% of the total synaptosome preparation. While our flow cytometry does not intentionally select synaptosomes based on size or fluorescence intensity, the overall approach may still be subject to technical biases, including differential AAV transduction efficiency across cell types, varying nuclear RNA recovery rates that depend on neuronal size, and size- or shape-dependent differences in synaptosome purification yields. These potential sources of bias should be considered when interpreting the observed distribution of synapse types.

Single-nucleus RNA sequencing of the sorted GFP-positive nuclei revealed a diverse array of cell types in both the pons and cerebellum, with their proportions closely matching the expected distributions based on previous studies. In total, we recovered 180,468 and 129,187 nuclei profiles from pons and cerebellum with a median transcript capture of 1,852 and 1,346 unique molecular identifiers per profile respectively across all six individual brain samples. Using established cell type markers from published dataset, we performed comprehensive cell typing in each brain region using gene expression data from snRNA-seq.

Through unbiased clustering analysis (Extended Data Fig. 7A, B) followed by quality control to remove low-quality clusters (Extended Data Fig. 7C, D), we identified several distinct neuronal and non-neuronal populations in the pons and cerebellum. Initial cell type annotation was performed using canonical marker genes (Extended Data Fig. 7E, F) and further refined using the Allen Institute's MapMyCell annotation tool (Extended Data Fig. 7G, H). Following quality control, we identified 110,923 and 75,796 high-quality nuclei profiles from pons and cerebellum (Fig. 5G, H), with cell type identities validated using marker gene expression across merged clusters (Extended Data Fig. 7I, J). In the pons, neuronal populations comprised approximately 82% of total cells, with glutamatergic neurons being the most abundant (75,375 cells, including 35,818 Fat2 Glut and 39,557 Hoxb5 Glut populations), followed by GABAergic neurons (8,932 cells, including 2,034 Otp Pax3 Gaba, 6,154 Gata3 Gly-Gaba, and 744 Onecut1 Gaba populations) and serotonergic neurons (2,435 cells). These populations were classified based on their molecular markers, including *Slc17a6* (Vglut2) and *Slc17a7* (Vglut1) for glutamatergic neurons, *Gad1*, *Gad2*, and *Slc32a1* (Vgat) for GABAergic neurons, and *Slc6a4* and *Tph2* for serotonergic neurons. The non-neuronal cells included oligodendrocytes (ODC, 12,800 cells), astrocytes (6,384 cells), oligodendrocyte precursor cells (OPC, 1,562 cells), and smaller populations of endothelial cells

(2,439 cells), microglia (536 cells), and macrophages (460 cells). In the cerebellum, neurons constituted approximately 92% of all cells, with granule cells being the predominant population (49,650 cells, ~65% of total cells). The remaining neuronal populations included molecular layer interneurons (MLI1: 7,230 cells; MLI2: 1,586 cells), Purkinje cells (7,462 cells), Purkinje layer interneurons (PLI, 1,925 cells), Golgi cells (1,454 cells), and unipolar brush cells (UBC, 284 cells). Among non-neuronal cells, we identified Bergmann glia (1,992 cells), oligodendrocytes (906 cells), astrocytes (1,034 cells), ependymal cells (101 cells), choroid plexus cells (1,549 cells), and smaller populations of endothelial cells (435 cells), microglia (489 cells), and macrophages (444 cells). Cell type-specific marker genes were selected based on previously published cerebellar atlas data[38]. While our identified cell types are consistent with those reported in other single-cell databases including the Allen Brain Atlas[39], the relative abundance shows a bias towards neurons, particularly glutamatergic neurons, which is a known limitation of AAV1/2-based cell targeting approaches. Notably, the substantial reduction in cerebellar nuclei counts following quality control (from 129,187 to 75,796) likely reflects the selective loss of granule cells, which are particularly susceptible to exclusion during quality filtering due to their small size and typically lower RNA content.

Importantly, we detected barcodes in a high percentage of the sequenced nuclei, with 98.9% of pontine nuclei containing PreRNA and 98.9% of cerebellar nuclei containing PostRNA (Fig. 4I, J). This high efficiency of mRNA and barcode co-detection was crucial for comprehensive connectivity mapping, as they serve as the source pool of all synaptic barcodes to match with. Overall, these results confirmed that our modified library construction strategy to allow both mRNA and barcode capture generated sufficient data to accurately determine barcoded cell types, and that our viral strategy recapitulated the natural distribution of different cell types in each targeted region as reported by other unbiased single cell sequencing experiments.

Analysis of the synaptosomal sequencing data revealed the presence of both PreRNA and PostRNA, with approximately 23.6% (142,728) of synaptosomes containing both barcodes (Extended Data Fig. 8A). While this co-occurrence rate suggests successful capture of synaptic connections, the majority of synaptosomes contained either PreRNA (17.9%, 108,377) or PostRNA (58.5%, 354,196) alone. The lower abundance of PreRNA compared to PostRNA likely reflects the complex biological transport process required for PreRNA: these barcodes must travel long distances from the presynaptic neuron cell body through the axon before reaching the synaptic region, during which they may dissociate from their transporter proteins. In contrast, PostRNA has a relatively shorter transport distance within the postsynaptic region. This biological constraint is further confounded by the technical limitations of the 10x Genomics single cell sequencing workflow including high dropout rates and variation in barcode detection efficiency (Capture Sequence 2 is slightly more efficient than Capture Sequence 1), which ultimately contributed to the observed differential detection rates.

Further analysis revealed distinct patterns in N30 diversity and UMI distributions. The number of unique N30 sequences scaled linearly with total UMI counts (Extended Data Fig. 8B), with both PreRNA and PostRNA showing similar scaling relationships despite different absolute numbers. The N30 diversity distribution (Extended Data Fig. 8C) showed a characteristic right-skewed pattern, indicating that most synaptosomes contained a modest number of unique N30 sequences, with a long tail of higher-diversity cases. Read count analysis demonstrated that most UMIs were

associated with relatively few reads (Extended Data Fig. 8D), with both PreRNA and PostRNA showing similar distributions centered around 1-10 reads per UMI. The comparison of Pre/Post UMIs per synaptosome (Extended Data Fig. 8E) revealed a complex relationship between PreRNA and PostRNA abundance, with most synaptosomes showing asymmetric distributions of the two barcode types. Detailed read count distributions for both PreRNA (Extended Data Fig. 8F) and PostRNA (Extended Data Fig. 8G) followed similar patterns but with different scales, reflecting the overall higher abundance of PostRNA in the dataset.

5. Reconstruction of the pontocerebellar connectome

With the efficient barcoding of nuclei and synaptosomes in the pontocerebellar circuit using Connectome-seq constructs, we next focused on reconstructing the connectome from our sequencing data. We developed a computational pipeline to extract, error-correct, and match barcodes from both nuclear and synaptosomal sequencing data (Extended Data Fig. 9A). The pipeline implements multi-level error correction including CB-UMI-based filtering and network-based 30-mer correction, followed by stringent barcode matching criteria. We first collapsed barcodes from each dataset by allowing up to 1 mismatch among all detected unique N30 sequences. Next, we performed inter-sample matching by allowing up to 2 mismatches between the barcodes from nuclei and synaptosomes to account for PCR and sequencing introduced errors, balancing coverage and accuracy based on a thorough analysis across different allowed mismatch numbers (Extended Data Fig. 9B, C).

Our analysis revealed high barcode diversity in the synaptosomal fraction, with 81,452 unique PreRNA and 147,164 unique PostRNA sequences detected from ~600k barcode-containing synaptosomes (Extended Data Fig. 9B). Of these, 23,788 (29.2%) PreRNA were uniquely matched to their source in the presynaptic nuclear pool, while 27,886 (19.2%) PostRNA were uniquely matched to the postsynaptic nuclear source at Hamming distance 2. Despite technical factors affecting barcode recovery and matching rates, the matched barcodes provided sufficient coverage to reconstruct major circuit features, as demonstrated by the cell-type specific connectivity patterns described below.

When analyzing PreRNA and PostRNA matches separately, we found that 3,618 pons nuclei and 1,945 cerebellar nuclei were matched with corresponding synaptosomes (Fig. 5A, B). To assess the reliability of these matches, we first evaluated their consistency with known pontocerebellar circuit architecture (Fig. 5C). The canonical circuit consists of mossy fibers from pontine neurons that project through the middle cerebellar peduncle to target three main cell types in the cerebellar cortex: granule cells, Golgi cells, and unipolar brush cells (UBC)[40]. Recent studies have revealed additional connectivity patterns, such as mossy fiber inputs onto candelabrum cells, a type of Purkinje cell layer interneurons (PLI)[41]. Transient mossy fiber inputs to Purkinje cells during development were also long known to exist[42]. Given that our molecular approach has the potential to uncover novel connectivity patterns, where unexpected connections may represent genuine biological findings rather than technical artifacts, we focused our false discovery rate (FDR) analysis on the pontine side, where glutamatergic neurons are well-established as the classical projection neurons. When excluding all GABAergic neurons, the FDR was 9.8% (354/3,618). Using more stringent criteria where only glutamatergic neurons (Fat2 Glut and Hoxb5 Glut) are considered true positives (excluding both GABAergic and serotonergic neurons), the estimated

FDR increases to 12.4% (447/3,618). It is worth noting that serotonergic neurons in the pons, particularly from raphe nuclei, may also project to the cerebellum[43], though definitive molecular evidence for this connection remains to be established.

We next analyzed two key metrics of the connected populations: (1) the relative abundance of each cell type within the matched population (percentage of matched cells) and (2) the connection rate for each cell type (percentage of total cells of that type that were matched to synaptosomes. In the pons (Fig. 5D), among the 3,618 matched neurons, glutamatergic neurons were the predominant population, with Fat2 Glut (1,922 cells, 53.1% of matched; 5.37% match rate) and Hoxb5 Glut cells (1,249 cells, 34.5% of matched; 3.16% match rate) together accounting for 87.6% of all matched neurons. The remaining populations included Gata3 Gly-GABA (284 cells, 7.8% of matched; 4.61% match rate), Tph2 Glut-Sero (93 cells, 2.6% of matched; 3.82% match rate), Otp Pax3 Gaba (52 cells, 1.4% of matched; 2.56% match rate), and Onecut1 Gaba (18 cells, 0.5% of matched; 2.42% match rate). The high proportion and match rates of glutamatergic neurons align with their established role as the primary projection neurons in pontocerebellar circuits. In the cerebellum (Fig. 5E), the 1,945 matched neurons showed a diverse cellular composition reflecting the layered architecture of cerebellar cortex. Granule cells, despite their small size, formed the largest population (651 cells, 33.5% of matched; 1.31% match rate), followed by Purkinje cells (574 cells, 29.5% of matched; 7.69% match rate), MLI1 interneurons (340 cells, 17.5% of matched; 4.70% match rate), Golgi cells (194 cells, 10.0% of matched; 13.34% match rate), PLI (106 cells, 5.4% of matched; 5.51% match rate), MLI2 (76 cells, 3.9% of matched; 4.79% match rate), and UBC (5 cells, 0.3% of matched; 1.76% match rate). The varying match rates likely reflect a combination of biological and technical factors, with the notably low match rate for granule cells potentially attributed to their small soma size (~5-8 μm) compared to Purkinje cells (~20-40 μm) and Golgi cells (~20 μm), which would affect both viral infection efficiency and RNA capture in single-nucleus sequencing.

To reconstruct the full connectome, we analyzed synaptosomes containing both PreRNA and PostRNA that matched with our filtered nuclear populations, allowing us to map connections between individual pons and cerebellar neurons. The resulting clustered connectivity matrix (Fig. 5F) revealed distinct connectivity patterns at single cell level organized by cell types, where 281 pons neurons formed connections with 150 cerebellar neurons. Each red box in the matrix represents synaptosome read counts (Log2 scale) supporting specific neuron pairs, with a small subset of neuron pairs (~5%) supported by multiple independent synaptosomes, suggesting potentially stronger connections. The hierarchical clustering reveals modules of connectivity dominated by glutamatergic inputs and more distributed outputs. Quantification of these connections by cell type (Fig. 5G) shows that among the 338 total unique synaptosomes, pontine glutamatergic neurons (Fat2 and Hoxb5 Glut) formed the majority (184 and 96 unique synaptosomes respectively). The remaining populations - Gata3 Gly-GABA, Tph2 Glut-Sero, Otp Pax3 Gaba, and Onecut1 Gaba - showed more modest connectivity (22, 15, 8, and 3 unique synaptosomes respectively). On the cerebellar side, Golgi cells received the highest number of unique synaptosomal connections (108), followed by Purkinje cells (83), granule cells (63), MLI1 (43), PLI (28), and MLI2 (13). The substantial numbers of Glut-to-granule cell and Glut-to-Golgi cell connections align with known feedforward excitatory and inhibitory pontocerebellar circuitry, while the similar number of Glut-to-Purkinje cell connections represents an unexpected finding that warrants further investigation. The organized, cell-type specific patterns observed in both the

matrix and the quantified connections suggest these represent genuine circuit features rather than random noise, though some connections may still reflect technical artifacts given current understanding of pontocerebellar circuits.

6. Analysis and validation of a novel pons-to-Purkinje cell connectivity

Having established the technical framework of Connectome-seq and verified its ability to capture known connectivity patterns, we wanted to evaluate its potential in revealing new connections in well-established neural circuits. We decided to follow up on our intriguing discovery of the consistent connection between pontine neurons and Purkinje cells. The existence of such connectivity has been debated in the field, with previous studies only reporting its presence in the developing brain, not in adults[42,44].

To investigate the molecular characteristics of these connected neurons, we isolated all the Purkinje cells from our cerebellar snRNA-seq data and further grouped them into eleven subclusters (Fig. 6A). Intriguingly, Purkinje cells with pons connection identified by Connectome-seq (Cseq-positive) showed distinct clustering patterns, concentrating in subclusters 1, 6, and 10, suggesting potential molecular programs determining their connectivity (Fig. 6B). Quantification revealed highest enrichment of connected cells in cluster 10 (79.3%), followed by cluster 1 (23.1%), with decreasing proportions in other clusters (Extended Data Fig. 10A). To identify connectivity-determining marker genes, we performed differential gene expression analysis between Cseq-positive and Cseq-negative Purkinje cells. Our analysis revealed several highly enriched genes in connected neurons, including *Slc1a6*, *Pde1c*, *Cfap100*, *Gm11867*, *AY036118*, and *Sema3c* (Fig. 6C, Extended Data Fig. 10B-E). The enriched genes fell into distinct functional categories: cytoskeletal and structural organization (*Cfap100*), axon guidance (*Sema3c*), signaling regulation (*Pde1c*), and neurotransmitter regulation (*Slc1a6*). Interestingly, *Slc1a6*/EAAT4, a high-affinity glutamate transporter specifically expressed in Purkinje cells, is known to be differentially expressed across Purkinje cell populations following a parasagittal banding pattern that correlates with zebrin II expression[45]. Its higher expression in pons-connected cells suggests these neurons may require enhanced glutamate clearance capacity, which could be important for regulating their spontaneous firing patterns and long-term survival, as loss of EAAT4 leads to irregular Purkinje cell firing and eventual cell death[46]. Several other highly enriched markers (*Gm11867*, *AY036118*) represent poorly characterized genes that may reveal novel mechanisms of neural connectivity. The co-enrichment of established connectivity regulators like *Sema3c* alongside genes involved in diverse cellular processes suggests a complex molecular program integrating multiple pathways.

To validate this unexpected finding, we performed cell type-specific anterograde tracing with AAV1, which spreads trans-synaptically through a VAMP2-dependent synaptic vesicle release mechanism[47]. We developed a Cre-dependent AAV1-mScarlet virus (Fig. 6D), incorporating both a self-complementary AAV genome and a short CAG promoter to achieve high expression levels of mScarlet in the postsynaptic neurons. This design was based on previous work demonstrating that these two features enhanced protein expression from trans-synaptically trafficked viral particles[48]. We injected this Cre-dependent virus into the pons of adult Pcp2-Cre mice, which express Cre specifically in Purkinje cells (Fig. 6E). Remarkably, we observed clear mScarlet expression in a subset of Purkinje cells across the cerebellum, providing strong evidence for the

persistence of pons-to-Purkinje cell connections in the adult brain (Fig. 6F, G and Extended Data Fig. 10F).

To validate our findings of connectivity-marker genes, we performed RNA sequencing on these AAV1-labeled Purkinje cells, providing an independent dataset to verify the molecular signature of Cseq-positive Purkinje cells. We performed differential gene expression analysis comparing AAV1-labeled, Cseq-positive, and Cseq-negative Purkinje cells. This analysis demonstrated consistent enrichment patterns of our identified markers across these populations, with *Sema3c* showing the strongest enrichment (Fig. 6H). Based on this finding, we performed SEMA3C immunostaining on cerebellar sections from the same AAV1-DIO-mScarlet injected Pcp2-Cre mice used for connectivity tracing. This analysis confirmed co-expression of SEMA3C protein in the AAV1-labeled, pons-connected Purkinje cells, providing additional validation at the protein level (Fig. 6I). RNA in situ hybridization data from Allen Brain Atlas database also confirmed the presence of *Sema3c* RNA in a subset of Purkinje cells, further supporting our hypothesis that *Sema3c* is a predictor of pons connectivity in Purkinje cells (Extended Data Fig. 10G). Together, these results demonstrated the ability of Connectome-seq to reveal novel circuit features and relevant molecular markers while highlighting the importance of molecular context in circuit organization.

Discussion

Understanding neural circuit organization at single-cell resolution remains a fundamental challenge in neuroscience. In this study, we developed Connectome-seq, demonstrating its ability to map neuronal connectivity while simultaneously capturing molecular identities of connected neurons. Applied to the mouse pontocerebellar circuit, this approach validated both established connectivity patterns and suggested unexpected circuit features that warrant further investigation.

The observation of potential pons-to-Purkinje cell connectivity in our dataset raises intriguing questions about cerebellar circuit organization. Through molecular characterization of these connected neurons, we identified several markers enriched in pons-connected Purkinje cells, with *Sema3c* emerging as a particularly strong predictor of connectivity. This finding is especially intriguing given the known role of semaphorin signaling in cerebellar development, where *Sema3a* expressed by Purkinje cells acts as a repulsive cue preventing pontocerebellar mossy fibers from making inappropriate connections[49]. The presence of *Sema3c* in connected Purkinje cells suggests a potential molecular switch mechanism where different semaphorins could determine target specificity. This is reminiscent of the cardiac neural crest development system, where *Sema3a* provides repulsion while *Sema3c* acts as an attractant to guide precise migration[50]. However, further validation through complementary approaches, particularly electrophysiological recordings, will be essential to confirm these connections and understand their functional significance. The spatial intermingling of pontine axons targeting different cerebellar cell types suggests a more complex organization of the pontocerebellar circuit than previously appreciated, raising intriguing questions about the developmental mechanisms establishing these precise connectivity patterns.

Connectome-seq offers several advantages over existing approaches to neural circuit mapping. While serial electron microscopy reconstruction has provided foundational insights into circuit architecture at synaptic resolution, as demonstrated by recent studies in the cerebellar cortex[15], it

faces inherent limitations in scalability and throughput for mapping long-distance circuits across multiple brain regions. Light microscopy approaches such as array tomography and mGRASP allow synaptic resolution imaging but are limited by the need for thin sectioning and reconstruction of large volumes. Recent sequencing-based methods have improved scalability, though each faces distinct limitations: Sindbis virus-based methods enable high-throughput projection mapping but cannot reveal post-synaptic partners, while rabies virus-based methods allow synaptic connectivity mapping but struggle with barcode mixing in densely connected regions. Both viruses also face severe toxicity issues that may compromise circuit integrity. In contrast, Connectome-seq achieves comprehensive circuit mapping using minimally toxic AAV vectors while maintaining molecular precision, as demonstrated by our mapping result of the pontocerebellar system.

Through systematic optimization of molecular constructs, isolation protocols, and computational strategies, we achieved matching between synaptosomal barcodes and their nuclear sources at rates of 29.2% for PreRNA and 19.2% for PostRNA. These rates reflect several technical challenges in barcode recovery and matching. Source nuclei can be lost during tissue processing, particularly during dissection, homogenization, and flow cytometry enrichment where nuclear aggregates and compromised nuclei are excluded. Single-cell sequencing dropout effects may prevent capture of low-abundance barcodes in both nuclear and synaptosomal samples. Additionally, sequence mismatches are introduced during PCR amplification, especially given the higher number of PCR cycles required for synaptosomal samples. The impact of PCR-induced errors is evidenced by progressively increased matching rates when allowing more mismatches in the matching criteria, though this must be balanced against the introduction of false positives. The relatively low read counts per synaptosome also make the method susceptible to contamination from ambient RNA, potentially generating spurious connectivity calls. For the current study, we focused our analysis and validation efforts on high-confidence connections supported by multiple synaptosomal matches, particularly the consistent pons-to-Purkinje cell connectivity. Improving signal-to-noise ratios through better synaptosome isolation protocols and more stringent computational filtering will be crucial for future applications requiring high-confidence connectivity mapping. Our analysis identified several potential sources of technical artifacts: the detection of GABAergic neurons in pons projecting to cerebellum (9.8% of matched populations) likely represent either misassignment of barcodes to false source nuclei contaminated by RNA from true source nuclei, or false-positive matches from our barcode collapsing and matching algorithm. Using known non-projecting cells as negative controls, we estimated false discovery rates between 9.8%-12.4%.

To estimate the method's sensitivity, we calculated the theoretical synaptic populations accessible to our sequenced cells. Given our sampling of 12.5% of total glutamatergic pontine neurons (75,375 out of 603,082, according to published dataset[51]) and 0.25% of cerebellar granule cells (49,650 out of ~20 million from one third of cerebellum), combined with the known ~200 billion total mossy fiber (MF)→GrC synapses in the cerebellum[15], we estimate approximately 12.5% × 0.25% × 200 billion = 62.0 million theoretical MF→GrC synapses could be formed by our sequenced population. Similarly, with 1,454 cerebellar Golgi cells sampled and approximately 228 MF synapses per Golgi cell[52], we estimate 12.5% × 1,454 × 228 = 41,439 theoretical MF→Golgi synapses in our sampled population. Our detection of 51 MF→GrC and 89 MF→Golgi connections therefore represent drastically different sensitivities of 0.0001% and 0.215% respectively for these connection types. This difference in sensitivity likely reflects distinct cellular architectures - the larger soma size and different synaptic organization of Golgi cells may facilitate

both viral infection and RNA capture. While current sampling rates are extremely sparse due to technical limitations, several aspects of the method could be substantially improved. The 23.6% rate of dual-RNA capture from synaptosomes suggests room for optimization in synaptosomal isolation and RNA preservation protocols. The drop from 29.2% PreRNA and 19.2% PostRNA matching rates to 7.7% final validated connections (281 + 150 nuclei in the final connectivity matrix over 3618 + 1945 nuclei from the single-side-matched populations) indicates significant potential for improving barcode design, library construction, and matching algorithms. Additionally, cell-type specific differences in detection sensitivity point to opportunities for optimizing viral delivery and gene expression strategies. As these technical aspects get further improved, the detection of highly validated connections like MF→GrC and MF→Golgi circuits, along with potentially novel MF→Purkinje connectivity patterns and molecular markers such as *Sema3c* that correlate with specific connections, suggests Connectome-seq could become a powerful tool for high-throughput circuit mapping across diverse brain regions.

The development of Connectome-seq represents a notable advance in our ability to understand neural circuit organization across multiple scales. By enabling systematic mapping of long-distance circuits with molecular precision, this approach bridges a critical gap between local circuit mapping and systems-level connectivity. Beyond the pontocerebellar system studied here, future applications of Connectome-seq could reveal organizational principles of other long-range circuits, particularly when integrated with complementary approaches like spatial transcriptomics and electrophysiological validation. As the technical limitations are addressed and sensitivity improved, this molecular understanding of circuit-specific connectivity patterns will be essential for deciphering how neural networks process information across brain regions, how circuits are established during development, and how their architecture may be disrupted in neurological disorders.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Chicken anti-V5 tag antibody	Abcam	Cat# ab9113, RRID:AB_307022
Rabbit anti-HA tag antibody	Abcam	Cat# ab9110, RRID:AB_307019
Mouse Anti-Green Fluorescent Protein (GFP) Monoclonal Antibody, Unconjugated, Clone GFP-20	Sigma-Aldrich	Cat# G6539, RRID:AB_259941
V5 Tag Monoclonal Antibody, Alexa Fluor™ 647	Thermo Fisher Scientific	Cat# 451098, RRID:AB_2532221
HA-Tag (C29F4) Rabbit mAb (PE Conjugate)	Cell Signaling Technology	Cat# 14904, RRID:AB_2798643
SEMA3C Polyclonal Antibody	Thermo Fisher Scientific	Cat# PA590429, RRID:AB_2806070
Goat Anti-Mouse IgG1 Antibody, Alexa Fluor 488 Conjugated	Thermo Fisher Scientific	Cat# A-21121, RRID:AB_2535764
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Thermo Fisher Scientific	Cat# A-11011, RRID:AB_143157
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat# A-11008, RRID:AB_143165
Goat anti-Chicken IgY (H+L) Secondary Antibody, Alexa Fluor 647	Thermo Fisher Scientific	Cat# A-21449, RRID:AB_2535866
Bacterial and virus strains		
Endura DUO Electrocompetent Cells	Biosearch Technologies	Cat# 60242
NEB Stable Competent E. coli (High Efficiency)	New England Biolabs	Cat# C3040
AAV1/2 viruses described in Table S3	This paper	See Table S3
Critical commercial assays		
Chromium Next GEM Single Cell 3′ Kit v3.1, 16 rxns	10x Genomics	Cat# PN-1000268
Chromium Next GEM Chip G Single Cell Kit, 48 rxns	10x Genomics	Cat# PN-1000120
Dual Index Kit NT Set A, 96 rxns	10x Genomics	Cat# PN-1000242
Dual Index Kit TT Set A, 96 rxns	10x Genomics	Cat# PN-1000215

BaseScope Duplex Assay	ACD Bio	Cat# 323800
Deposited data		
Raw and analyzed sequencing data	This paper	GEO: GSE285978
Experimental models: Cell lines		
HEK 293T Cells	ATCC	Cat# CRL-3216, RRID:CVCL_0063
Experimental models: Organisms/strains		
Mouse: Wildtype: C57BL/6J	The Jackson Laboratory	JAX: 000664, RRID:IMSR_JAX:000664
Mouse: CAG-Sun1/sfGFP; B6;129-*Gt(ROSA)26Sor*$^{tm5(CAG\text{-}Sun1/sfGFP)Nat}$/J	The Jackson Laboratory	JAX: 021039, RRID:IMSR_JAX:021039
Mouse: BAC-Pcp2-IRES-Cre; B6.Cg-Tg(Pcp2-cre)3555Jdhu/J	The Jackson Laboratory	JAX: 010536, RRID:IMSR_JAX:010536
Oligonucleotides		
Oligos and primers described in Table S2	This paper	See Table S2
Recombinant DNA		
Plasmids described in Table S1	This paper	See Table S1
Software and algorithms		
ImageJ/Fiji	Schneider et al.[7]	https://fiji.sc/
Cellranger	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
Samtools	Li et al.[9]	http://samtools.sourceforge.net/
MapMyCells	Allen Institute	RRID:SCR_024672
RStudio markdown files for analyzing sequencing data	This paper	https://github.com/BoxuanZhaoLab/connectome-seq-code

Other		
BD FACSAria II Cell Sorter	BD	RRID:SCR_018934

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Boxuan Zhao (zbx@illinois.edu).

Materials Availability

All unique reagents generated in this study are available from the lead contact.

Data and Code Availability

The accession number for the single-nucleus and single-synaptosome sequencing data reported in this paper is GEO: GSE285978. Code supporting the current study is available at https://www.github.com/BoxuanZhaoLab/connectome-seq-code and are available from the corresponding author on request.

Experimental Model and Subject Details

Cell Lines

HEK 293T cells (ATCC CRL-3216) were maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco), penicillin-streptomycin (Corning, 5000 units/mL penicillin and 5000 µg/mL streptomycin) at 37°C, 5% CO2. This cell line has not been authenticated.

Animals

All animal procedures were approved by Stanford University's Institutional Animal Care and Use Committee (APLAC Protocol #: 14007) and University of Illinois Institutional Animal Care and Use Committee (IACUC Protocol #: 23169). For Connectome-seq experiments, 6-8 weeks old male/female Sun1-sfGFP-Myc (Jackson Laboratory, Stock #021039) transgenic mice were used. For imaging experiments, 6-8 weeks old male/female C57BL/6J (Jackson Laboratory, Stock #000664) mice were used. For AAV1 tracing experiments, 6-8 weeks old male/female Pcp2-Cre (Jackson Laboratory, Stock #010536) transgenic mice were used.

Mice were group-housed (2-5 per cage) on a 12h light/dark cycle with *ad libitum* access to food and water. Temperature was maintained at 20 ± 2°C and humidity at 50 ± 10%. Both male and female mice were used for experiments.

Detailed Materials and Methods

Molecular Biology and Viral Constructs

Plasmid Construction

Constructs were cloned into the pAAV viral vector (a gift from F. Zhang, MIT). Table S1 listed all plasmids used in this study. For all constructs, PCR fragments were amplified using Q5 High-

Fidelity 2X Master Mix (NEB), and vectors were double digested with NEB restriction enzymes. Both fragments were purified with DNA gel and assembled using Gibson assembly. Assembled products were introduced into competent NEB Stable bacteria via heat shock transformation and correct clones were verified with Sanger sequencing.

Multiplexed Plasmid Library Construction

Barcoded RNA libraries were generated by converting single-stranded oligonucleotides containing random N30 sequences into double-stranded DNA through primer extension. Following exonuclease treatment to remove residual ssDNA, the purified dsDNA inserts were cloned into the vector backbone using Golden Gate Assembly. Assembly conditions were optimized using various vector:insert molar ratios (1:5 to 1:20). Large-scale library construction was performed using 3 µg vector DNA with the optimal ratio of insert in 150 µL reactions containing T4 DNA ligase buffer and Golden Gate Assembly enzyme mix. The ligated products were transformed into Endura electrocompetent cells (Lucigen), recovered, and expanded at 30°C. Library complexity was assessed through colony counting and Sanger sequencing of individual clones.

HEK 293T Cell Culture and Transfection

HEK 293T cells (ATCC) were maintained in High Glucose DMEM (Thermo) supplemented with 10% FBS (Corning), 1% (v/v) penicillin-streptomycin (Corning) at 37°C with 5% CO2. For transfections, cells were plated in 6-well plates at 5×10^5 cells/well 24h before transfection. Cells at 70-80% confluency were transfected using polyethylenimine (PEI, 1 mg/mL in water, pH 7.3) at 3:1 PEI:DNA mass ratio. For each well, 2 µg total plasmid DNA was mixed with 6 µg PEI in 200 µL Opti-MEM (Thermo), incubated for 10 min at room temperature, then added dropwise to cells. Media was changed after 6 hours. For cis-assay, cells were co-transfected with 1:1 PreSynBar and PostSynBar plasmids. For trans-assay, cells were transfected with PreSynBar or PostSynBar separately, after 24 hours both populations of cells were lifted and replated into a separate pre-coated glass-bottom dish, with PostSynBar-transfected cells replated first and PreSynBar-transfected cells replated on top of them one hour later. The total density of cells was around 2/3 of the confluent amount of the new dish. Replated cells were incubated for another 24 hours followed by imaging.

HEK 293T Immunofluorescence Imaging

Cells to be imaged were plated on glass coverslips coated with 25 µg/mL poly-D-lysine in PBS for 1 h at 37 °C. After transfection and replating, cells were fixed with 4% (v/v) paraformaldehyde in PBS for 15 min, washed two times with PBS, and permeabilized with ice-cold methanol at 4 °C for 10 min or with 0.1% Triton X-100 in PBS at room temperature for 10 min. Cells were then washed two times with PBS and blocked for 1 hour with 2% BSA (w/v) in PBS at room temperature. Samples were incubated with primary antibodies diluted 1:1000 in 1% BSA in PBS at 4°C overnight. We used anti-V5 antibody to detect PreSynBar, anti-HA antibody for PostSynBar, and anti-GFP antibody for split GFP. After primary antibody incubation, cells were washed three times in PBS and then incubated with secondary antibodies conjugated to Alexa Fluor-488/568/647 and DAPI diluted 1:1000 in 1% BSA in PBS for 90 min at room temperature. Cells were washed three times with PBS and mounted on glass slides and imaged by confocal fluorescence microscopy (Zeiss). Images were collected with Slidebook (Intelligent Imaging Innovations).

Concentrated AAV Production

Concentrated AAV virus was prepared for *in vivo* use as described previously[53]. Three 15-cm plates of HEK 293T cells with fewer than fifteen passages were transfected at 80% confluence. For each 15-cm plate, we combined 5.2 μg plasmid containing transgenes, 4.35 μg of AAV1 capsid plasmid, 4.35 μg of AAV2 capsid plasmid, and 10.4 μg pDF6 adenovirus helper plasmid with 130 μL PEI in 500 μL of serum-free DMEM for 10 min at room temperature. The media in the plates was then removed by aspiration and replaced with 20 mL of complete growth media plus the DNA mixture. HEK 293T cells were incubated for 48 h at 37 °C, supernatant was discarded, and the cell pellet was pooled together in PBS and collected by centrifugation at 800*g* at room temperature for 10 min. The pellet was resuspended in 20 mL of TBS solution (100 mM NaCl, 20 mM Tris, pH = 8.0), freshly made sodium deoxycholate solution (10% in H_2O) was added to the resuspended cells to a final concentration of 0.5%, and benzonase nuclease (Sigma) was added to a final concentration of 50 U/mL. The mixture was shaken at 37 °C for 1 hour and cleared by centrifugation at 3,000*g* for 15 min. The supernatant was loaded using a peristaltic pump (Gilson MP4) at 1 mL/min flow rate onto a HiTrap heparin column (Cytiva) that was pre-equilibrated with 10 mL of TBS. The column was washed with 10 mL of TBS using the peristaltic pump, followed by washing with 1 mL of 200 mM NaCl, 20 mM Tris, pH = 8.0 and 1 mL of 300 mM NaCl, 20 mM Tris, pH = 8.0 using a 5 mL syringe. The virus was eluted using 5 mL syringes with 1.5 mL of 400 mM NaCl, 20 mM Tris, pH = 8.0; 3.0 mL of 450 mM NaCl, 20 mM Tris, pH = 8.0 and 1.5 mL of 500 mM NaCl, 20 mM Tris, pH = 8.0. The eluted virus was concentrated using Amicon Ultra 15 mL centrifugal units (100,000 MWCO, Millipore) at 2,000× g for 2 min, to a final volume of less than 500 μL. Then 1 mL sterile PBS was added to the filter unit and the column was centrifuged again for 1 min, followed by another wash with 500 μL PBS with 0.05% of PF-68 (Sigma) and centrifugation until the virus volume was ∼100 μL. The concentrated AAV virus was divided into 5 μL aliquots in PCS strips and flash frozen by liquid nitrogen followed by storage at −80 °C.

To titer the concentrated viruses, 2 μL of virus was incubated with 1 μL DNase I (NEB), 4 μL DNase I buffer, and 33 μL H_2O at 37°C for 30 min followed by 75°C for 15 min. 5 μL of this reaction was then added to 14 μL H_2O and 1 μL of Proteinase K (NEB) at 50°C for 30 min followed by 98°C for 10 min. qPCR reaction was prepared by adding 2 μL of this reaction to 5 μL of SYBR Green master mix (2x), 0.06 μL each of the forward and reverse primers (50 μM stock), and 2.88 μL of H_2O. Primers were designed against the synapsin promoter and WPRE listed in Table S2. Standardized curves were generated using a purified linearized AAV transgene plasmid that contained synapsin promoter and WPRE at 0.05 ng, 0.1 ng, or 0.2 ng per μL. The titer of each viral sample was calculated in reference to the standard curve as fully described in ref [53].

Primary Rat Cortical Neuron Culture and AAV Infection

Rat cortical neurons were isolated from E18 embryos following protocols approved by Stanford University's IACUC. Culture plates and/or glass coverslips were coated with 0.001% (w/v) poly-l-ornithine (Sigma-Aldrich) in DPBS (Gibco) at room temperature for 2 hours, washed twice with DPBS, and subsequently coated with 5 μg/mL mouse laminin (Gibco) in DPBS at 37 °C overnight. Cortical tissue was enzymatically digested in papain (Worthington) and DNase I (Roche) for 30 min at 37°C, then filtered through a 40 μm nylon cell strainer. Neurons were plated in 24-well plates at 1.25×10^6 cells/well on glass coverslips or directly on the plate

surface and maintained in complete neurobasal media (GIBCO) containing 2% B27 (Life Technologies), 1% Glutamax (Life Technologies), and 1% penicillin-streptomycin (VWR, 5 units/mL penicillin, 5 µg/mL streptomycin) at 37°C with 5% CO2. At DIV4, glial proliferation was suppressed by replacing 300 µL media with 500 µL complete neurobasal media supplemented with 10 µM FUDR (Sigma-Aldrich). Thereafter, 50% media changes with fresh complete neurobasal media were performed every 3 days. For viral transduction, neurons were infected at DIV7 with concentrated AAV at 8×10^9 viral genomes per well in 24-well plates. Media was changed 24 hours post-infection, and cells were maintained for at least 7 additional days to allow for robust transgene expression before experimental use.

Mouse stereotaxic surgeries

For stereotaxis surgeries, mice were anesthetized with 1.5%–2.0% isoflurane, and placed in a stereotaxic apparatus (Kopf Instruments) on a heating pad. The fur was removed from the scalp and a midline incision was made. We then drilled open a ~0.3 mm diameter craniotomy over the injection sites. Using a glass pipette (30–50 µm tip diameter), we injected 500 nL of AAV at 100 nL/min to each injection site. The following coordinates (in mm relative to lambda unless otherwise noted) were used for viral injections: Bilateral pons: +0.2 A/P, ±0.5 M/L, -5.55 D/V; Right cerebellum (right vermis of cerebellar lobule V): +0.2 mm A/P anterior to the post-lambda fissure on the cranial midline. -3.0 M/L, -0.5 mm D/V below dura. We allowed the virus to incubate for 2-3 weeks before experiments. The virus titers used for each experiment are listed in Table S3.

Mouse histology and immunohistochemistry

Following virus injection and expression, mice were heavily anesthetized with isoflurane and then perfused with 20 mL of cold phosphate-buffered saline (PBS) followed by 20 mL of cold 4% paraformaldehyde (PFA) in PBS. The brain was extracted from the skull and incubated in PFA at 4°C overnight, and then transferred to 30% sucrose for another incubation at 4°C overnight. The brain was sliced on a vibratome (Leica) or a cryotome (Leica) in 100-µm sections and mounted on glass slides with coverslips for imaging. For immunohistochemistry of PreSynBar, PostSynBar, or SEMA3C, slices were blocked and permeabilized in PBS with 0.3% Triton-X (PBST) and 5% normal goat serum (NGS) at room temperature for 2 hours. Slices were stained in primary antibodies in PBST+5% NGS at 4°C overnight, followed by three washes with PBST, and staining with secondary antibodies in PBST+5% NGS at room temperature for 90 min. Slices were washed three times in PBST and mounted on glass slides with mounting media containing DAPI for imaging. Imaging was performed using similar methods described for HEK 293T cells.

BaseScope In Situ Hybridization and Immunofluorescence

We combined the workflow of the standard BaseScope Duplex Detection kit (Advanced Cell Diagnostics) and the tech note workflow of RNAscope or BaseScope RED Assay combined with Immunohistochemistry - Integrated Co-Detection Workflow (ICW) to enable co-detection of both barcodes and SynBar proteins. Briefly, 10 µm Cryosections were fixed in 4% PFA for 1 hour at room temperature after brief heat treatment (100°C, 2 min). Primary antibodies were applied in co-detection diluent overnight at 4°C. Following PBST washes and post-fixation (4% PFA, 30 min), sections were treated with hydrogen peroxide for 5 min and BLOXALL Endogenous Blocking Solution (Vector Laboratories) for 10 min, followed by protease IV

treatment for 30 min at 40°C. Probes (Human PPIB-C1 control probes for PreRNA, DapB-C2 control probes for PostRNA) were hybridized for 2 hours at 40°C followed by sequential amplification steps using the BaseScope Duplex Detection kit. All 12 amplification steps and both Fast Red development (2 min, room temperature) and Green development (40°C, 5 min) were used. All washing steps between amplifications used BaseScope wash buffer at room temperature. Co-detection blocking was carried out (40°C, 15 min), followed by secondary antibodies incubation overnight at 4°C. Sections were mounted and imaged immediately after final PBST washes.

For RNA-protein colocalization analysis using FIJI (ImageJ): Binary masks were generated for both RNA (FISH) and protein (immunofluorescence) channels using automatic Triangle thresholding for protein signals and either 40% of maximum intensity (PreRNA) or automatic Triangle thresholding (PostRNA) for RNA signals. For proximity analysis, RNA masks were dilated by 1 pixel using the Maximum filter. Direct overlap was calculated as the intersection of RNA and protein masks. Proximity overlap was determined as the intersection between dilated RNA masks and protein masks. The percentage of RNA in protein was calculated as (overlap area / total RNA area) × 100. For each condition, three independent samples were analyzed. Statistical significance was assessed using two-tailed t-tests.

Parallel Nuclear and Synaptosomal Isolation

Fresh brain tissue was extracted and processed immediately after perfusing the mice with 10 mL ice-cold PBS. Target brain regions (right 1/3 of the cerebellum or bilateral pons) were dissected with new blades and homogenized in 1 mL SET buffer (320 mM sucrose, 5 mM Tris-HCl pH 7.5, 1 mM EDTA, 1:100 diluted Protease Inhibitor Cocktail (Sigma), 0.2 U/μL RNase inhibitor (Lucigen)) on ice using a glass dounce homogenizer (15 strokes) followed by additional homogenization with another 0.4 mL fresh SET buffer (5 strokes). The homogenate was combined and centrifuged at 1,500g for 8 min at 4°C to separate the nuclear pellet (P1) and synaptosomal supernatant (S1).

For nuclear isolation, the P1 fraction was briefly rinsed with nuclear lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40, 1:100 diluted Protease Inhibitor Cocktail (Sigma), 0.2 U/μL RNase inhibitor (Lucigen)) and then resuspended in 600 μL fresh lysis buffer. Cerebellar samples were further homogenized with 30 strokes using a plastic pestle, while pontine samples were gently mixed. Following lysis, samples were diluted with an equal volume of lysis buffer and incubated on ice for 10 min. Nuclear suspensions were filtered and pelleted at 1,500g for 5 min. The nuclear pellet was washed once with 800 μL nuclear lysis buffer, then resuspended in 800 μL washing and staining buffer (1X PBS, 1% BSA, 0.2 U/μL RNase inhibitor (Lucigen)). After 3 min incubation on ice, nuclei were pelleted (1,500g, 5 min), resuspended in 600 μL washing and staining buffer containing SiR-DNA at 1:200 final dilution ratio, filtered, and sent for flow cytometry.

For synaptosome isolation, the S1 fraction was centrifuged at 12,600g for 16 min at 4°C to obtain the crude synaptosomal pellet (P2). The P2 fraction was gently resuspended in 600 μL washing and staining buffer, 200 μL crude synaptosomes were taken out and blocked by adding 2 μL Fc receptor blocker (1:100) for 5 min on ice. Synaptosomes were then stained with anti-V5-Alexa Fluor 647 (1:50) and anti-HA-PE (1:50) antibodies for 1 hour at 4°C by rotation incubation. After washing with 400 μL washing and staining buffer (6,000g, 5 min), samples were resuspended in 500 μL washing and staining buffer. 200 μL of the suspension was diluted

in 800 µL washing and staining buffer containing 1:10,000 diluted CellBrite 405 membrane dye, filtered, and sent for flow cytometry.

Parallel Nuclear and Synaptosomal Flow Cytometry

Single nucleus sorting: Nuclei were sorted on a BD FACSAria II flow cytometer using SiR-DNA (1:200) for DNA content and Sun1-Tag GFP for infected neuron identification. Sequential gating was performed to first select singlet nuclei based on SiR-DNA intensity, followed by GFP-positive event selection. Sorted nuclei were collected directly into pre-wet PCR strip tubes containing 5 µL collection buffer (PBS with 1% BSA and 4U/µL Protector RNase Inhibitor (Roche)) for downstream processing.

Synaptosomes were sorted using a reconfigured BD FACSAria II flow cytometer. To optimize synaptosome detection, the standard optical configuration was modified: the bandpass filter was removed from the V450 detector to create a second side scatter channel (SSC), and the V525 detector was fitted with a 450/50 bandpass filter and a 410 longpass filter to enable violet fluorescence detection of CellBrite 405. Sequential gating was employed to enrich synaptosomes. First, events were gated based on dual side scatter measurements to identify events above the noise floor threshold, comparing with a buffer-only negative control sample. Second, singlet synaptosomes were selected based on their forward scatter (FSC) and side scatter (SSC) ratios, corresponding to the expected size and internal complexity of synaptosomes. Third, membrane integrity was assessed using CellBrite Steady 405 dye, with intact synaptosomes selected based on the top ~90% intensity of the CellBrite 405 signal. Finally, double-positive synaptosomes, representing connected pre- and post-synaptic terminals, were identified based on V5 and HA fluorescence signals. Thresholds for V5 (<0.5% false positive) and HA (<1% false positive) were empirically determined using wildtype control samples to minimize the inclusion of non-specific events. Approximately 20,000 double-positive synaptosomes were collected per hour into 5 µL of collection buffer for downstream processing. The BD FACSAria II, with its cuvette-based flow cell enabling longer laser exposure time and gel-coupled light collection, along with its continuous spectral detection range, was chosen for its superior ability to consistently detect and isolate singlet synaptosome populations compared to other platforms with segmented spectral detection or plastic-based flow cells.

Single-Nucleus and Single-Synaptosome Sequencing Libraries Construction and Sequencing

Libraries were generated using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1, Chromium Next GEM Chip G, and 10x Chromium Controller (10x Genomics) according to manufacturer's protocol with modifications. For each experiment, approximately 20,000-30,000 individually sorted GFP-positive nuclei or double-positive synaptosomes were loaded into each channel of the chip. For nuclear libraries, cDNA was amplified, size-selected using SPRIselect beads to separate transcriptome and barcode components, and independently constructed into mRNA libraries and barcode libraries. Synaptosome libraries were constructed using barcode-specific amplification without template switching oligonucleotides. The resulting libraries were quality-controlled using Bioanalyzer High Sensitivity DNA Analysis (Agilent), quantified by Qubit 4 Fluorometer (ThermoFisher), and pooled for sequencing on an Illumina NovaSeq X platform.

Computational Analysis

Raw Data Processing and Barcode Extraction

Raw sequencing data was processed using Cell Ranger (v7.1.0) with custom genome references including SynBar and Cre genes. Nuclear libraries were processed using force-cells parameter of actual loaded amount, while synaptosomal libraries used expect-cells parameter of actual loaded amount. Barcode sequences were extracted from filtered nuclear and unfiltered synaptosomal BAM files using custom Python scripts. Sequences were collapsed using Hamming distance 1 within samples to account for sequencing errors.

Single Nucleus RNA-seq Analysis

Initial data processing was performed for each sample independently. Raw data was loaded using Read10X, including both gene expression and synapse barcodes (SynBarpost, SynBarpre). Seurat objects were created with gene expression data as the default assay, while synapse barcodes were stored as separate assay. Quality control metrics were calculated including the percentage of mitochondrial genes (pattern "^mt-"), number of features, and total counts per cell. Cells were filtered to retain those with >100 and <8000 features and <5% mitochondrial reads. Data normalization was performed using LogNormalize with a scale factor of 10,000, followed by identification of 2,000 variable features using variance-stabilizing transformation (VST) method. After scaling all genes, principal component analysis was performed. The optimal number of dimensions (15) was determined using both JackStraw (100 replicates) and elbow plots.

For multi-sample analysis, the six independent samples were integrated using Harmony[54] to remove batch effects. For cerebellum data, after Harmony integration, clustering was performed using the standard Seurat pipeline[55], with gene expression data normalized using LogNormalize with scale factor 10,000, followed by identification of 2,000 variable features using variance-stabilizing transformation. After data scaling, principal component analysis was performed and the optimal number of dimensions (15) was determined using JackStraw and elbow plots. UMAP embedding was carried out at multiple resolutions (0.1-0.5) with 0.5 selected as optimal resolution based on cluster separation and biological relevance. Pons nuclei data were processed with a custom pipeline modified from a published single nucleus sequencing pipeline[56]. Briefly, variable genes were identified using a binomial model that compared expected versus observed expression frequencies across cells, selecting genes where the observed proportion was at least 5% less than the expected proportion based on bulk expression. Expression data was square root transformed, and a k-nearest neighbor graph was constructed using cosine distance on the first 30 harmony-corrected principal components. Clustering was performed using the Louvain algorithm across multiple resolution parameters (0.1-2.0), with the optimal resolution (1.0) selected based on cluster size criteria (minimum 500 cells) and cluster separation.

For both datasets, small clusters (<10 cells) were merged with their most transcriptionally similar neighbors based on correlation of average expression profiles. Clusters with median gene counts below 1,000 were filtered out (clusters 0 and 23 for pons; clusters 0 and 23 for cerebellum). Cell types were initially annotated using canonical marker genes, including neurotransmitter genes for pons described in ref[39] and published cell type markers for cerebellum described in ref[38]. Cluster identities were further annotated using Allen Institute's MapMyCell annotation tool, with each cluster assigned to the cell type showing the highest annotation counts. Initial subclusters were merged into major cell types based on shared marker expression patterns and consistent MapMyCell annotations.

Connectivity Analysis

Nuclear and synaptosomal barcode sequences from neurons were matched using a stepwise filtering approach. Initial pre-filtering removed ambiguous nuclear barcodes where the highest count was not at least 5 times (for pons) or 2 times (for cerebellum) greater than the second highest count, and synaptosomes associated with more than 100 unique barcode sequences. Matching between nuclear and synaptosomal barcodes was performed iteratively, allowing 0-3 hamming distance mismatches. Potential artifacts were removed through multiple filtering steps: nuclear and synaptosomal barcode reads with identical 10x Cell Barcodes were eliminated to avoid misassignment, barcodes associated with more than 100 synapses were removed to control for ambient RNA, and for cases with one synaptosome matching to multiple nuclei, only the pair with the highest synaptic count followed by highest nuclear count was retained. The filtered matches were used to construct cell-type grouped connectivity matrices based on shared barcodes between pre- and post-synaptic populations.

Differential Expression Analysis

Connectivity-associated marker gene expression was analyzed using Seurat's FindMarkers function to compare Cseq-positive and Cseq-negative populations. Analysis parameters required genes to be expressed in at least 25% of cells in either population with a minimum log fold-change threshold of 0.25. Statistical significance was determined by using adjusted p-values to account for multiple comparisons. Top differentially expressed genes were identified for each comparison and visualized using normalized expression data across conditions.

Statistical Analysis

Statistical analyses were performed with at least two biological replicates, with most key experiments using three or more replicates. Specific statistical tests included: Unpaired two-tailed Student's t-test for comparing expression levels between two conditions (e.g., RNA Polymerase II vs III-driven expression); One-way ANOVA followed by Dunnett's multiple comparisons test for comparing multiple conditions to a control (e.g., different RNA targeting sequences); Paired t-test for comparing high and low abundance samples in PCR analysis; Hierarchical clustering using Ward's method (ward.D2) for connectivity analysis; For differential gene expression analysis, statistical significance was assessed using adjusted p-values to control for multiple comparisons. Data are presented as mean ± SEM unless otherwise noted in figure legend. Statistical significance was typically indicated as: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; ns, not significant. No statistical methods were used to pre-determine sample size or fit with assumptions of statistical tests. No subjects with successful library construction were excluded from the study. Nuclei not passing quality control metrics defined in “Single Nucleus RNA-seq Analysis” were omitted from the bioinformatic analysis.

Acknowledgments

We are deeply grateful to Drs. Alice Y. Ting and Liqun Luo (Stanford University) for their foundational contributions to this work, including the initial conception of the method, guidance throughout the experimental development, and critical feedback on the manuscript. Their insights were instrumental in shaping both the technical approach and the scientific rigor of this study. We thank Dr. Steve R. Quake (Stanford University) for providing guidance to the project and critical sequencing resources. We thank Dr. Etienne Herzog (The University of Bordeaux) and Dr. Meredith Weglarz (Stanford University - Shared FACS Facility) for assistance with protocol optimization for single synaptosome sorting experiments. Nuclei and synaptosome sorting/flow cytometry analysis for this project was done on an instrument in the Stanford Shared FACS Facility (RRID: SCR_017788) obtained using NIH S10 Shared Instrument Grant S10RR025518-01. B.S.Z. was supported by the Life Sciences Research Foundation Shurl and Kay Curci Foundation Postdoctoral Fellowship and Wu Tsai Neurosciences Institute Interdisciplinary Scholar Award. This research was conducted with support from Stanford ChEM-H Postdocs at the Interface Seed Grant (B.S.Z. and A.I.) and the Neuro-omics Initiative grant from Wu Tsai Neurosciences Institute of Stanford University.

Contributions

B.S.Z. conceived and designed the study, performed all experiments with D.C. and A.I., analyzed the data with A.I., and wrote the manuscript. B.S.Z. and A.I. designed the sequencing workflow, performed sequencing experiments, and built the bioinformatic pipeline with help from Y.W. D.C. performed validation experiments on cell type specific connectivity. A.I. and W.J.Z. helped with data analysis and visualization. M.J.W. helped design AAV injection scheme and perform stereotaxic injections. Y.W. helped with Basescope RNA imaging experiments. All authors reviewed the manuscript and provided critical feedback.

References

1. White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1997). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B Biol. Sci. *314*, 1–340. https://doi.org/10.1098/rstb.1986.0056.

2. Dorkenwald, S., Matsliah, A., Sterling, A.R., Schlegel, P., Yu, S., McKellar, C.E., Lin, A., Costa, M., Eichler, K., Yin, Y., et al. (2024). Neuronal wiring diagram of an adult brain. Nature *634*, 124–138. https://doi.org/10.1038/s41586-024-07558-y.

3. Scheffer, L.K., Xu, C.S., Januszewski, M., Lu, Z., Takemura, S., Hayworth, K.J., Huang, G.B., Shinomiya, K., Maitlin-Shepard, J., Berg, S., et al. (2020). A connectome and analysis of the adult Drosophila central brain. eLife *9*, e57443. https://doi.org/10.7554/eLife.57443.

4. Schlegel, P., Yin, Y., Bates, A.S., Dorkenwald, S., Eichler, K., Brooks, P., Han, D.S., Gkantia, M., dos Santos, M., Munnelly, E.J., et al. (2024). Whole-brain annotation and multi-connectome cell typing of Drosophila. Nature *634*, 139–152. https://doi.org/10.1038/s41586-024-07686-5.

5. Zheng, Z., Lauritzen, J.S., Perlman, E., Robinson, C.G., Nichols, M., Milkie, D., Torrens, O., Price, J., Fisher, C.B., Sharifi, N., et al. (2018). A Complete Electron Microscopy Volume of

the Brain of Adult Drosophila melanogaster. Cell *174*, 730-743.e22. https://doi.org/10.1016/j.cell.2018.06.019.

6. Briggman, K.L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. Nature *471*, 183–188. https://doi.org/10.1038/nature09818.

7. Euler, T., Detwiler, P.B., and Denk, W. (2002). Directionally selective calcium signals in dendrites of starburst amacrine cells. Nature *418*, 845–852. https://doi.org/10.1038/nature00931.

8. Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature *500*, 168–174. https://doi.org/10.1038/nature12346.

9. Schmidt, H., Gour, A., Straehle, J., Boergens, K.M., Brecht, M., and Helmstaedter, M. (2017). Axonal synapse sorting in medial entorhinal cortex. Nature *549*, 469–475. https://doi.org/10.1038/nature24005.

10. Motta, A., Berning, M., Boergens, K.M., Staffler, B., Beining, M., Loomba, S., Hennig, P., Wissler, H., and Helmstaedter, M. (2019). Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science *366*, eaay3134. https://doi.org/10.1126/science.aay3134.

11. Gour, A., Boergens, K.M., Heike, N., Hua, Y., Laserstein, P., Song, K., and Helmstaedter, M. (2021). Postnatal connectomic development of inhibition in mouse barrel cortex. Science *371*, eabb4534. https://doi.org/10.1126/science.abb4534.

12. Turner, N.L., Macrina, T., Bae, J.A., Yang, R., Wilson, A.M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A.L., et al. (2022). Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. Cell *185*, 1082-1100.e24. https://doi.org/10.1016/j.cell.2022.01.023.

13. Ding, Z., Fahey, P.G., Papadopoulos, S., Wang, E.Y., Celii, B., Papadopoulos, C., Chang, A., Kunin, A.B., Tran, D., Fu, J., et al. (2023). Functional connectomics reveals general wiring rule in mouse visual cortex. Preprint, https://doi.org/10.1101/2023.03.13.531369 https://doi.org/10.1101/2023.03.13.531369.

14. Bodor, A.L., Schneider-Mizell, C.M., Zhang, C., Elabbady, L., Mallen, A., Bergeson, A., Brittain, D., Buchanan, J., Bumbarger, D.J., Dalley, R., et al. (2023). The Synaptic Architecture of Layer 5 Thick Tufted Excitatory Neurons in the Visual Cortex of Mice. Preprint, https://doi.org/10.1101/2023.10.18.562531 https://doi.org/10.1101/2023.10.18.562531.

15. Nguyen, T.M., Thomas, L.A., Rhoades, J.L., Ricchi, I., Yuan, X.C., Sheridan, A., Hildebrand, D.G.C., Funke, J., Regehr, W.G., and Lee, W.-C.A. (2023). Structured cerebellar connectivity supports resilient pattern separation. Nature *613*, 543–549. https://doi.org/10.1038/s41586-022-05471-w.

16. Cook, S.J., Jarrell, T.A., Brittin, C.A., Wang, Y., Bloniarz, A.E., Yakovlev, M.A., Nguyen, K.C.Q., Tang, L.T.-H., Bayer, E.A., Duerr, J.S., et al. (2019). Whole-animal connectomes of both Caenorhabditis elegans sexes. Nature *571*, 63–71. https://doi.org/10.1038/s41586-019-1352-7.

17. Witvliet, D., Mulcahy, B., Mitchell, J.K., Meirovitch, Y., Berger, D.R., Wu, Y., Liu, Y., Koh, W.X., Parvathala, R., Holmyard, D., et al. (2021). Connectomes across development reveal principles of brain maturation. Nature *596*, 257–261. https://doi.org/10.1038/s41586-021-03778-8.

18. Vu, A.T., Auerbach, E., Lenglet, C., Moeller, S., Sotiropoulos, S.N., Jbabdi, S., Andersson, J., Yacoub, E., and Ugurbil, K. (2015). High resolution whole brain diffusion imaging at 7 T for the Human Connectome Project. NeuroImage *122*, 318–331. https://doi.org/10.1016/j.neuroimage.2015.08.004.

19. Zhu, D., Zhang, T., Jiang, X., Hu, X., Chen, H., Yang, N., Lv, J., Han, J., Guo, L., and Liu, T. (2014). Fusing DTI and fMRI data: A survey of methods and applications. NeuroImage *102*, 184–191. https://doi.org/10.1016/j.neuroimage.2013.09.071.

20. Lo, L., and Anderson, D.J. (2011). A Cre-Dependent, Anterograde Transsynaptic Viral Tracer for Mapping Output Pathways of Genetically Marked Neurons. Neuron *72*, 938–950. https://doi.org/10.1016/j.neuron.2011.12.002.

21. Micheva, K.D., and Smith, S.J. (2007). Array Tomography: A New Tool for Imaging the Molecular Architecture and Ultrastructure of Neural Circuits. Neuron *55*, 824. https://doi.org/10.1016/j.neuron.2007.08.007.

22. Xu, X., Holmes, T.C., Luo, M.-H., Beier, K.T., Horwitz, G.D., Zhao, F., Zeng, W., Hui, M., Semler, B.L., and Sandri-Goldin, R.M. (2020). Viral Vectors for Neural Circuit Mapping and Recent Advances in Trans-synaptic Anterograde Tracers. Neuron *107*, 1029–1047. https://doi.org/10.1016/j.neuron.2020.07.010.

23. Kim, J., Zhao, T., Petralia, R.S., Yu, Y., Peng, H., Myers, E., and Magee, J.C. (2012). mGRASP enables mapping mammalian synaptic connectivity with light microscopy. Nat. Methods *9*, 96–102. https://doi.org/10.1038/nmeth.1784.

24. Tsetsenis, T., Boucard, A.A., Araç, D., Brunger, A.T., and Südhof, T.C. (2014). Direct Visualization of Trans-Synaptic Neurexin–Neuroligin Interactions during Synapse Formation. J. Neurosci. *34*, 15083. https://doi.org/10.1523/JNEUROSCI.0348-14.2014.

25. Kebschull, J.M., Garcia da Silva, P., Reid, A.P., Peikon, I.D., Albeanu, D.F., and Zador, A.M. (2016). High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. Neuron *91*, 975–987. https://doi.org/10.1016/j.neuron.2016.07.036.

26. Chen, X., Sun, Y.-C., Zhan, H., Kebschull, J.M., Fischer, S., Matho, K., Huang, Z.J., Gillis, J., and Zador, A.M. (2019). High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. Cell *179*, 772-786.e19. https://doi.org/10.1016/j.cell.2019.09.023.

27. Huang, L., Kebschull, J.M., Fürth, D., Musall, S., Kaufman, M.T., Churchland, A.K., and Zador, A.M. (2020). BRICseq Bridges Brain-wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals. Cell *182*, 177-188.e27. https://doi.org/10.1016/j.cell.2020.05.029.

28. Clark, I.C., Gutiérrez-Vázquez, C., Wheeler, M.A., Li, Z., Rothhammer, V., Linnerbauer, M., Sanmarco, L.M., Guo, L., Blain, M., Zandee, S.E.J., et al. (2021). Barcoded viral tracing of single-cell interactions in central nervous system inflammation. Science *372*, eabf1230. https://doi.org/10.1126/science.abf1230.

29. Saunders, A., Huang, K.W., Vondrak, C., Hughes, C., Smolyar, K., Sen, H., Philson, A.C., Nemesh, J., Wysoker, A., Kashin, S., et al. (2022). Ascertaining cells' synaptic connections and RNA expression simultaneously with barcoded rabies virus libraries. Nat. Commun. *13*, 6993. https://doi.org/10.1038/s41467-022-34334-1.

30. Zhang, A., Jin, L., Yao, S., Matsuyama, M., van Velthoven, C.T., Sullivan, H.A., Sun, N., Kellis, M., Tasic, B., Wickersham, I., et al. (2024). Rabies virus-based barcoded neuroanatomy resolved by single-cell RNA and in situ sequencing. eLife *12*, RP87866. https://doi.org/10.7554/eLife.87866.

31. Shin, D., Urbanek, M.E., Larson, H.H., Moussa, A.J., Lee, K.Y., Baker, D.L., Standen-Bloom, E., Ramachandran, S., Bogdanoff, D., Cadwell, C.R., et al. (2024). High-Complexity Barcoded Rabies Virus for Scalable Circuit Mapping Using Single-Cell and Single-Nucleus Sequencing. bioRxiv, 2024.10.01.616167. https://doi.org/10.1101/2024.10.01.616167.

32. Arora, A., Castro-Gutierrez, R., Moffatt, C., Eletto, D., Becker, R., Brown, M., Moor, A.E., Russ, H.A., and Taliaferro, J.M. (2022). High-throughput identification of RNA localization elements in neuronal cells. Nucleic Acids Res. *50*, 10626–10642. https://doi.org/10.1093/nar/gkac763.

33. Goering, R., Arora, A., Pockalny, M.C., and Taliaferro, J.M. (2023). RNA localization mechanisms transcend cell morphology. eLife *12*, e80040. https://doi.org/10.7554/eLife.80040.

34. Luquet, E., Biesemann, C., Munier, A., and Herzog, E. (2017). Purification of Synaptosome Populations Using Fluorescence-Activated Synaptosome Sorting. In Synapse Development: Methods and Protocols, A. Poulopoulos, ed. (Springer New York), pp. 121–134. https://doi.org/10.1007/978-1-4939-6688-2_10.

35. Whittaker, V.P. (1993). Thirty years of synaptosome research. J. Neurocytol. *22*, 735–742. https://doi.org/10.1007/BF01181319.

36. Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R., et al. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369–1384. https://doi.org/10.1016/j.neuron.2015.05.018.

37. Hafner, A.-S., Donlin-Asp, P.G., Leitch, B., Herzog, E., and Schuman, E.M. (2019). Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. Science *364*, eaau3644. https://doi.org/10.1126/science.aau3644.

38. Kozareva, V., Martin, C., Osorno, T., Rudolph, S., Guo, C., Vanderburg, C., Nadaf, N., Regev, A., Regehr, W.G., and Macosko, E. (2021). A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature *598*, 214–219. https://doi.org/10.1038/s41586-021-03220-z.

39. Yao, Z., Van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature *624*, 317–332. https://doi.org/10.1038/s41586-023-06812-z.

40. De Zeeuw, C.I., Lisberger, S.G., and Raymond, J.L. (2021). Diversity and dynamism in the cerebellum. Nat. Neurosci. *24*, 160–167. https://doi.org/10.1038/s41593-020-00754-9.

41. Osorno, T., Rudolph, S., Nguyen, T., Kozareva, V., Nadaf, N.M., Norton, A., Macosko, E.Z., Lee, W.-C.A., and Regehr, W.G. (2022). Candelabrum cells are ubiquitous cerebellar cortex interneurons with specialized circuit properties. Nat. Neurosci. *25*, 702–713. https://doi.org/10.1038/s41593-022-01057-x.

42. Takeda, T., and Maekawa, K. (1989). Transient direct connection of vestibular mossy fibers to the vestibulocerebellar Purkinje cells in early postnatal development of kittens. Neuroscience *32*, 99–111. https://doi.org/10.1016/0306-4522(89)90110-3.

43. Alonso, A., Merchán, P., Sandoval, J.E., Sánchez-Arrones, L., Garcia-Cazorla, A., Artuch, R., Ferrán, J.L., Martínez-de-la-Torre, M., and Puelles, L. (2013). Development of the serotonergic cells in murine raphe nuclei and their relations with rhombomeric domains. Brain Struct. Funct. *218*, 1229–1277. https://doi.org/10.1007/s00429-012-0456-8.

44. Mason, C., and Gregory, E. (1984). Postnatal maturation of cerebellar mossy and climbing fibers: transient expression of dual features on single axons. J. Neurosci. *4*, 1715. https://doi.org/10.1523/JNEUROSCI.04-07-01715.1984.

45. Dehnes, Y., Chaudhry, F.A., Ullensvang, K., Lehre, K.P., Storm-Mathisen, J., and Danbolt, N.C. (1998). The Glutamate Transporter EAAT4 in Rat Cerebellar Purkinje Cells: A Glutamate-Gated Chloride Channel Concentrated near the Synapse in Parts of the Dendritic Membrane Facing Astroglia. J. Neurosci. *18*, 3606. https://doi.org/10.1523/JNEUROSCI.18-10-03606.1998.

46. Perkins, E.M., Clarkson, Y.L., Suminaite, D., Lyndon, A.R., Tanaka, K., Rothstein, J.D., Skehel, P.A., Wyllie, D.J.A., and Jackson, M. (2018). Loss of cerebellar glutamate transporters EAAT4 and GLAST differentially affects the spontaneous firing pattern and survival of Purkinje cells. Hum. Mol. Genet. *27*, 2614–2627. https://doi.org/10.1093/hmg/ddy169.

47. Zingg, B., Peng, B., Huang, J., Tao, H.W., and Zhang, L.I. (2020). Synaptic Specificity and Application of Anterograde Transsynaptic AAV for Probing Neural Circuitry. J. Neurosci. *40*, 3250–3267. https://doi.org/10.1523/JNEUROSCI.2158-19.2020.

48. Faress, I., Khalil, V., Yamamoto, H., Sajgo, S., Yonehara, K., and Nabavi, S. (2023). Recombinase-independent AAV for anterograde transsynaptic tracing. Mol. Brain *16*, 66. https://doi.org/10.1186/s13041-023-01053-7.

49. Rabacchi, S.A., Solowska, J.M., Kruk, B., Luo, Y., Raper, J.A., and Baird, D.H. (1999). Collapsin-1/Semaphorin-III/D Is Regulated Developmentally in Purkinje Cells and Collapses Pontocerebellar Mossy Fiber Neuronal Growth Cones. J. Neurosci. *19*, 4437. https://doi.org/10.1523/JNEUROSCI.19-11-04437.1999.

50. Toyofuku, T., Yoshida, J., Sugimoto, T., Yamamoto, M., Makino, N., Takamatsu, H., Takegahara, N., Suto, F., Hori, M., Fujisawa, H., et al. (2008). Repulsive and attractive semaphorins cooperate to direct the navigation of cardiac neural crest cells. Dev. Biol. *321*, 251–262. https://doi.org/10.1016/j.ydbio.2008.06.028.

51. Erö, C., Gewaltig, M.-O., Keller, D., and Markram, H. (2018). A Cell Atlas for the Mouse Brain. Front. Neuroinformatics *12*.

52. Ito, M. (2009). Cerebellar Microcircuitry. In Encyclopedia of Neuroscience, L. R. Squire, ed. (Academic Press), pp. 723–728. https://doi.org/10.1016/B978-008045046-9.01306-1.

53. Wang, W., Wildes, C.P., Pattarabanjird, T., Sanchez, M.I., Glober, G.F., Matthews, G.A., Tye, K.M., and Ting, A.Y. (2017). A light- and calcium-gated transcription factor for imaging and manipulating activated neurons. Nat. Biotechnol. *35*, 864–871. https://doi.org/10.1038/nbt.3909.

54. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

55. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.

56. Langlieb, J., Sachdev, N.S., Balderrama, K.S., Nadaf, N.M., Raj, M., Murray, E., Webber, J.T., Vanderburg, C., Gazestani, V., Tward, D., et al. (2023). The molecular cytoarchitecture of the adult mouse brain. Nature *624*, 333–342. https://doi.org/10.1038/s41586-023-06818-7.

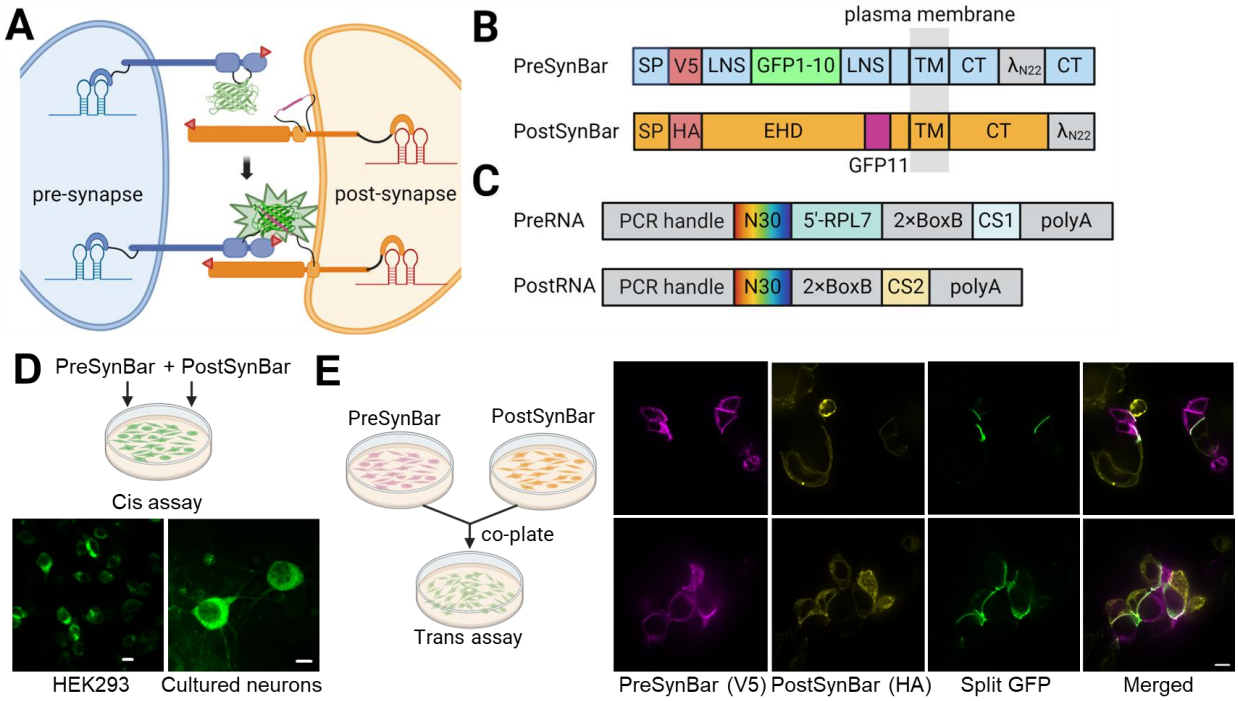


Figure 1: Design and validation of SynBar constructs

(A) Schematic illustration of the SynBar system at synapses. PreSynBar and PostSynBar proteins interact to form split-GFP and bring RNA barcodes together.

(B) Schematic representation of PreSynBar and PostSynBar proteins, showing the neurexin/neuroligin base, split-GFP components, and RNA-binding domains. GFP1–10 and GFP11 denote the split-GFP moieties containing β-strands 1–10 and 11, respectively. SP, Signal peptide; EHD, Esterase homology domain; LNS, laminin/neurexin/sex hormone binding domain; TM, transmembrane domain; CT, cytoplasmic tail.

(C) Design of PreRNA and PostRNA RNA. Key functional elements include PCR handles, N30 random region for unique cell of origin identification, BoxB sites for λN22 binding, and specific capture sequences (CS1/2) for 10X library construction. RPL7 5'UTR in PreRNA enhances distal transport.

(D) Cis validation of SynBar interaction in HEK 293 cells and rat cortical neurons showing reconstituted split GFP signals (green) inside cells expressing both PreSynBar and PostSynBar (cis-interaction). Scale bars: 10 µm.

(E) Trans validation showing specific GFP reconstitution (green) at contact surface between HEK 293 cells expressing complementary PreSynBar (magenta, via V5 staining) and PostSynBar (yellow, via HA staining) constructs. Two sets of representative images shown. Scale bar: 10 µm.

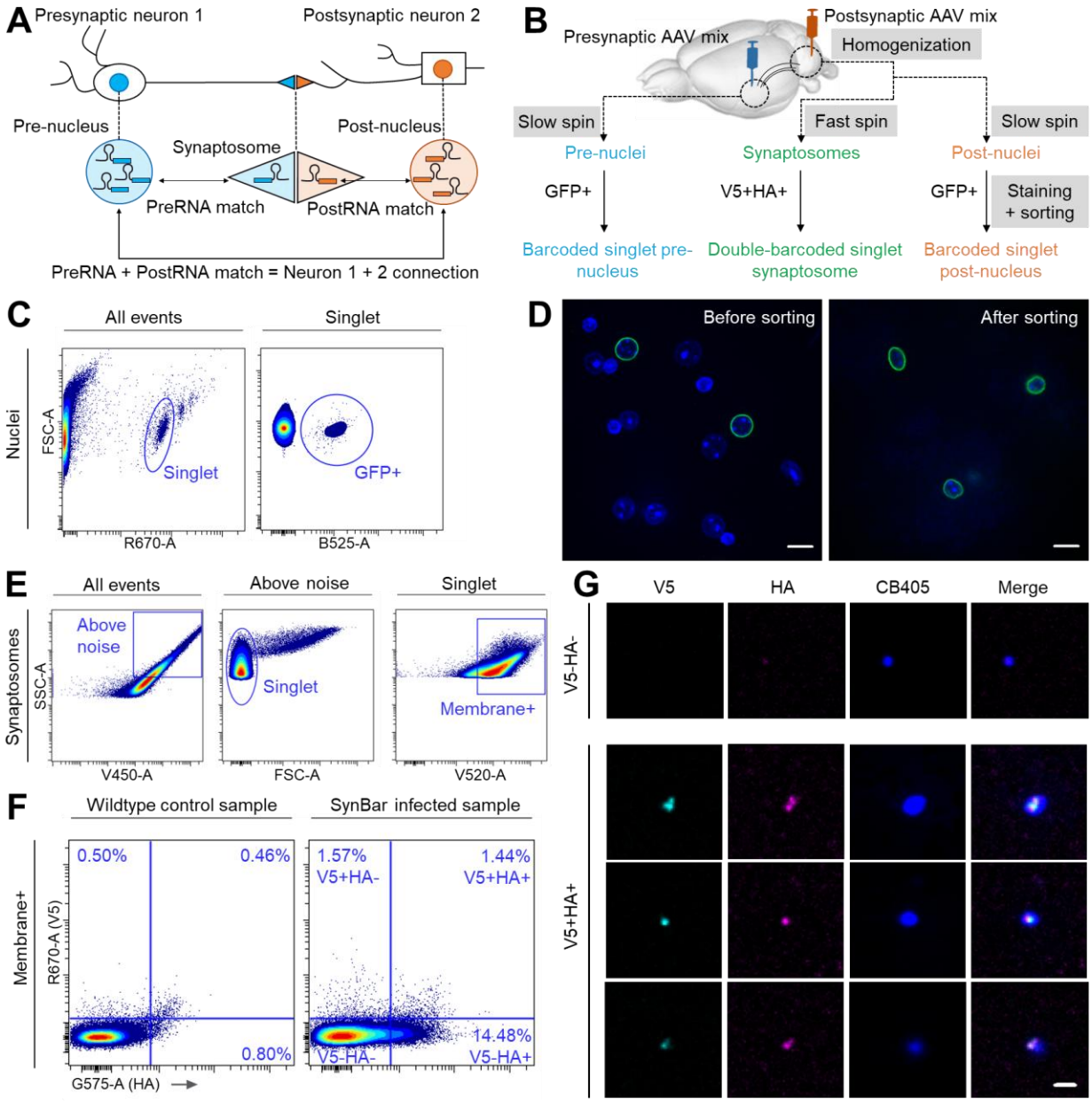


Figure 2: Nuclear and Synaptosome Isolation Strategy

(A) Schematic of the barcode matching strategy. Pre-nucleus contains PreRNA and post-nucleus contains PostRNA from their respective neurons. When both PreRNA and PostRNA sequences from nuclei match those found in a single synaptosome, this indicates a synaptic connection between the two neurons.

(B) Workflow for parallel isolation of nuclei and synaptosomes. Following AAV injection into pre- and postsynaptic regions, brain tissue is processed through differential centrifugation. Slow spins yield crude pre- and post-nuclei, while a fast spin of the supernatant of the slow spin produces

the synaptosome fraction. Nuclei and synaptosomes are enriched by FACS based on GFP signal (nuclei) or V5+HA antibody staining (synaptosomes).

(C) Nuclear isolation strategy. Left: Initial FACS plot showing all events, with gating on SiR-DNA signal identifying singlet nuclei. Right: Gating on singlet nuclei to select GFP+ population.

(D) Validation of nuclear sorting. Confocal images showing pre-sort mixture of GFP+ and GFP- nuclei (left) and post-sort enrichment of GFP+ singlet nuclei (right). Scale bars: 10 μm.

(E) Sequential gating strategy for synaptosome isolation. Left to right: Selection of events above noise threshold, selection of singlet population based on scattering properties, and identification of membrane-intact singlet synaptosomes based on Cellbrite 405 signal.

(F) Synaptosome enrichment by antibody staining. FACS plots comparing wildtype control (left) versus SynBar-infected samples (right), showing enriched V5+/HA+ double-positive population in infected samples.

(G) Imaging validation of sorted synaptosomes. Confocal images of sorted populations showing V5-HA- double-negative and V5+HA+ double-positive synaptosomes. Blue is from universal membrane stain with Cellbrite 405 (CB405). Cyan is PreSynBar signal from staining with anti-V5 antibody. Magenta is PostSynBar signal from staining with anti-HA antibody. Scale bars: 2 μm.

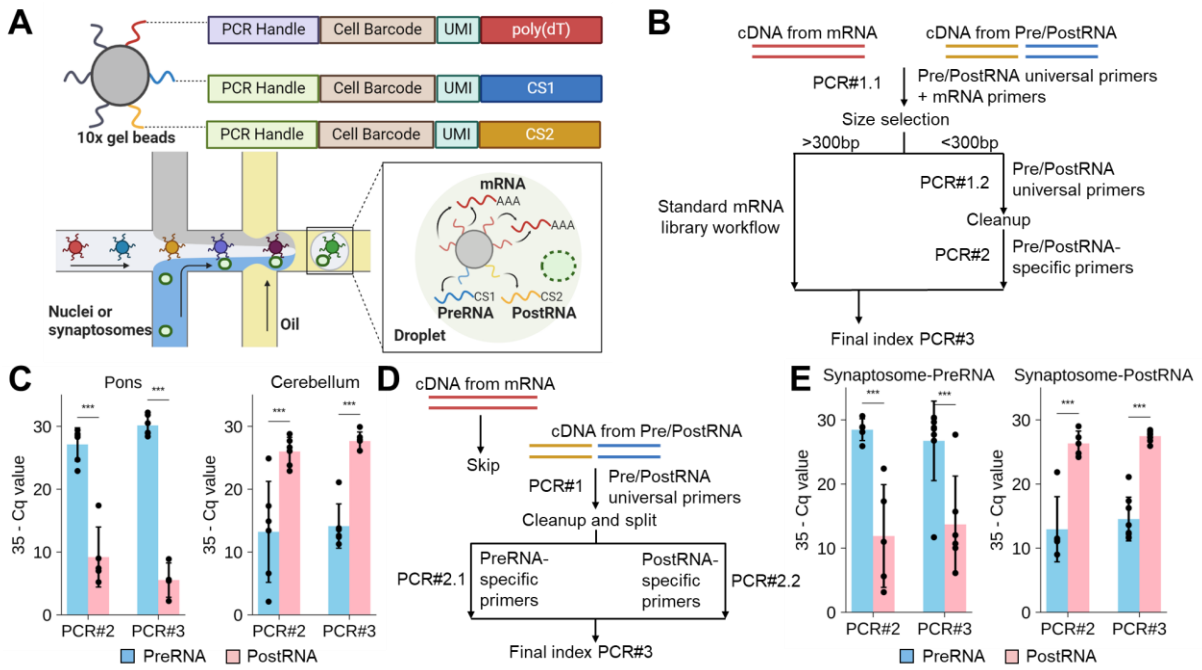


Figure 3. Development of parallel single-nucleus and single-synaptosome sequencing

(A) Adapting 10x Genomics gel bead design for simultaneous capture of transcriptomes and barcodes. Top: Structure of three capture oligonucleotides on gel beads - poly(dT) for mRNA (red), Capture Sequence 1 (CS1) for PreRNA (blue), and Capture Sequence 2 (CS2) for PostRNA (orange). Bottom: Schematic of microfluidic co-encapsulation showing nuclei/synaptosomes with gel beads in oil emulsion droplets. Inset shows molecular components within a droplet.

(B) Nuclear library construction workflow. The protocol separates mRNA-derived cDNA (red) from Pre/PostRNA-derived cDNA (blue/orange) through size selection after initial PCR#1.1. Large fragments (>300bp) undergo standard mRNA library preparation, while small fragments (<300bp) are processed through PreRNA- or PostRNA-specific amplification steps. PCR#2 uses barcode-specific primers, followed by final indexing PCR#3.

(C) qPCR analysis showing selective amplification of PreRNA and PostRNA barcodes from nuclei samples from pons and cerebellum. Bar plots show Taqman qPCR values (35-Cq) for PreRNA (blue) and PostRNA (pink) after PCR#2 and PCR#3. Higher values indicate greater abundance. PreRNA barcodes showed consistently higher abundance in pons samples whereas PostRNA barcodes showed consistently higher abundance in cerebellum samples, demonstrating specificity of the amplification strategy. Error bars represent standard deviation; individual data points shown as black dots. *** $p < 0.001$, paired t-test.

(D) Synaptosome library construction workflow optimized for barcode recovery. mRNA capture is omitted, and both PreRNA and PostRNA cDNA undergo direct amplification with universal primers (PCR#1). The protocol then splits into separate reactions for PreRNA- and PostRNA-specific amplification (PCR#2.1 and PCR#2.2) before final indexing.

(E) qPCR validation of Pre- and PostRNA barcode enrichment in synaptosome samples. Bar plots show Taqman qPCR values (35-Cq) comparing PreRNA-enriched (blue) and PostRNA-enriched (pink) samples after PCR#2 and PCR#3. Synaptosome-PreRNA samples show specific enrichment of PreRNA barcodes, while Synaptosome-PostRNA samples show specific enrichment of PostRNA barcodes, confirming successful library construction of each type of barcode. Error bars represent standard deviation; individual replicates shown as black dots. *** $p < 0.001$, paired t-test.

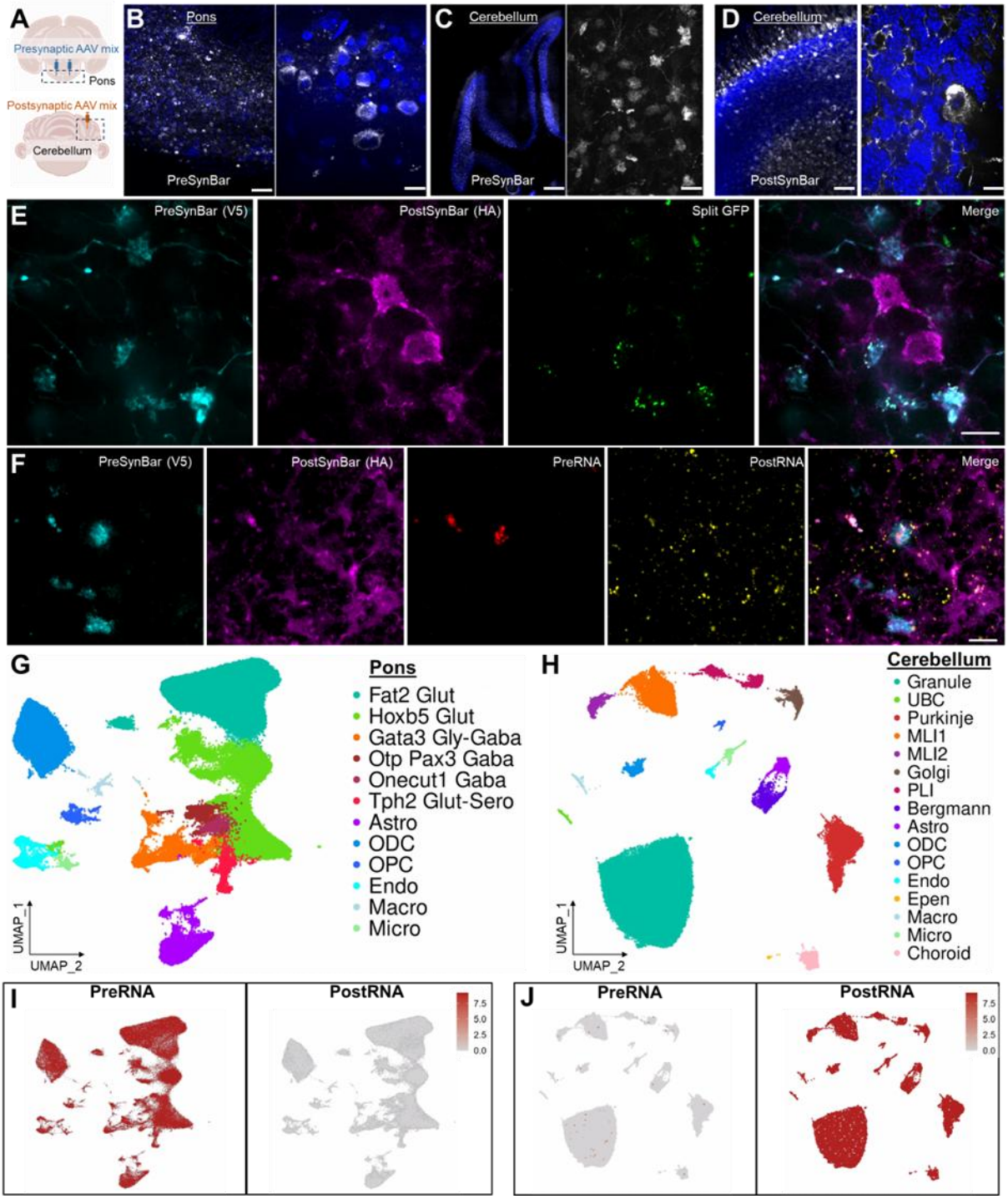


Figure 4. *In vivo* application and validation of Connectome-seq in the mouse pontocerebellar circuit

(A) Schematic showing bilateral injection of presynaptic AAV mix (blue) into the pons and unilateral injection of postsynaptic AAV mix (orange) into the cerebellum.

(B) Confocal images of pons neurons expressing PreSynBar. Left: low magnification showing widespread expression of PreSynBar in pons neurons. Right: high magnification showing cellular expression pattern. Blue: DAPI; White: PreSynBar signal stained with anti-V5 antibody. Scale bars: left panel 50 μm, right panel 10 μm.

(C) PreSynBar projection to the cerebellum. Left: low magnification showing cerebellar lobules with PreSynBar-positive mossy fiber terminals from pons (white, anti-V5 antibody staining) in the granule cell layer outlined by strong DAPI signal (blue). Right: high magnification showing individual mossy fiber terminals expressing PreSynBar. Left panel 100 μm, right panel 10 μm.

(D) PostSynBar expression across different cerebellar cell types. Left: low magnification view showing distinct PostSynBar expression patterns in molecular layer interneurons, Purkinje cells, and granule cells based on their characteristic locations relative to the granule cell layer. Right: high magnification showing cellular morphology details with dense granule cell dendrites and one Purkinje cell soma. Blue: DAPI; White: PostSynBar signal stained with anti-HA antibody. left panel 50 μm, right panel 10 μm.

(E) Split-GFP reconstitution at synaptic contacts. From left to right: PreSynBar-expressing mossy fiber terminals (cyan), PostSynBar-expressing granule cells (magenta), split GFP signals marking connected synapses (green), and merged image showing colocalization. Scale bar: 10 μm.

(F) RNA fluorescence in situ hybridization (FISH) validating synaptic barcode localization in the cerebellum. From left to right: PreSynBar protein (cyan), PostSynBar protein (magenta), PreRNA RNA (red), PostRNA RNA (yellow), and merged image showing colocalization. Scale bar: 10 μm.

(G) UMAP visualization of major cell populations in the pons, showing distinct clusters of neuronal populations with different neurotransmitter expression, including glutamatergic (Fat2 Glut and Hoxb5 Glut), GABAergic (Gata3 Gly-Gaba, Onecut1 Gaba, Otp Pax3 Gaba), serotonergic (Tph2 Glut-Sero), as well as various glial cell types including astrocytes (Astro), oligodendrocytes (ODC), oligodendrocyte precursor cells (OPC), endothelial cells (Endo), macrophage (Macro), and microglia (Micro).

(H) UMAP visualization of major cerebellar cell populations, highlighting granule cells, Purkinje cells, Golgi cells, unipolar brush cells (UBC), molecular layer interneurons (MLI1 and MLI2), Purkinje cell layer interneurons (PLI), and various non-neuronal cell types including Bergmann glia, astrocytes (Astro), oligodendrocytes (ODC), oligodendrocyte precursor cells (OPC), endothelial cells (Endo), ependymal cell (Epen), macrophage (Macro), microglia (Micro), and choroid plexus cells (Choroid).

(I) UMAP visualization showing PreRNA expression distribution across all pontine cell populations. PostRNA has virtually no expression as expected. Scale bars indicate log2 relative barcode expression levels.

(J) UMAP visualization showing PostRNA expression distribution across all cerebellar cell populations. PreRNA has virtually no expression as expected. Scale bars indicate log2 relative barcode expression levels.

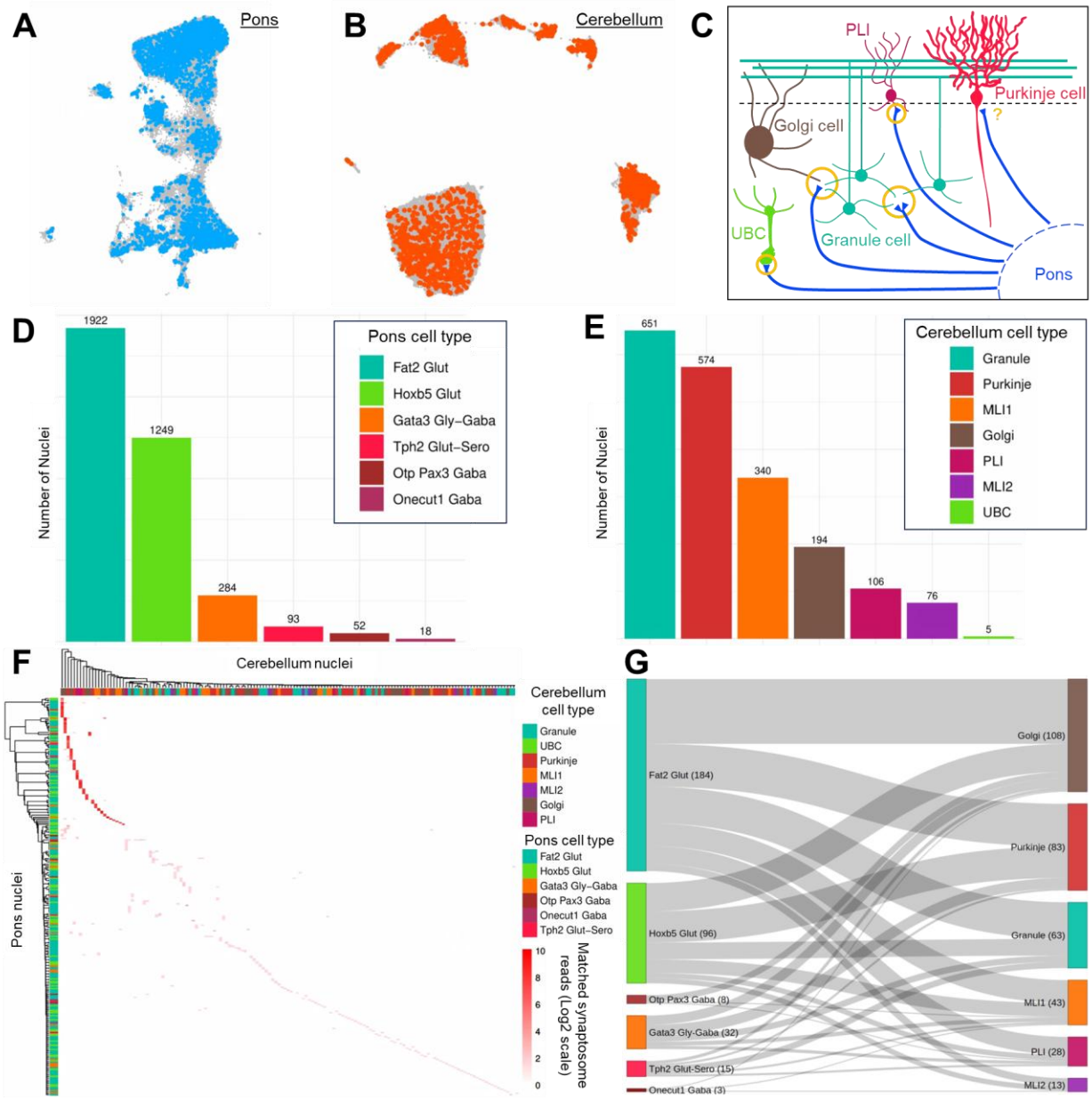


Figure 5. Single-cell resolution mapping of the pontocerebellar circuit

(A) UMAP visualization of pons nuclei colored by matched status (blue = synaptosomally matched nuclei, gray = unmatched nuclei).

(B) UMAP visualization of cerebellar nuclei colored by matched status (orange = synaptosomally matched nuclei, gray = unmatched nuclei).

(C) Schematic illustration of the canonical pontocerebellar circuit architecture. Major cell types are shown with their characteristic morphologies and relative positions. Major cellular targets of pontine axons (blue lines) include granule cells (teal) and Golgi cells (brown), as well as other targets including UBCs (unipolar brush cells, green) and PLI (Purkinje layer interneurons,

magenta). Purkinje cell (red) is shown with its characteristic dendritic arbors, while its status as a direct pontine target is under investigation.

(D) Quantification of matched cell type distribution in the pons (n=3,618 nuclei). Cell types include Fat2 Glut (n=1,922), Hoxb5 Glut (n=1,249), Gata3 Gly-GABA (n=284), Tph2 Glut-Sero (n=93), Otp Pax3 Gaba (n=52), and Onecut1 Gaba (n=18).

(E) Quantification of matched cell type distribution in the cerebellum (n=1,945 nuclei). Distribution shows Granule cells (n=651), Purkinje cells (n=574), MLI1 (n=340), Golgi cells (n=194), PLI (n=106), MLI2 (n=76), and UBC (n=5).

(F) Cell type-specific connectivity matrix between pons neurons (rows, n=281) and cerebellar neurons (columns, n=150). Color intensity indicates synaptosome read counts (Log2 scale) supporting each neuron pair connection. Hierarchical clustering reveals organized patterns of connectivity.

(G) Sankey diagram quantifying synaptosomal connections between major cell types. Width of flows represents the number of unique synaptosomes between each cell type pair. Numbers indicate total unique matched synaptosomes detected for each cell type: Fat2 Glut (184), Hoxb5 Glut (96), Gata3 Gly-GABA (22), Tph2 Glut-Sero (15), Otp Pax3 Gaba (8), and Onecut1 Gaba (3) from pons connecting to Golgi (108), Purkinje (83), Granule (63), MLI1 (43), PLI (28), and MLI2 (13) in cerebellum.

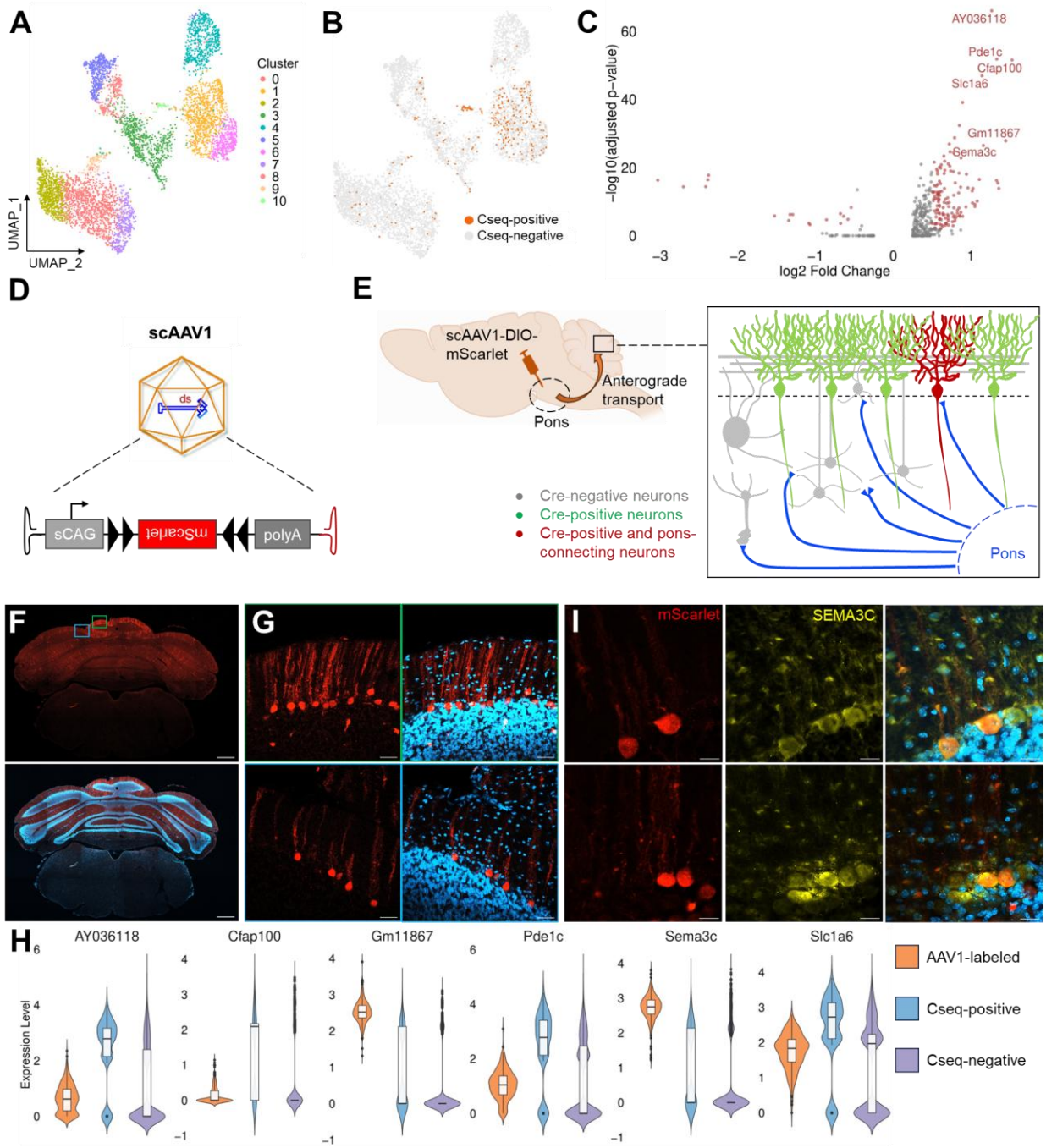


Figure 6. Discovery and validation of direct pons-to-Purkinje cell connectivity

(A) UMAP visualization of single-nucleus RNA sequencing data from cerebellar Purkinje cells, revealing 11 distinct molecular populations (clusters 0-10) based on transcriptional profiles.

(B) UMAP plot of Purkinje cells identified through Connectome-seq as receiving direct pontine inputs (Cseq-positive, orange) or lacking pontine connections (Cseq-negative, gray), demonstrating non-random distribution of pontine-connected cells across molecular subtypes.

(C) Volcano plot showing differential gene expression analysis between pons-connected and non-connected Purkinje cells. Top differentially expressed genes (*Slc1a6*, *Pde1c*, *Cfap100*, *Gm11867*, *AY036118*, *Sema3c*) are highlighted in red text. X-axis shows log2 fold change and y-axis shows -log10(adjusted p-value).

(D) Design of self-complementary AAV1 (scAAV1) construct for anterograde trans-synaptic tracing. Vector contains self-complementary AAV genome with short CAG promoter driving Cre-dependent mScarlet expression, flanked by inverted terminal repeats (ITRs).

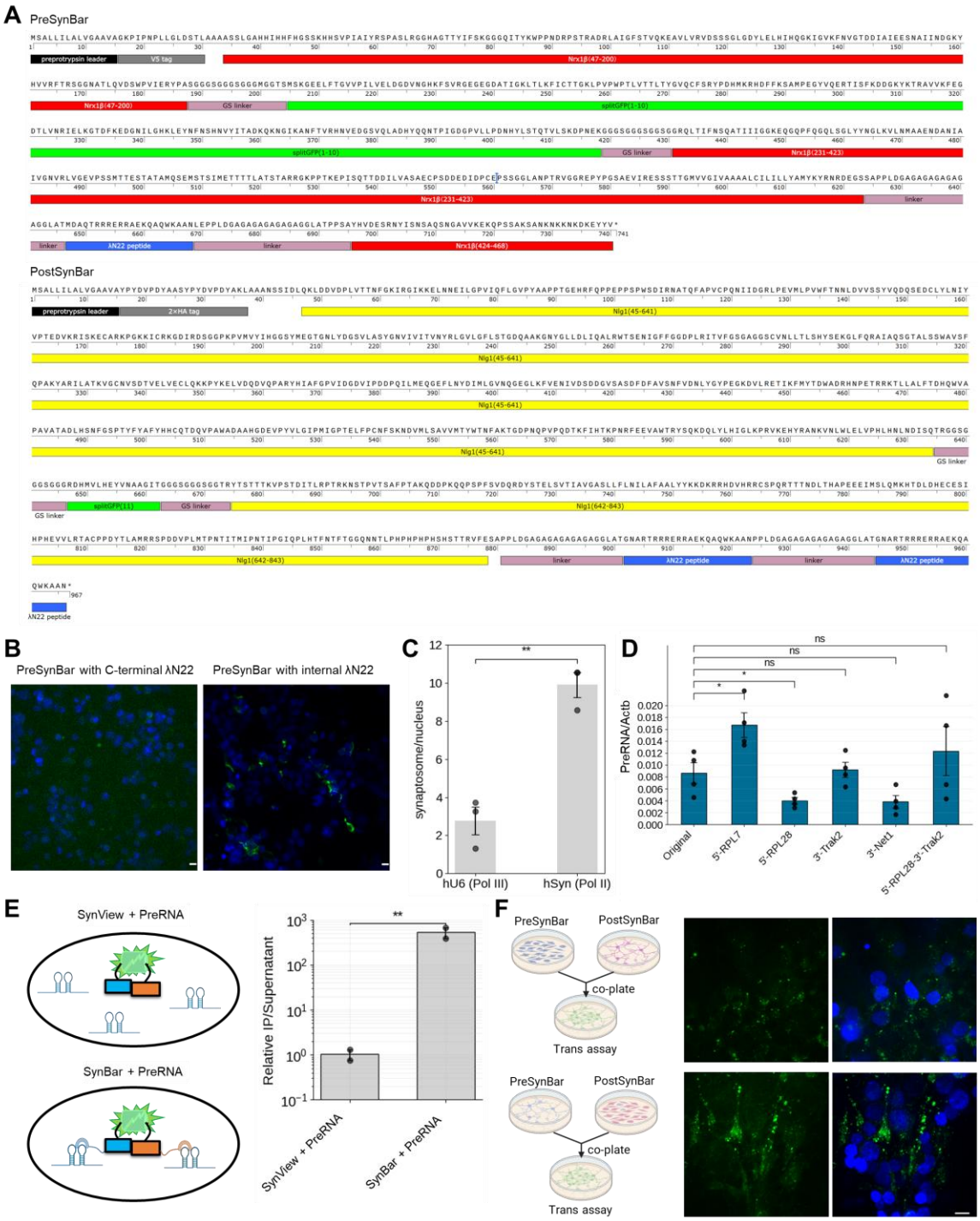
(E) Injection scheme of scAAV1-DIO-mScarlet into pons of Pcp2-Cre mouse and expected distribution of RFP-positive neurons.

(F) Anterograde trans-synaptic tracing confirms direct pontine inputs to adult Purkinje cells. Following injection of Cre-dependent AAV1-mScarlet virus into the pons of adult Pcp2-Cre mice, labeled Purkinje cells (red) are visible across cerebellar lobules, validating persistent pons-to-Purkinje connectivity. Nuclei counterstained with DAPI (blue). Scale bar: 600 μm.

(G) Representative high-magnification images of AAV1-labeled Purkinje cells. Left: mScarlet fluorescence (red) reveals detailed cellular morphology of pons-connected Purkinje cells. Right: Merged images with DAPI (blue) show characteristic positioning of Purkinje cell bodies and their elaborate dendritic trees extending into the molecular layer. Scale bar: 40 μm.

(H) Expression levels of top marker genes in three different groups of Purkinje cells: AAV1-labeled (orange), Cseq-positive (blue), and Cseq-negative (purple), shown as violin plots with boxplots overlay.

(I) Immunofluorescence validation of marker gene expression. Left: mScarlet-labeled pontine-connected Purkinje cells (red); Middle: SEMA3C immunostaining (yellow); Right: Merged images with DAPI (cyan) showing colocalization. Two representative fields shown. Scale bars: 20 μm.



Extended Data Figure 1: Design and Optimization of SynBar Protein-RNA Complex

(A) Protein domain organization of PreSynBar and PostSynBar constructs. PreSynBar contains modified neurexin 1β (red) with embedded split-GFP(1-10) (green), V5 tag (grey), and λN22 RNA-binding domain (blue). PostSynBar contains modified neuroligin 1 (yellow) with embedded

split-GFP(11) (green), HA tag (grey), and λN22 domain (blue). Numbers indicate amino acid positions.

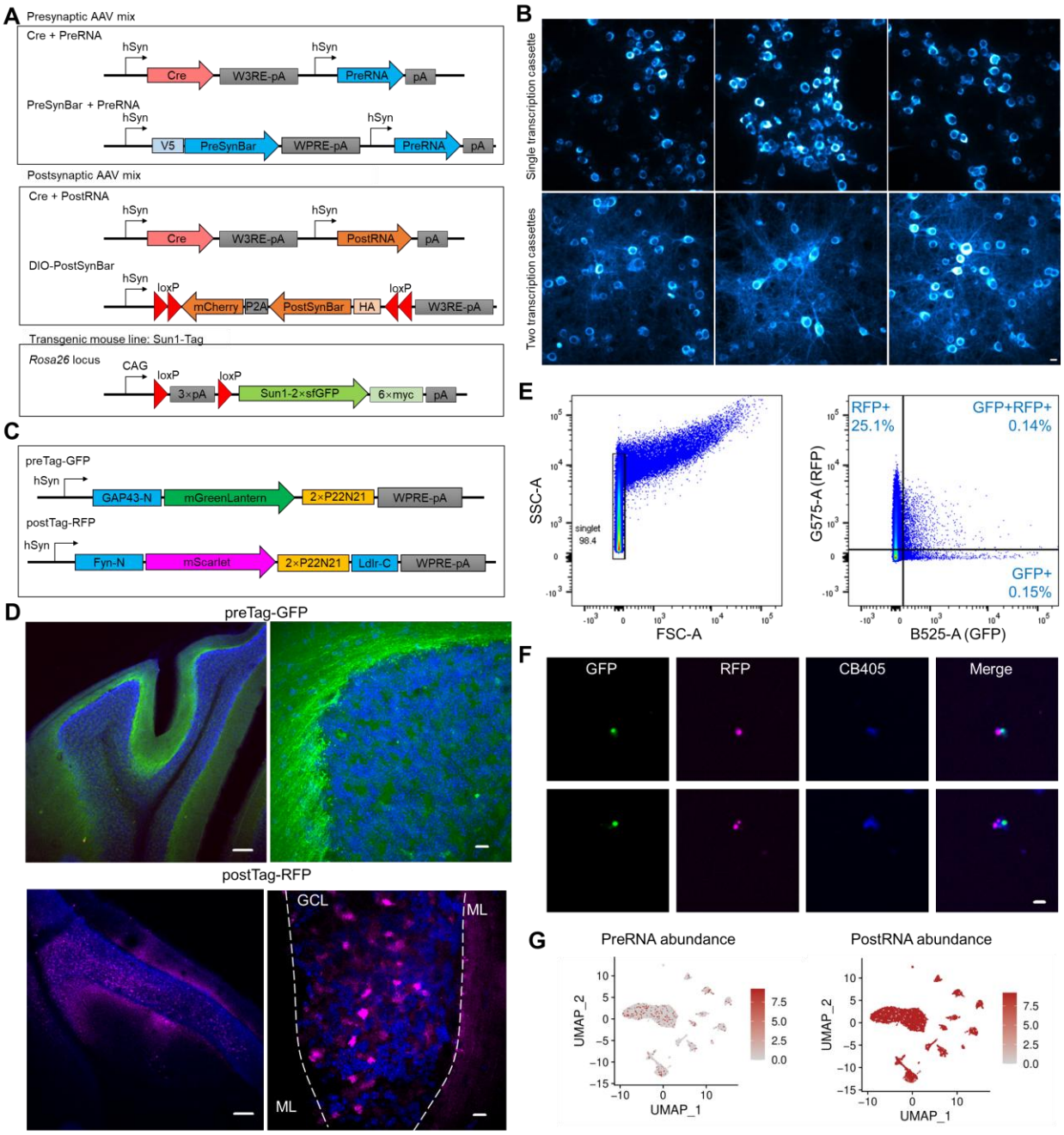
(B) Trans-interaction assay in HEK 293 cells demonstrating the importance of λN22 positioning in PreSynBar. Left: PreSynBar with C-terminal λN22 failed to show proper interaction with PostSynBar, so no split GFP signal was observed. Right: PreSynBar with internal C-terminal λN22 showed robust trans-synaptic interaction with PostSynBar evidenced by clear split GFP signals. Scale bars: 10 μm. Blue: DAPI; Green: reconstituted split GFP signal.

(C) Comparison of barcode expression levels under different promoters. RNA Polymerase II-driven expression (hSyn promoter) shows higher barcode abundance in the synaptosomal fraction (normalized to nuclear fraction) compared to Polymerase III-driven expression (hU6 promoter). Data are represented as mean ± SEM from n = 3 biological replicates. Statistical significance was determined using unpaired two-tailed Student's t-test (**$p < 0.01$).

(D) Optimization of PreRNA localization using various targeting sequences. After co-injecting AAVs expressing PreSynBar and different PreRNA designs into presynaptic brain regions, the abundance of PreRNA in postsynaptic synaptosomal fraction were measured by qPCR (normalized to housekeeping gene Actb) for five different barcode designs: Original unmodified PreRNA, RPL7 5'-UTR, RPL28 5'-UTR, Trak2 3'-UTR, Net1 3'-UTR, and combined RPL28 5'-UTR + Trak2 3'-UTRs. RPL7 5'-UTR shows highest enrichment. Data are represented as mean ± SEM from n = 4 biological replicates. Statistical significance was determined using one-way ANOVA followed by Dunnett's multiple comparisons test comparing all conditions to Original (*$p < 0.05$; ns, not significant).

(E) RNA immunoprecipitation (RIP) demonstrating stable λN22-BoxB RNA interaction. Left: Schematic showing experimental design – HEK 293 cells were co-transfected with either PreSynBar+PostSynBar or PreSynView+PostSynView (control lacking λN22 domains) along with PreRNA containing BoxB sequences. Right: Quantification of PreRNA enrichment in anti-GFP immunoprecipitates versus supernatant, shown as log10 ratio. SynBar shows >500-fold higher RNA enrichment compared to SynView control, confirming stable and specific BoxB-λN22 interaction. Data are represented as mean ± SEM from n = 2 biological replicates. Statistical significance was determined using unpaired two-tailed Student's t-test on log-transformed values (**$p < 0.01$).

(F) Trans-interaction assay validating SynBar interaction between different cell types. Left: Experimental schematic showing separate plating and co-culture conditions: HEK 293 cells expressing PreSynBar with neurons expressing PostSynBar (top), and neurons expressing PreSynBar with HEK 293 cells expressing PostSynBar (bottom). Right: Representative immunofluorescence images showing specific GFP reconstitution (green) at contact points between the different cell types in both configurations. Scale bar: 10 μm. Blue: DAPI nuclear staining.



Extended Data Figure 2: AAV Construct Design, Optimization, and Alternative Fluorescent Protein Approach

(A) Design of viral constructs and transgenic mouse line used in Connectome-seq. Top: Presynaptic AAV mix containing two viruses - a Cre-PreRNA virus and a PreSynBar-PreRNA virus, both driven by two hSyn promoter cassettes. Middle: Postsynaptic AAV mix containing Cre-PostRNA virus and Cre-dependent (DIO) PostSynBar virus. Bottom: Sun1-Tag transgenic mouse line showing the *Rosa26* locus with Cre-dependent expression of Sun1-2×sfGFP fusion protein.

(B) Confocal microscopy images illustrating the optimization of PreSynBar expression and localization in cultured neurons. Upper: Representative image of neurons infected with an AAV carrying a single transcriptional cassette driving the expression of V5-tagged PreSynBar. The protein shows predominantly somatic localization with limited distribution into neuronal processes. Lower: Representative image of neurons infected with an AAV carrying two transcriptional cassettes: one driving V5-tagged PreSynBar expression and another driving PreRNA expression. This design demonstrates more extensive localization of PreSynBar throughout neuronal processes. Note the comparable expression levels of PreSynBar in the soma between the two conditions. Pseudo color: V5 staining. Scale bars: 10 μm.

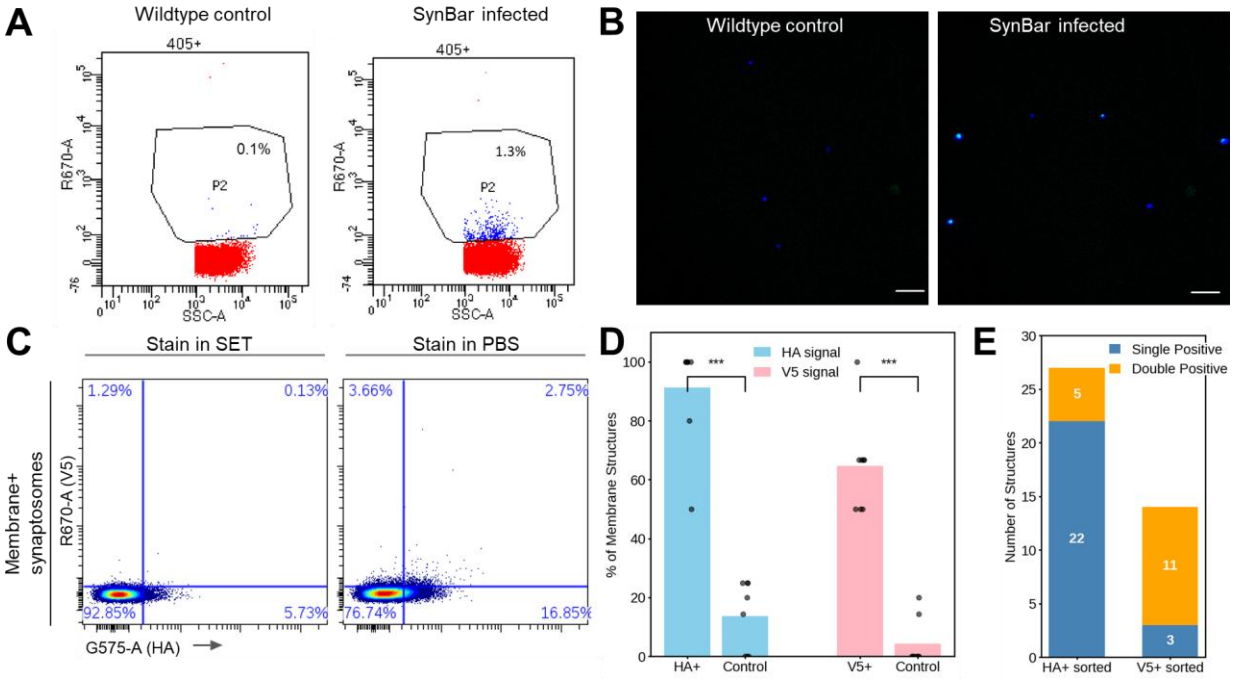
(C) Design of preTag-GFP and postTag-RFP constructs. PreTag-GFP contains GAP43-N targeting sequence, mGreenLantern, and RNA-binding domain (2×P22N21). PostTag-RFP contains Fyn-N and Ltdr-C targeting sequences, mScarlet, and RNA-binding domain. Both constructs are driven by hSyn promoter.

(D) In vivo expression pattern of fluorescent protein constructs. Left: PreTag-GFP shows strong axonal localization in cerebellar molecular layer (ML). Right: PostTag-RFP demonstrates dendritic targeting in granule cell layer (GCL). Blue: DAPI. Scale bars: left panels 100 μm, right panels 10 μm.

(E) Synaptosome sorting strategy using native fluorescence. Left: Initial singlet gating based on scatter properties. Right: Identification of GFP+, RFP+, and double-positive populations.

(F) Imaging validation of sorted synaptosomes. Representative confocal images of sorted double-positive synaptosomes showing membrane staining (CellBrite405, blue), RFP signal (red), and GFP signal (green). Scale bars: 2 μm.

(G) Analysis of barcode abundance in cerebellar nuclei expressing fluorescent constructs. UMAP visualization of single-nucleus RNA-seq data showing PreRNA (left) and PostRNA (right) expression levels. Cerebellar nuclei were infected with PostRNA-expressing virus, and PreRNA signal came from contamination of ambient PreRNA from pons projections. Color intensity indicates barcode abundance.



Extended Data Figure 3: Optimization of Synaptosome Immunostaining

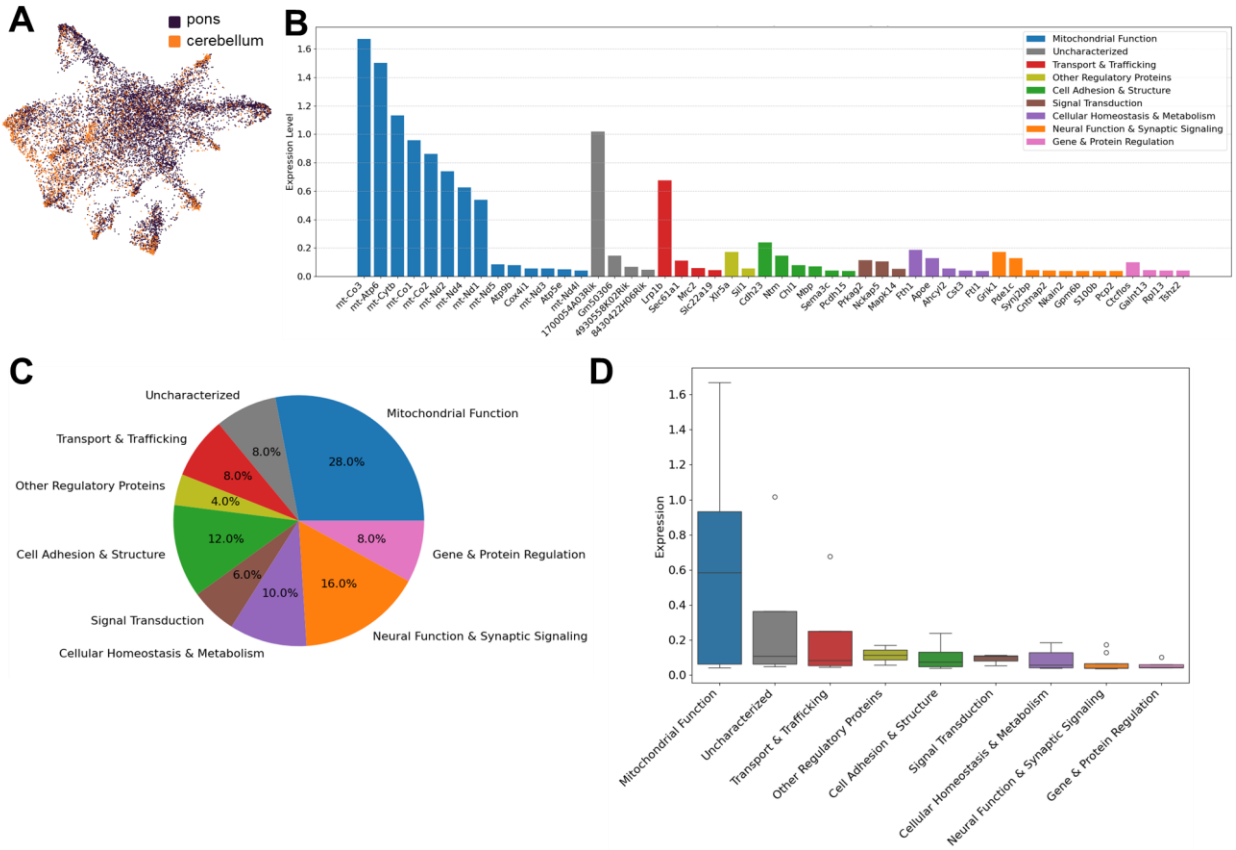
(A) Flow cytometry detection of reconstituted GFP using membrane+ gating. Wildtype control (left) versus SynBar-infected sample (right) shows specific detection of GFP+ synaptosomes (1.3% of membrane+ population) in infected tissue only.

(B) Confocal imaging validation of sorted GFP+ synaptosomes from wildtype and SynBar-infected samples. Blue: CellBrite405 membrane dye; Green: anti-GFP signal. Scale bar: 10 µm.

(C) Comparison of antibody staining efficiency in different buffers. Staining in sucrose-based SET buffer (left) shows reduced signal compared to PBS-based buffer (right), with higher detection of positive populations in PBS (2.75% vs 0.13% double-positive).

(D) Quantification of marker presence in sorted populations. Bar graph showing the percentage of membrane-intact structures (based on CellBrite 405 signal) that are positive for HA or V5 in HA+ sorted populations versus control populations or V5+ sorted populations versus control populations. Error bars represent SEM; ***$p < 0.001$.

(E) Analysis of single versus double-positive structures across sorted populations. Bar graph showing the number of structures that are either single positive (blue) or double positive (orange) for V5 and HA markers in HA+ sorted and V5+ sorted populations.



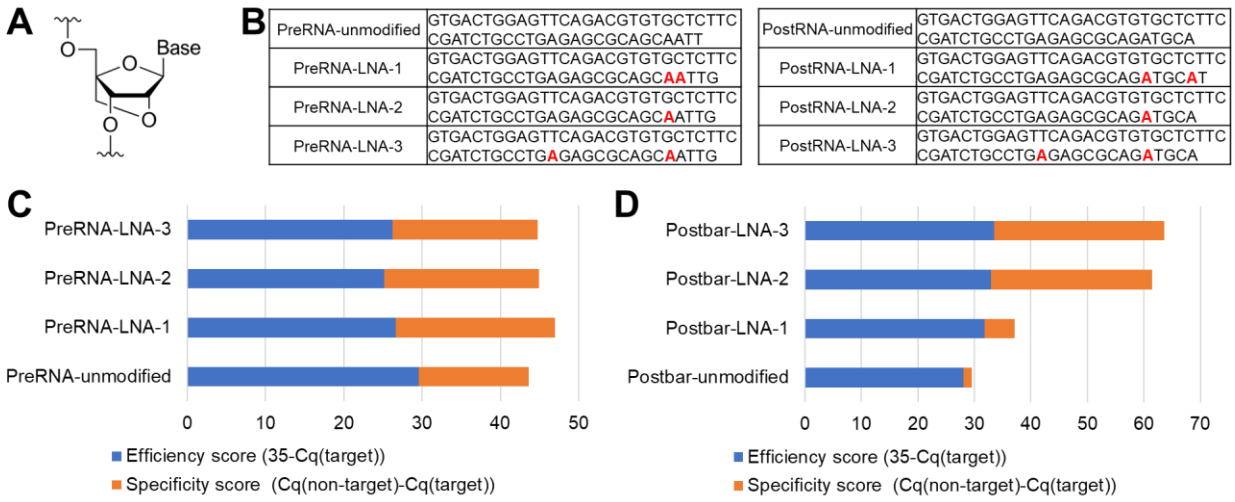
Extended Data Figure 4. Analysis of synaptosomal gene expression from single-synaptosome mRNA sequencing data

(A) UMAP visualization of single-synaptosome RNA sequencing data from pons (purple) and cerebellum (orange). The extensive mixing of synaptosomes from different brain regions indicates insufficient gene detection sensitivity to distinguish synaptic origins based on transcriptome alone.

(B) Expression levels of top 50 most abundant transcripts detected in synaptosomes, colored by gene category. Mitochondrial transcripts (blue) dominate the synaptosomal RNA pool, with limited representation of transport and synapse-related genes.

(C) Proportion of different gene categories among top 50 expressed genes in synaptosomes. Mitochondrial genes comprise 28% of highly expressed transcripts.

(D) Expression level distributions across gene categories, demonstrating higher expression of mitochondrial transcripts compared to other categories. Boxes show quartiles, whiskers extend to 1.5× interquartile range, and points indicate outliers.



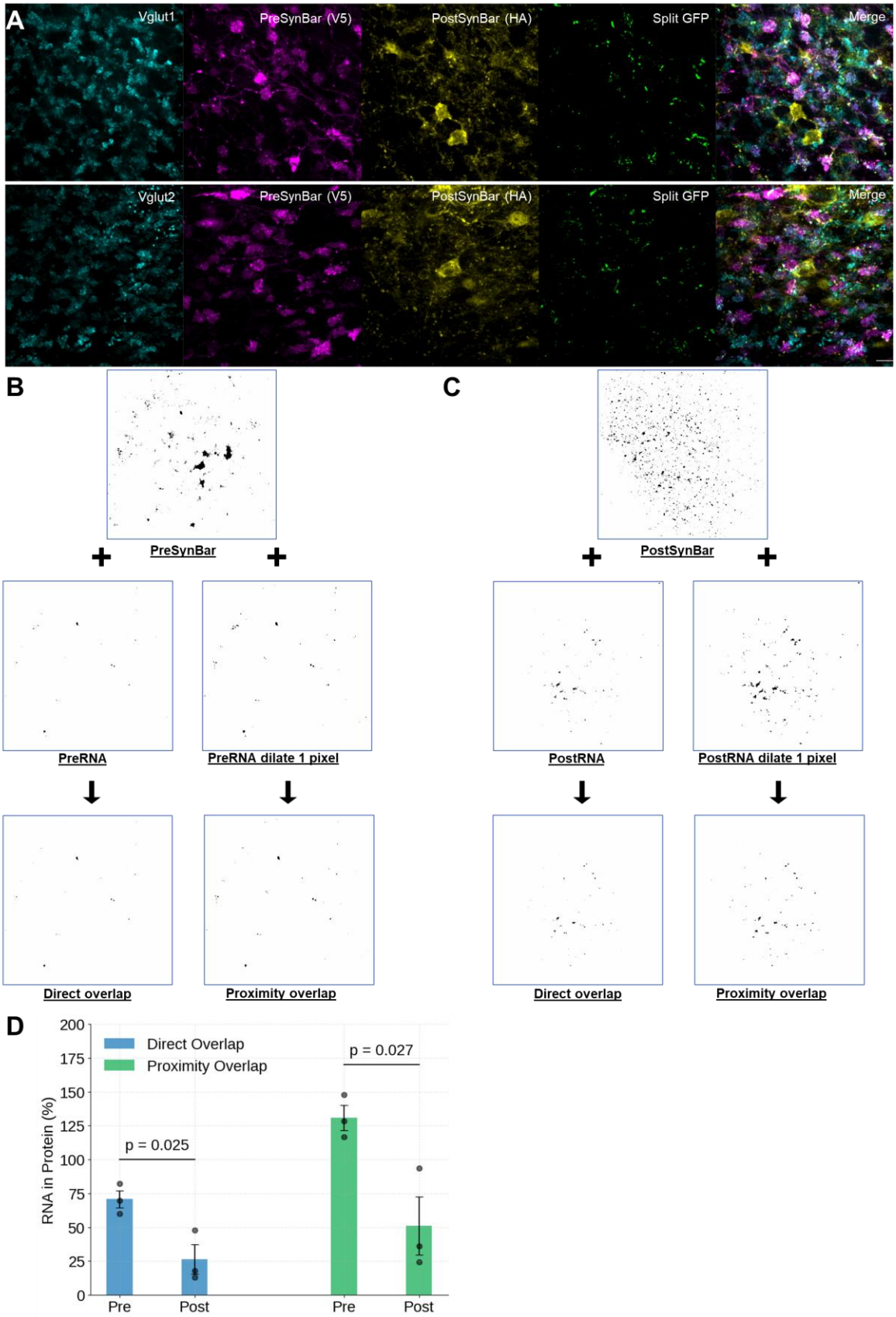
Extended Data Figure 5. Design and optimization of LNA-modified primers for balanced barcode amplification

(A) Chemical structure of locked nucleic acid (LNA) nucleotide showing the methylene bridge between 2' oxygen and 4' carbon that "locks" the ribose in C3'-endo conformation.

(B) Sequence comparison of unmodified and LNA-modified primers for PreRNA (left) and PostRNA (right) amplification. LNA modifications (shown in red) were systematically introduced at different positions to optimize primer performance. Each version (LNA-1, LNA-2, LNA-3) represents different modification patterns.

(C) Performance evaluation of PreRNA primers. Bar plots show efficiency score (35-Cq(target), blue) and specificity score (Cq(non-target)-Cq(target), orange) for unmodified and LNA-modified primers. Higher scores indicate better performance. PreRNA-LNA-1 was chosen as the optimized version.

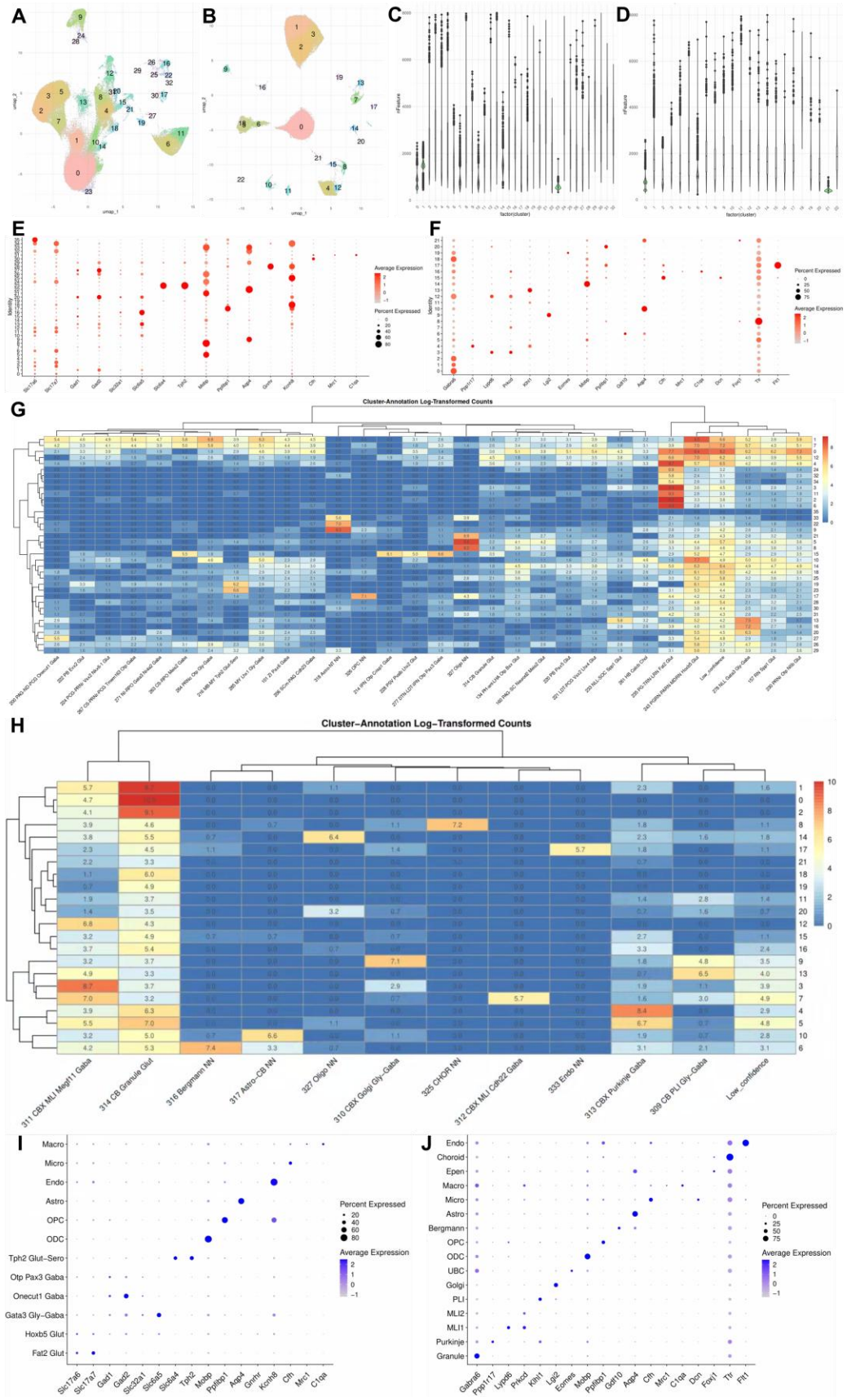
(D) Performance evaluation of PostRNA primers showing improved amplification efficiency and specificity with LNA modifications, particularly for LNA-2 and LNA-3 versions. Note the substantial improvement in specificity score compared to unmodified primers. PostRNA-LNA-3 was chosen as the optimized version.


A
Vglut1
PreSynBar (V5)
PostSynBar (HA)
Split GFP
Merge
Vglut2
PreSynBar (V5)
PostSynBar (HA)
Split GFP
Merge
B
PreSynBar
PreRNA
PreRNA dilate 1 pixel
Direct overlap
Proximity overlap
C
PostSynBar
PostRNA
PostRNA dilate 1 pixel
Direct overlap
Proximity overlap
D
Direct Overlap
Proximity Overlap
p = 0.025
p = 0.027
RNA in Protein (%)
200
175
150
125
100
75
50
25
0
Pre
Post
Pre
Post

Extended Data Figure 6. Validation of SynBar protein expression and RNA-protein colocalization

(A) Validation of PreSynBar expression with endogenous presynaptic markers, together with split GFP reconstitution. Upper panel, from left to right: Vglut1 immunostaining (cyan), PreSynBar signal (anti-V5, magenta), PostSynBar signal (anti-HA, yellow), Split GFP signals marking connected synapses (green), and merged image showing colocalization. Lower panel shows the same arrangement but with Vglut2 instead of Vglut1 in the leftmost channel. Representative max z projection field-of-view shown. Scale bar: 10 μm.

(B-D) Quantitative analysis of RNA-protein colocalization. (B) Representative images showing the workflow for PreRNA-PreSynBar colocalization analysis. Binary masks were generated for RNA (FISH signal) and protein (immunofluorescence) channels. Direct overlap was calculated from the intersection of RNA and protein masks. Proximity overlap was determined after dilating the RNA mask by 1 pixel. (C) Similar analysis workflow for PostRNA-PostSynBar colocalization. (D) Quantification of RNA-protein colocalization (n=3 independent samples). Both direct overlap and proximity analysis showed significantly higher colocalization for PreRNA with PreSynBar compared to PostRNA with PostSynBar (direct overlap: $p=0.025$; proximity overlap: $p=0.027$, two-tailed t-test).


A
B
C
D
E
F
G
Cluster-Annotation Log-Transformed Counts
H
Cluster-Annotation Log-Transformed Counts
311 CBX MLI Megf11 Gaba
314 CB Granule Glut
316 Bergmann NN
317 Astro-CB NN
327 Oligo NN
310 CBX Golgi Gly-Gaba
325 CHOR NN
312 CBX MLI Cdh22 Gaba
333 Endo NN
313 CBX Purkinje Gaba
309 CB PLI Gly-Gaba
Low_confidence
I
Macro
Micro
Endo
Astro
OPC
ODC
Tph2 Glut-Sero
Otp Pax3 Gaba
Onecut1 Gaba
Gata3 Gly-Gaba
Hoxb5 Glut
Fat2 Glut
Percent Expressed
Average Expression
J
Endo
Choroid
Epen
Macro
Micro
Astro
Bergmann
OPC
ODC
UBC
Golgi
PLI
MLI2
MLI1
Purkinje
Granule
Percent Expressed
Average Expression

Extended Data Figure 7. Cell type annotation of nuclei from pons and cerebellum.

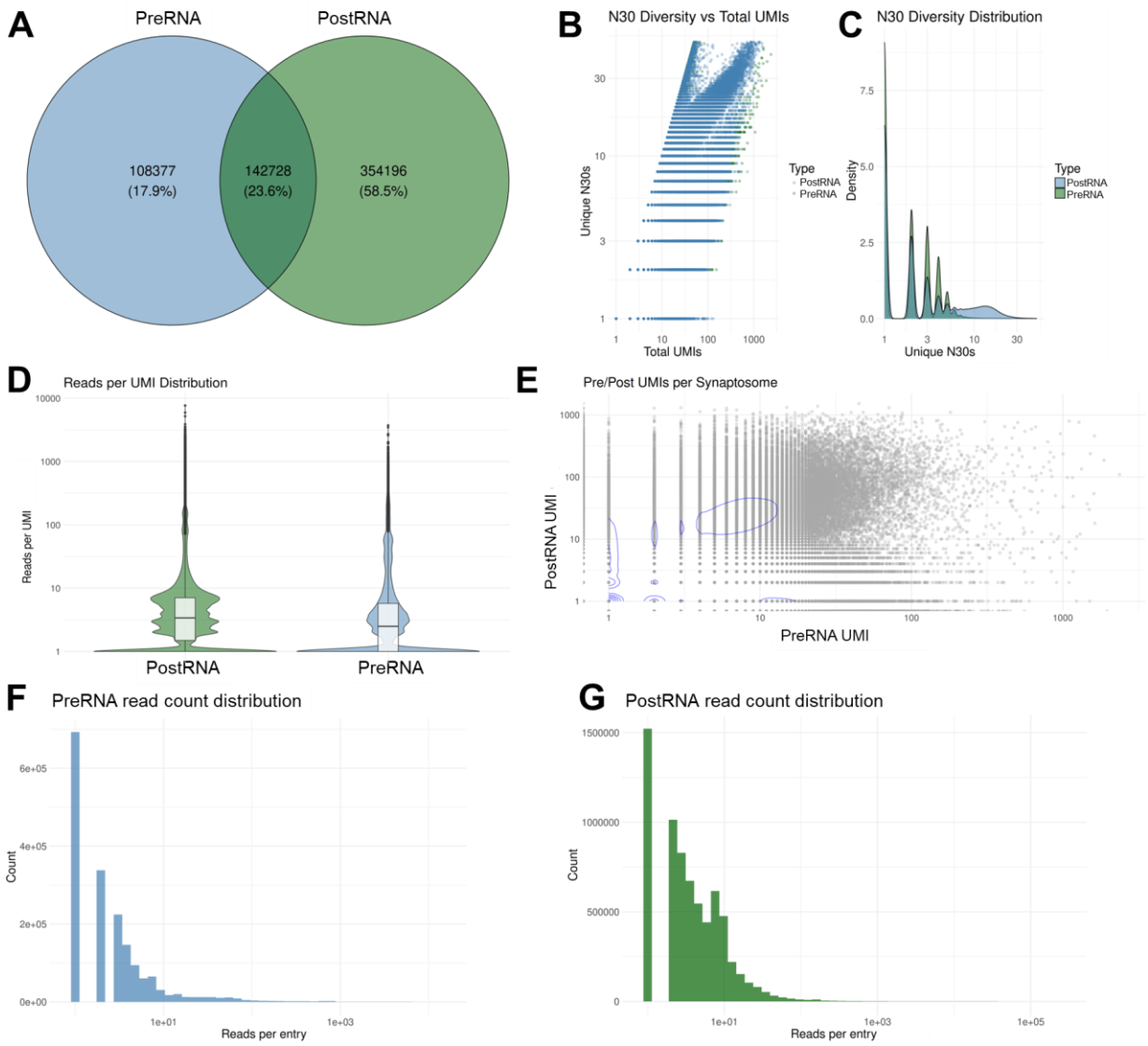
(A-B) Initial UMAP visualization of cell clusters in the pons (A) and cerebellum (B), with clusters numbered.

(C-D) Violin plots showing gene counts per cluster, used to identify and remove low-quality clusters. Clusters 0, 23 were filtered out from pons (C), and clusters 0, 21 were removed from cerebellum (D) due to low gene counts (<1,000).

(E-F) Dot plots for initial cell type annotation using canonical marker genes. For pons (E), these include neurotransmitter transporters (*Slc17a6, Slc17a7, Gad1, Gad2, Slc32a1, Slc6a5, Slc6a4*), synthesizing enzyme (*Tph2*), and non-neuronal cell markers. For cerebellum (F), markers include those for granule cells, interneurons (Golgi, UBC, MLI1/2, PLI), Purkinje cells, and non-neuronal cells. Dot size indicates percentage of expressing cells (20-80% for pons, 0-75% for cerebellum) and color intensity represents average expression level (-1 to 2).

(G-H) Allen Institute's MapMyCell annotation results shown as hierarchical clustering heatmaps for pons (G) and cerebellum (H), displaying the proportion of each annotated cell type within individual clusters to assign cluster identity.

(I-J) Final dot plots showing expression of the same canonical marker genes across merged and annotated cell populations in pons (I) and cerebellum (J), confirming the consistency between marker gene expression and final cluster annotations. Dot size indicates percentage of expressing cells and color intensity shows average expression level.



Extended Data Figure 8. Analysis of synaptosomal sequencing data reveals distinct patterns of PreRNA and PostRNA distribution.

(A) Venn diagram showing the overlap between synaptosomes containing PreRNA and PostRNA. Numbers indicate unique synaptosomes in each category, with 143,034 (23.6%) containing both barcodes.

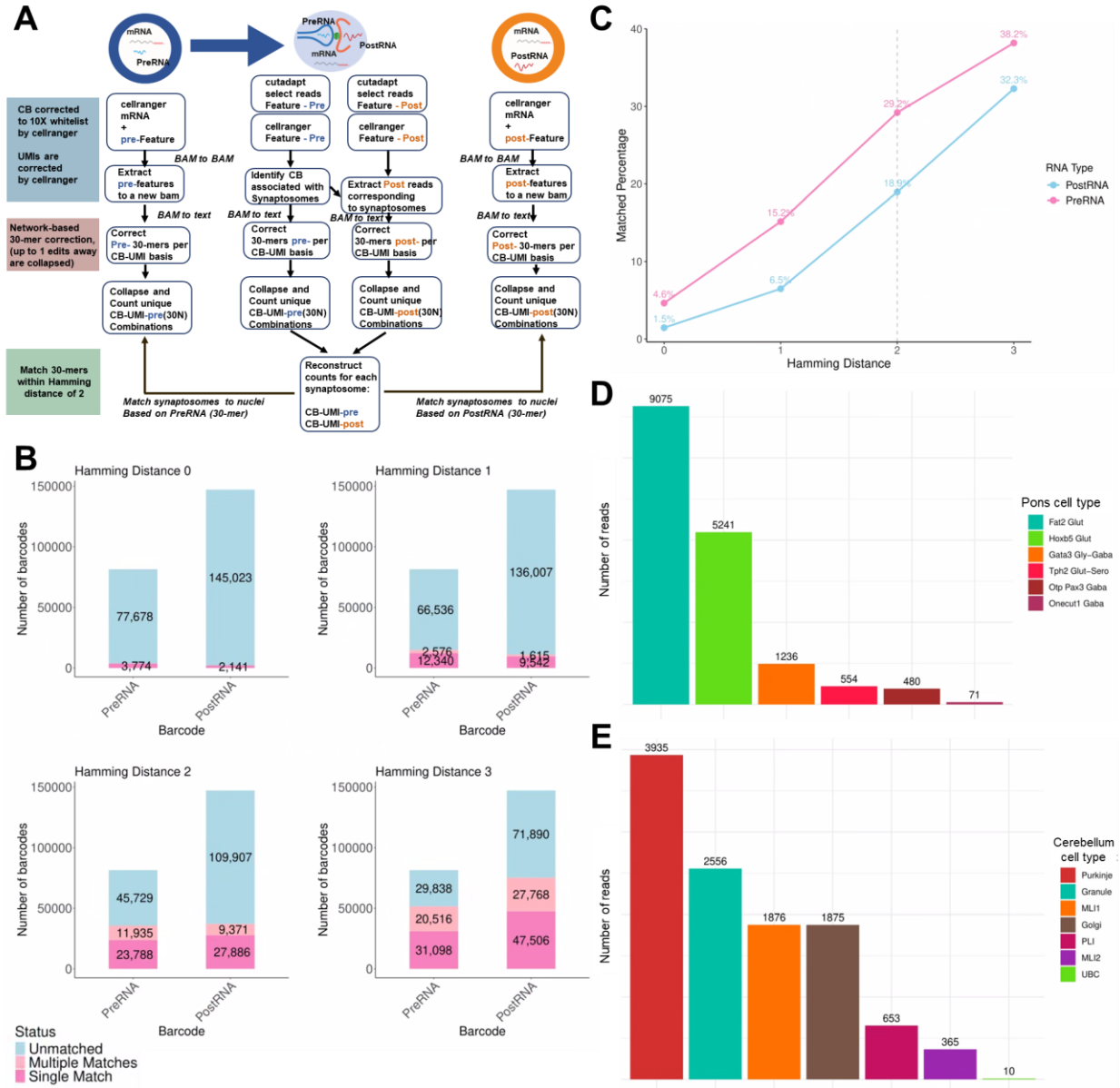
(B) Relationship between N30 diversity and total UMI counts per synaptosome, shown on log10 scales. Each point represents a single synaptosome, with PreRNA (blue) and PostRNA (green) showing similar scaling patterns.

(C) Distribution of N30 diversity (unique N30s per synaptosome) for PreRNA and PostRNA populations, demonstrating the frequency of different complexity levels in the synaptosomal population.

(D) Violin plots showing the distribution of reads per UMI for PostRNA (left) and PreRNA (right). Box plots within violins indicate median and quartile values, with extensions showing the full range excluding outliers.

(E) Scatter plot comparing PreRNA and PostRNA UMI counts per synaptosome (log10 scales). Grey points represent individual synaptosomes, with density contours highlighting the distribution pattern.

(F-G) Read count distributions for PreRNA (F) and PostRNA (G) entries, showing the frequency of different read depths on a log10 scale. Note the different y-axis scales between the two plots, reflecting the higher abundance of PostRNA in the dataset.



Extended Data Figure 9. Computational pipeline and analysis for pontocerebellar connectome reconstruction

(A) Workflow of the computational pipeline showing parallel processing of nuclear mRNAs and barcodes. Key steps include CB-UMI-based error correction, feature extraction, network-based 30-mer correction, and barcode matching across compartments.

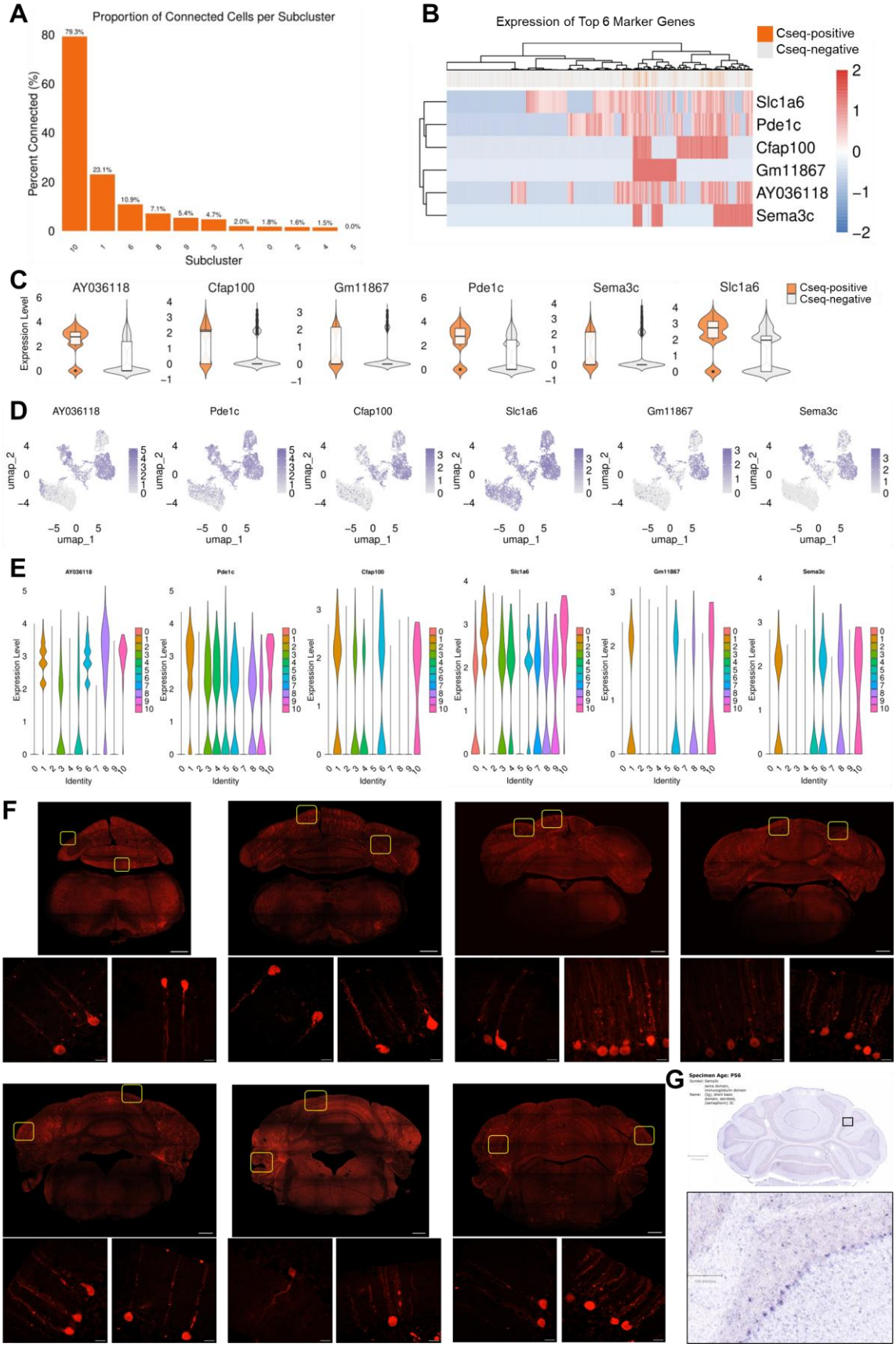
(B) Quantification of matched and unmatched barcodes at different Hamming distances (0-3). Each plot shows the distribution between PostRNA and PreRNA, with barcodes matching with a unique nucleus (Single Match) in pink and unmatched in blue, while barcodes matching with multiple nuclear references (Multiple Matches) shown in light pink. Only Single Match barcodes

were passed down to downstream analysis. Increasing matching rates were observed at higher Hamming distances.

(C) Graph showing the relationship between single matched percentage and Hamming distance for both PostRNA (blue) and PreRNA (pink). The dashed line indicates the chosen Hamming distance = 2 for subsequent analyses.

(D) Distribution of total matched reads across pons cell types. Each barcode-containing read from synaptosomes that matched to a nucleus was counted, showing the total read coverage per cell type: Fat2 Glut (9,075 reads), Hoxb5 Glut (5,241 reads), Gata3 Gly-GABA (1,236 reads), Tph2 Glut-Sero (554 reads), Otp Pax3 Gaba (480 reads), and Onecut1 Gaba (71 reads). This represents the raw read counts from the synaptosomal PreRNA that matched with nuclei at Hamming distance 2.

(E) Distribution of total matched reads across cerebellar cell types. Each barcode-containing read from synaptosomes that matched to a nucleus was counted, showing read coverage per cell type: Purkinje cells (3,935 reads), Granule cells (2,556 reads), MLI1 (1,876 reads), Golgi cells (1,875 reads), PLI (653 reads), MLI2 (365 reads), and UBC (10 reads). This represents the raw read counts from the synaptosomal PostRNA that matched with nuclei at Hamming distance 2.


A
Proportion of Connected Cells per Subcluster
Percent Connected (%)
Subcluster
B
Expression of Top 6 Marker Genes
Cseq-positive
Cseq-negative
Slc1a6
Pde1c
Cfap100
Gm11867
AY036118
Sema3c
C
AY036118
Cfap100
Gm11867
Pde1c
Sema3c
Slc1a6
Expression Level
D
umap_1
umap_2
E
Identity
F
G
Specimen Age: P56

Extended Data Figure 10. Detailed molecular analysis of pons-connected Purkinje cells

(A) Quantification of pons-connected Purkinje cells across different subclusters, showing highest enrichment in cluster 10 (79.3%), followed by cluster 1 (23.1%), with decreasing proportions in other clusters, demonstrating preferential connectivity in specific molecular subtypes.

(B) Hierarchical clustering of expression patterns for top 6 marker genes (*Slc1a6*, *Pde1c*, *Cfap100*, *Gm11867*, *AY036118*, *Sema3c*) in Cseq-positive (orange) versus Cseq-negative (gray) Purkinje cells. Color scale indicates normalized expression levels from -2 (blue) to 2 (red).

(C) Two-condition violin plots showing expression distribution of top 6 marker genes in Cseq-positive (orange) versus Cseq-negative (gray) Purkinje cells, demonstrating consistent upregulation in connected populations.

(D) UMAP feature plots showing expression patterns of top 6 marker genes across all Purkinje cells. Color intensity indicates expression level.

(E) Detailed expression patterns of top 6 marker genes across all Purkinje cell subclusters (0-10). Each violin plot shows expression distribution within a subcluster, colored by cluster identity. Note the cluster-specific expression patterns correlating with connectivity.

(F) Representative images from serial sectioned mouse brain (from posterior to anterior) showing consistent labeling of Purkinje cells by AAV1 anterograde tracing. For each panel: left shows low magnification view of cerebellar section (red: mScarlet signal); right shows high magnification of boxed region with merged mScarlet and DAPI (blue) signals. Scale bars: 600 μm (upper panels), 40 μm (lower panels).

(G) *Sema3c* RNA in situ hybridization data from the Allen Brain Atlas (P56 mouse). Left: Low magnification image of coronal cerebellar section. Right: High magnification of boxed region showing *Sema3c* expression in a subset of Purkinje cells, providing independent validation of cell-type specific expression pattern.

Table S1. Related to Methods. Plasmids used or generated in this study.

Name	Vector	Promoter	Description
hSyn-PreSynBar-hSyn-PreRNA barcode landing pad	pAAV	hSyn	SynBar protein and RNA construct used in HEK and neuron culture experiment, related to Fig. 1
hSyn-PostSynBar-hSyn-PostRNA barcode landing pad	pAAV	hSyn	SynBar protein and RNA construct used in HEK and neuron culture experiment, related to Fig. 1
hU6-presynaptic-barcode-3BoxB	pAAV	hU6	hU6 promoter driven barcode RNA with no polyA, related to Fig. S1C
hU6-postsynaptic-barcode-3BoxB	pAAV	hU6	hU6 promoter driven barcode RNA with no polyA, related to Fig. S1C
CAG- PreRNA-PPIB-original-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL7-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL28-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
CAG- PreRNA-PPIB-3'Trak2-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
CAG- PreRNA-PPIB-3'Net1-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL28-3'Trak2-hSyn-Cre	pAAV	CAG and hSyn	PreRNA with different cis-element to potentially improve its synaptic localization, related to Fig. S1D
pre-SynView	pAAV	hSyn	Original pre-SynView construct cloned into pAAV vector, related to Fig. S1E
post-SynView	pAAV	hSyn	Original post-SynView construct cloned into pAAV vector, related to Fig. S1E
hSyn-Cre-hSyn-PreRNA	pAAV	hSyn	The set of SynBar constructs used in the optimized connectome-seq experiments, design shown in Fig.

barcode landing pad (PPIB)			S2A. Also used for Basescope experiment shown in Fig. 4F
hSyn-PreSynBar-hSyn-PreRNA barcode landing pad	pAAV	hSyn	The set of SynBar constructs used in the optimized connectome-seq experiments, design shown in Fig. S2A
hSyn-Cre-hSyn-PostRNA barcode landing pad	pAAV	hSyn	The set of SynBar constructs used in the optimized connectome-seq experiments, design shown in Fig. S2A
hSyn-DIO-PostSynBar-P2A-mCherry	pAAV	hSyn	The set of SynBar constructs used in the optimized connectome-seq experiments, design shown in Fig. S2A
hSyn-DIO-axonTag-mGreenLantern-P22N21	pAAV	hSyn	Axon-targeted GFP with RNA binding domain, design shown in Fig. S3A
DIO-dendriteTag-oScarlet3-P22N21	pAAV	hSyn	Dendrite-targeted RFP with RNA binding domain, design shown in Fig. S3A
hSyn-Cre-W3SL-hSyn-PostRNA-dapB	pAAV	hSyn	Used for Basescope experiment shown in Fig. 4F
sCAG-DIO-oScarlet3	pAAV	sCAG	Used for AAV1 tracing experiment, design shown in Fig. S12A
AAV1 capsid	N/A	N/A	Used for AAV generation, Addgene #112862
AAV2 capsid	N/A	N/A	Used for AAV generation, Addgene #104963
pAdDeltaF6	N/A	N/A	Used for AAV generation, Addgene #112867

Table S2. Related to Methods. Oligos used in the paper.

Name	Sequence	Modifications	Description
N30-oligo	AATGATACGGCGACCACCGAGGTCTCCAATTGVHDBBDHVATVHDBNNBDHVTAVHDBBDHVATGCATGAGACCCTACACGCCTCCCTCGCGCCATCAG	Containing randomly mixed nucleotides, B=C or G or T D=A or G or T H=A or C or T V=A or C or G N=any base	For cloning highly complex PreRNA and PostRNA library
N30-oligo reverse primer	CTGATGGCGCGAGGGAGGCGTGTAG	none	For cloning highly complex PreRNA and PostRNA library
PreRNA-unmodified	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAGCAATT	none	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PreRNA-LNA-1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAGC**AA**TTG	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PreRNA-LNA-2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAGC**A**ATTG	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PreRNA-LNA-3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTG**A**GAGCGCAGC**A**ATTG	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PostRNA-unmodified	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAGATGCA	none	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PostRNA-LNA-1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAG**A**TGC**A**T	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PostRNA-LNA-2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGAGAGCGCAG**A**TGCA	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction
PostRNA-LNA-3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTG**A**GAGCGCAG**A**TGCA	Underline marks locked nuclei acid modified base	For optimizing best primers to amplify PreRNA and PostRNA during library construction

bar6-F	ggagTCGTGTCGTGCCTGAGA	With and without 5'biotinylation	For PCR #1 and PCR #1.1 during library construction shown in Fig. 3
1N	GC AGC GTC AGA TGT GTA TAA GAG ACA G	5'biotinylation	For PCR #1 and PCR #1.1 during library construction shown in Fig. 3
P5-1N	AA TGA TAC GGC GAC CAC CGA GAT CTA CAC TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG	5'biotinylation	For PCR #2, #2.1, and #2.2 during library construction, together with LNA primers, shown in Fig. 3
2boxB universal Taqman probe	AGTCCAACTACTAAACTGGGCTAAGTCC	5'-FAM/3'-BHQ-1	For Taqman qPCR assays
PPIB-F	acaggcccatggactcactt	none	For six version PreRNA comparison Taqman qPCR with universal probe, related to Fig. S1D
PPIB-R	ATAATATCCTCGATAGGGCCCTTCT	none	For six version PreRNA comparison Taqman qPCR with universal probe, related to Fig. S1D
PreRNA-F	CCTGAGAGCGCAGCAAT	none	For PreRNA Taqman qPCR with universal probe
PreRNA-R	TGCTAGGACCGGCCTTAAAG	none	For PreRNA Taqman qPCR with universal probe
PostRNA-F	GCCTGAGAGCGCAGATG	none	For PostRNA Taqman qPCR with universal probe
PostRNA-R	TAGCCGCTAATAGGTGAGC	none	For PostRNA Taqman qPCR with universal probe
Actb-F	AGCCATGTACGTAGCCATCC	none	For Actb qPCR
Actb-R	CTCTCAGCTGTGGTGGTGAA	none	For Actb qPCR
hSyn-F	CCTGCGTATGAGTGCAAGT	none	For AAV titering
hSyn-R	TTGGGTGCTTGTCCAGTG	none	For AAV titering
WPRE-F	CTATGTTGCTCCTTTTACGC	none	For AAV titering
WPRE-R	ACCAGGATTTATACAAGGAGG	none	For AAV titering

Table S3. Related to Methods. AAV used for different experiments.

Name	Capsid	Titer (vg/ul)	Description
hSyn-PreSynBar-hSyn-PreRNA barcode landing pad	AAV1/2 mix	3.98E+09	For neuron culture infection, related to Fig. 1
hSyn-PostSynBar-hSyn-PostRNA barcode landing pad	AAV1/2 mix	9.13E+08	For neuron culture infection, related to Fig. 1
CAG- PreRNA-PPIB-original-hSyn-Cre	AAV1/2 mix	4.55E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL7-hSyn-Cre	AAV1/2 mix	1.30E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL28-hSyn-Cre	AAV1/2 mix	3.66E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
CAG- PreRNA-PPIB-3'Trak2-hSyn-Cre	AAV1/2 mix	1.92E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
CAG- PreRNA-PPIB-3'Net1-hSyn-Cre	AAV1/2 mix	2.33E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
CAG- PreRNA-PPIB-5'RPL28-3'Trak2-hSyn-Cre	AAV1/2 mix	1.27E+11	For mouse brain injection to optimize PreRNA, related to Fig. S1D
hSyn-Cre-hSyn-PreRNA-library	AAV1/2 mix	1.35E+10	The set of SynBar viruses used for sequencing experiments, design shown in Fig. S2A
hSyn-PreSynBar-hSyn-PreRNA-library	AAV1/2 mix	1.22E+10	The set of SynBar viruses used for sequencing experiments, design shown in Fig. S2A
hSyn-Cre-hSyn-PostRNA barcode landing pad	AAV1/2 mix	2.95E+10	The set of SynBar viruses used for sequencing experiments, design shown in Fig. S2A
hSyn-DIO-PostSynBar-P2A-mCherry	AAV1/2 mix	4.32E+10	The set of SynBar viruses used for sequencing experiments, design shown in Fig. S2A
hSyn-DIO-axonTag-mGreenLantern-P22N21	AAV1/2 mix	1.74E+10	Axon-targeted GFP with RNA binding domain, design shown in Fig. S3A
DIO-dendriteTag-oScarlet3-P22N21	AAV1/2 mix	1.93E+10	Dendrite-targeted RFP with RNA binding domain, design shown in Fig. S3A

hSyn-PreSynBar-hSyn-PreRNA-PPIB	AAV1/2 mix	4.51E+09	For mouse brain injection for Basescope experiment shown in Fig. 4F
hSyn-Cre-W3SL-hSyn-PostRNA-dapB	AAV1/2 mix	1.97E+10	For mouse brain injection for Basescope experiment shown in Fig. 4F
sCAG-DIO-oScarlet3	AAV1	1.42E+11	Used for AAV1 tracing experiment, design shown in Fig. S12A