

# Ascertaining cells' synaptic connections and RNA expression simultaneously with barcoded rabies virus libraries

Received: 26 October 2021

Accepted: 21 October 2022

Published online: 16 November 2022

 Check for updates

Arpiar Saunders<sup>1,2,3</sup>✉, Kee Wui Huang<sup>4</sup>, Cassandra Vondrak<sup>1,2</sup>, Christina Hughes<sup>1,2</sup>, Karina Smolyar<sup>1,2</sup>, Harsha Sen<sup>1,2</sup>, Adrienne C. Philson<sup>4</sup>, James Nemesh<sup>1,2</sup>, Alec Wysoker<sup>1,2</sup>, Seva Kashin<sup>1,2</sup>, Bernardo L. Sabatini<sup>4</sup> & Steven A. McCarroll<sup>1,2</sup>✉

Brain function depends on synaptic connections between specific neuron types, yet systematic descriptions of synaptic networks and their molecular properties are not readily available. Here, we introduce SBARRO (Synaptic Barcode Analysis by Retrograde Rabies ReadOut), a method that uses single-cell RNA sequencing to reveal directional, monosynaptic relationships based on the paths of a barcoded rabies virus from its “starter” postsynaptic cell to that cell's presynaptic partners. Thousands of these partner relationships can be ascertained in a single experiment, alongside genome-wide RNAs. We use SBARRO to describe synaptic networks formed by diverse mouse brain cell types *in vitro*, finding that different cell types have presynaptic networks with differences in average size and cell type composition. Patterns of RNA expression suggest that functioning synapses are critical for rabies virus uptake. By tracking individual rabies clones across cells, SBARRO offers new opportunities to map the synaptic organization of neural circuits.

The mammalian brain contains hundreds of cell types that connect with one another through synapses into intricate, and mostly uncharacterized, neural circuits. Traditional approaches for measuring synaptic connections and networks—such as whole-cell electrophysiological recordings and anatomical reconstructions from electron microscopy—sample only a few cells or small tissue volumes, do not readily scale to many animals or genotypes, and do not ascertain the molecular type and state of each cell.

Recent advances in single-cell transcriptomic profiling have made identifying cells and cell types within complex tissue routine<sup>1–4</sup>. Together with engineered proteins and viruses, additional cell features such as protein expression<sup>1</sup>, developmental origin<sup>2,3</sup>, axonal projection patterns<sup>4</sup> and physical interactions<sup>5</sup> can be decoded from RNA data. In the nervous system, rabies virus spreads from cell-to-cell in a retrograde fashion, from a neuron's dendrites into the axons of its

presynaptic partners<sup>6</sup>. Prevailing models suggest such transmission events occur at synapses, likely due to the presence of viral entry receptors<sup>7</sup> and high rates of membrane turnover. While the synaptic phenomenology of rabies virus transmission has been used for decades to discover neural pathways<sup>8</sup>, inefficient conversion of plasmid DNA into infective RNA-containing particles has largely precluded using rabies and other lyssaviruses in genomic applications<sup>9</sup>.

Here, we introduce SBARRO (Synaptic Barcode Analysis by Retrograde Rabies ReadOut), which combines monosynaptic rabies virus tracing<sup>10</sup>, viral genomic barcoding and scRNA-seq to generate high-throughput descriptions of cell-type-resolved synaptic networks. In SBARRO, encapsidated rabies virus genomes are distinguished by unique, transcribed viral barcode sequences (VBCs), allowing thousands of independent monosynaptic networks to be reconstructed in

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. <sup>3</sup>Vollum Institute, Oregon Health & Science University, Portland, OR 97239, USA. <sup>4</sup>Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA. ✉ e-mail: [saundear@ohsu.edu](mailto:saundear@ohsu.edu); [mccarroll@genetics.med.harvard.edu](mailto:mccarroll@genetics.med.harvard.edu)

parallel by tracking paths of clonal infections which originate in individual postsynaptic starter cells and spread to those cells' presynaptic partners.

Monosynaptic rabies virus tracing differs from labeling synaptically connected brain cells with unmodified rabies virus in two key respects<sup>10</sup>. First, starter postsynaptic cells are uniquely labeled and functionalized to enable founder infections. Starter cell infection targeting occurs through the over-expression of a fluorescently-tagged exogenous receptor that selectively binds a non-native viral spike protein in the rabies particle membrane, conferring cell entry. SBARRO takes advantage of a widely-used version of this technique, in which the TVA receptor is fused to a red fluorescent protein<sup>11</sup> and rabies virus genomes encoding a green fluorescent protein are pseudotyped with the cognate EnvA coat protein. Second, to limit rabies transmission to cells directly synapsing on the postsynaptic starter cell, the rabies glycoprotein gene is absent from the genome and is instead over-expressed in *trans* in the starter cells. Thus only rabies virus particles budding off of starter cell membranes are competent for uptake<sup>12</sup>. Importantly, active rabies replication and transcription drives high levels of viral RNA expression.

Unlike traditional monosynaptic tracing, in which fluorescence microscopy is used to distinguish populations of starter and presynaptic cells, with SBARRO, critical features of monosynaptic networks are encoded in each cell's RNA and are read out at single-cell resolution using high-throughput scRNA-seq (Fig. 1a). Because individual founder infections clonally replicate in starter cells before spreading into each presynaptic network, the presence of shared VBCs can be used to reconstruct many monosynaptic networks in parallel by tracking the infectivity path of each viral clone. In sampling cellular RNAs alongside VBCs, our approach distinguishes postsynaptic from presynaptic cell identities (via RNAs encoding the TVA receptor) and cell types and molecular states (via thousands of host cell RNAs)<sup>5</sup>. In genomics, rabies virus has been previously used in RABID-seq<sup>5</sup>, in which rabies virus spread was used to infer putative direct physical contacts between glia, a previously unknown and uncharacterized type of rabies virus transmission.

Here we present a comprehensive experimental and analytical framework for using SBARRO to discover structural and molecular properties of cell-type-specific synaptic connectivity. We leverage dissociated mouse brain cells grown into synaptic networks *in vitro*, similar to systems used to identify synaptogenic molecules<sup>13,14</sup> and for which features of *in vivo* connectivity can remain<sup>15</sup>. We discover that, in this system, cells of different types have presynaptic networks with different average properties, including cell type composition and size. Finally, we discover that rabies virus spread associates with a molecular signature of synaptic maturation, suggesting that functioning synapses, in addition to entry receptors, are critical for uptake of rabies virus.

## Results

### Barcoding millions of rabies viral genomes

Synaptic tracing with barcoded rabies requires libraries of barcoded rabies virus particles that have high numbers of unique barcodes which are as uniform as possible in abundance, such that individual viral barcodes are introduced into no more than one starter cell (in an experiment) and can be used to define "unitary" clonal infectivity paths (uCIPs; Fig. 1b). Historically, successfully creating negative-stranded RNA viruses from cellular transfection of DNA plasmids expressing viral proteins and genomes occurs with very low probability (~1 in 10,000 correctly transfected cells<sup>9</sup>) and multiple prolonged stages of viral replication are necessary to achieve high viral titers. Inefficient rescue and disproportionate replication make generating complex rabies libraries challenging. To generate libraries encoding millions of barcodes, we developed molecular and computational methods to introduce, retain, and quantify

barcodes in DNA plasmids and in rescued RNA genomes (Fig. 1c, d; Supplementary Fig. 1, Supplementary Fig. 2 and Supplementary Data 1).

We first developed a PCR-based strategy to flexibly engineer bipartite barcodes, generated through combinatorial diversity, into circular DNA plasmids (Supplementary Fig. 1b), followed by transformation and plate-based growth conditions optimized to retain DNA plasmid barcode diversity (Supplementary Fig. 2b, c). We also created a rabies rescue system—achieving equivalent viral titers threefold faster than the current, field-standard protocol<sup>16</sup>—that minimized barcode loss and disproportionate amplification during viral replication (Supplementary Fig. 2d). To assess viral barcode diversity and distribution, we used single-molecule sequencing (Fig. 1e and Supplementary Fig. 1c), for which we developed analysis methods to identify and correct for PCR and sequencing mutations (Supplementary Fig. 2b and Methods).

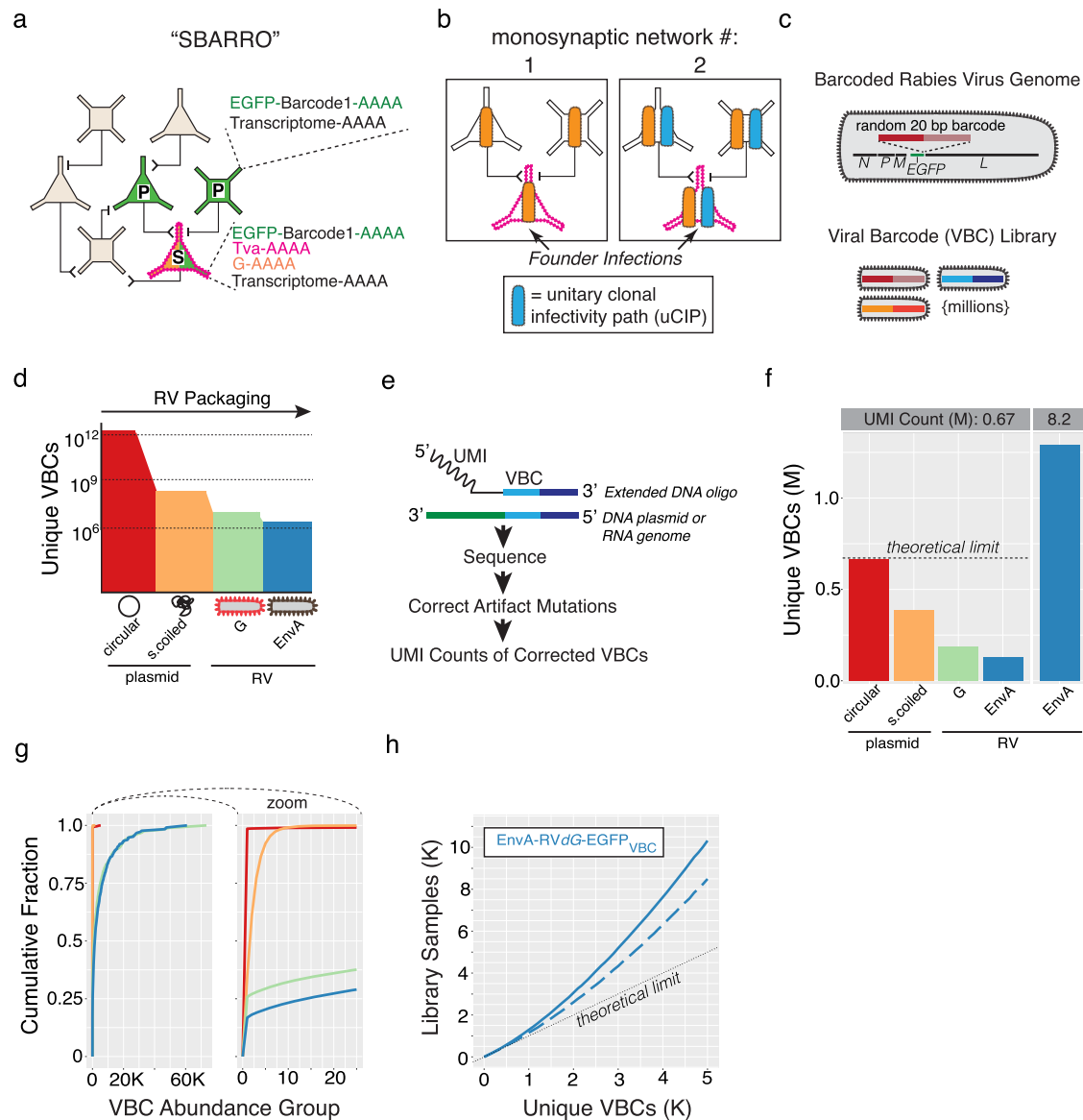
We used this approach to generate an EnvA-pseudotyped rabies library with a 20 bp randomer encoded in the 3' UTR of *EGFP* of SAD-dG-B19 (EnvA-RVdG-*EGFP*<sub>VBC</sub>; Supplementary Fig. 1a). We compared the total number of unique barcodes and their relative abundances across each production stage (Fig. 1d–g and Supplementary Fig. 1a). After PCR and circularization, nearly every sequenced plasmid contained a unique barcode. This diversity was reduced by bacterial amplification, though without substantially distorting representation of the retained barcodes. Rabies rescue induced barcode loss and abundance distortions and was mildly exacerbated by EnvA-pseudotyping. Deeper sequencing of the final EnvA-RVdG-*EGFP*<sub>VBC</sub> genomes (6.4 unique molecular identifiers (UMI) per viral barcode on average) quantified the relative abundances of 1.29 million unique, error-corrected barcodes (Fig. 1f).

To estimate the fraction of EnvA-RVdG-*EGFP*<sub>VBC</sub> founder infections that would be from viral particles with unique barcodes, we randomly sampled barcodes from the sequenced genomes of the infecting library and calculated the resulting number of unique barcodes (Fig. 1h). For these analyses, we used 50% unique barcodes as our benchmark, though the actual number of unique founder infections depends on properties of the infecting library and the number of founder infections in the experiment. Sampling up to ~8900 library genomes resulted in >50% unique barcodes; this could be increased to 15,500 library genomes by filtering out the 88 most abundant barcodes in the library, and to 83,600 library genomes by mixing nine equivalent libraries *in silico* (Supplementary Fig. 2e). These analyses suggested that our optimized protocols have helped overcome inefficiencies that previously limited rabies applications in scalable genomics research and suggested uCIPs can be efficiently generated from thousands of founder infections.

### Characterizing the barcoded rabies library

To directly determine the relationship between our barcoded EnvA-RVdG-*EGFP*<sub>VBC</sub> particles and the brain cells they infect, we infected three replicate cell cultures derived from embryonic mouse cortex. Infections were targeted to cells by recombinant adeno-associated virus (rAAV) expression of TVA; the host cells lacked the G protein necessary for rabies spread (Fig. 2a, b, Supplementary Data 2 and Methods). After 72 h, we collected 60,816 transcriptomes ( $n=6$  scRNA-seq libraries each from a single culture well) that captured both the cellular RNAs and the barcoded region of *EGFP* mRNAs (Supplementary Fig. 3 and Methods).

Naive analysis of barcode sequences from individual cells initially suggested large barcode "families" with many highly similar sequences (Fig. 2c). Reasoning that such relationships were largely created by PCR or sequencing errors, we developed an algorithm to collapse families of highly similar barcode sequences into the single barcode responsible for the putative founder infection (Methods). After collapse, barcode sequences associated with different inferred founder

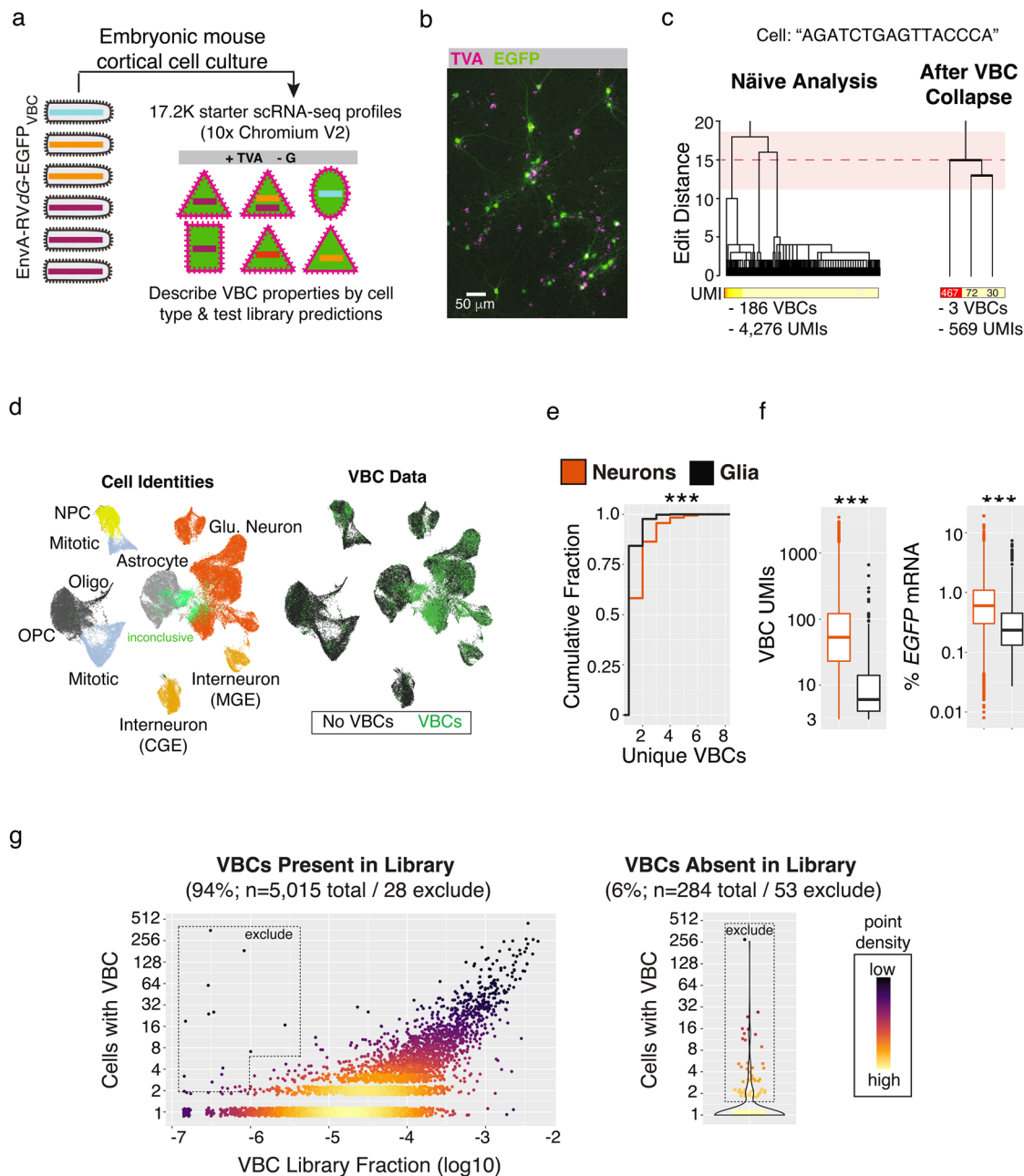


**Fig. 1 | Single-virion RNA tracking enabled by libraries of rabies virus particles encapsulating millions of uniquely barcoded genomes.** **a** Monosynaptic SBARRO. Shapes indicate neuron type. A TVA-expressing starter cell (“S”, magenta border) complemented in *trans* with rabies virus glycoprotein (G; orange text/fill) are selectively transduced by EnvA-RVdG-EGFP<sub>VBC</sub> (green text/fill) in which G is replaced by barcoded EGFP. G-complemented clonal particles spread a single retrograde synapse into presynaptic cells (“P”, black border). Uninfected tan cells make no starter cell synapses. Single-cell RNA profiles (dotted lines) inform (1) synaptic groupings (from viral barcode (VBC) sharing, green); (2) Starter or pre-synaptic status (from TVA mRNAs, magenta) and (3) host cell type (by capturing thousands of cellular mRNAs, black). **b** Monosynaptic relationships are inferred through unitary clonal infectivity paths (uCIPs). uCIPs are defined by VBCs sufficiently rare in the infecting rabies library to seed single founder infections in starter cells (magenta border). Example library-abundant (orange) or library-rare (blue) VBCs. Abundant VBCs infect multiple starter cells and thus conflate presynaptic networks one and two. Rare VBCs define a uCIP which corresponds to

monosynaptic network 2. **c** Viral particles are distinguished by a 20 bp VBC in the 3' UTR of EGFP. SBARRO libraries contain millions of uniquely barcoded genomes. **d** Schematic of VBC diversity during rabies packaging. **e** Sequencing-based VBC quantification using unique molecular identifiers (UMIs). **f–h** VBC diversity metrics (color-coded as in **d**). **f** Unique VBCs at each packaging stage after down-sampling to 0.67 million UMI counts (left) or with 8.2 million counts to more thoroughly characterize the final EnvA pseudotyped library (right). The dotted “theoretical limit” indicates a library in which every sampled VBC is unique. **g** Cumulative distribution of VBCs binned by “VBC abundance group” (AG) across packaging stages (0.67 M counts/stage). Counts for all VBCs sampled once belong to AG = 1; sampled twice belong to AG = 2, etc. **h** The relationship between unique VBCs and total VBC draws from the EnvA-RVdG-EGFP<sub>VBC</sub> library (8.2 M counts; blue line) and after removing the 88 most abundant VBCs (dashed blue line). The dotted line shows maximum diversity in which every VBC draw is unique. Source data are provided as a Source Data file.

infections in the same cells had the same distribution of similarity relationships (edit distances) as random barcodes did. Furthermore, there were (1) similar numbers of UMIs covering the barcode and EGFP transcript within the same cells and (2) independence in the number of unique barcodes and barcode-associated RNA counts (Supplementary Fig. 3a, b). All data presented hereafter have been computationally collapsed in this way.

Cells in which we detected at least one viral barcode also tended to have devoted a substantial fraction of their transcription to rabies genes (% of total UMIs mean  $\pm$  s.e.m.,  $15 \pm 0.14$ ), compared to cells that did not have a barcode ( $0.3 \pm 0.007\%$ , a rate consistent with background due to ambient cell-free RNA). This suggests that barcode ascertainment was sensitive, selective, and distinguished infected starter cells ( $n = 17,283$ ) from neighboring uninfected ( $n = 43,533$ ) cells



**Fig. 2 | Properties of barcoded library infection revealed through single-cell RNA profiling from mouse brain cultures lacking cell-to-cell viral spread.**

**a** Experimental schematic. EnvA-RVdG-EGFP<sub>VBC</sub> library transduced starter cells (+TVA) from which the rabies virus could replicate but not spread to other cells (-G). RNA profiles were captured ( $n = 60.8$  K cells), including from rabies-infected cells ( $n = 17.2$  K), allowing a direct analysis between how viral genomic barcodes are expressed in many single starter cells. **b** Representative image of dissociated mouse brain cell cultures (14 days in vitro) expressing TVA (magenta) and EGFP (green). TVA expression was induced by a cocktail of two high-titer rAAVs (CAG-Flex-TVA-mCherry: serotype 2-9,  $2.2 \times 10^{10}$  genomes/well; Syn1-EBFP-Cre: serotype 2-1,  $6 \times 10^9$  genomes/well). Cultures consist of TVA-/EGFP-, TVA+/EGFP- and TVA+/EGFP+ cells. **c** Inference of founder VBC sequences and accurate UMI-based counts from single cell RNA profiles in light of subsequent barcode mutations. A dendrograms illustrating VBC sequence relationships (top) and UMI counts (below) before (left) and after (right) "within-cell VBC collapse" for a single example RNA profile (Methods). The mean (red dotted line) and two standard deviations (pink shading) expected

from the distribution of edit distances across 20 base pair barcode sequences. **d** UMAP embedding of 60,816 scRNA profiles color-coded by molecular identity (left) or VBC ascertainment status (right) following LIGER analysis (Methods). A subset of infected scRNA profiles ( $n = 2635$ ) could not be definitively identified (light green). **e, f** Comparison of single cell VBC properties ascertained from RNA profiles of TVA+ neurons ( $n = 4222$ ) or glia ( $n = 914$ ; \*\*\* $p < 2.2e-16$ , two-sided Kolmogorov-Smirnov Test). Data from only the 1:10 EnvA-RVdG-EGFP<sub>VBC</sub> dilution (MOI - 1.5) are shown (Methods and Supplementary Fig. 3e). **e** Cumulative distribution of unique VBCs. **f** Total VBC UMIs (left) or % EGFP mRNA (right). **g** Critically evaluate the performance of the EnvA-RVdG-EGFP<sub>VBC</sub> through a corpus of 17.2 K starter cell RNA profiles. Left, for ascertained VBCs in the library (94%), the relationship between library abundance and the number of independent starter cell infections. Right, for library-absent VBCs (6%), the number of independent infections. VBCs observed in more starter cell RNA profiles than expected based on quantitative library abundance or library absence were flagged for exclusion (Methods). Source data are provided as a Source Data file.

(Supplementary Fig. 3c). Putatively infected cells (those cells for which >1% of total UMIs came from rabies genes) for which we failed to detect a barcode tended to have very small RNA profiles (<500 UMIs). In addition, <2% of cells with a barcode and a large RNA profile (>10,000 UMIs) lacked viral loads indicative of infection (<1% rabies RNA), suggesting spurious viral barcode associations were rare.

Using the above data, we investigated whether the properties of infection differed among starter cell types which we identified through a clustering analysis that integrated rabies-infected and neighboring uninfected cells (Fig. 2d and Methods). Consistent with our AAV-based strategy to express the TVA receptor (Supplementary Fig. 4a), we observed EnvA-pseudotyped rabies infections mainly in glutamatergic neurons, interneurons and neural precursor cells and less frequently observed in other glia types (astrocytes, polydendrocytes and oligodendrocytes) and cells undergoing mitosis (Fig. 2d). Downstream of infected TVA<sup>+</sup> cells, analysis revealed clear differences in infection properties: relative to infected glia, infected neurons tended to have more founder infections (unique barcodes mean  $\pm$  s.e.m.: Neurons,  $1.6 \pm 0.02$ ; Glia,  $1.2 \pm 0.02$ ), far more barcoded rabies transcripts detected per RNA profile (Neurons,  $156.5 \pm 4.0$ ; Glia,  $17.7 \pm 6.4$ ), and higher percentages of EGFP per RNA profile (Neurons,  $1.1 \pm 0.03$ ; Glia,  $0.5 \pm 0.04$ ) (Fig. 2e, f). These infection properties were unrelated to quantitative differences in TVA receptor expression (Supplementary Fig. 4b, c), suggesting they arose as a feature of intrinsic cell-type specific molecular differences downstream of the exogenous entry receptor following bath application of rabies virus.

In principle, the combination of multiple founder infections in the same starter cell could help define uCIPs through coupled presynaptic spread, but in practice, cell biological constraints might limit the number of founder infections. To evaluate this, we leveraged the viral barcodes to quantify the multiplicity of infection (MOI) at single-cell resolution and to relate this to the titer of the infecting library (Supplementary Fig. 3d, e). At the lowest titer we tried (MOI, -0.15), more than 97% of neuron and astrocyte RNA profiles were associated with a single VBC (unique VBCs mean  $\pm$  s.e.m. for neurons/astrocytes: MOI -0.15,  $1.07 \pm 0.009/1.03 \pm 0.01$ ). Infections with 10-fold higher titer resulted in more multiply infected cells (with two or more viral barcodes) (MOI -1.5,  $1.42 \pm 0.01/1.14 \pm 0.01$ ). However, we saw only minimal further increases at 100-fold higher titer (MOI -15,  $1.6 \pm 0.01/1.2 \pm 0.02$ ). At a biological level, these data suggest an intrinsic asymptote in the number of founder infections individual cells will meaningfully sustain –perhaps, for example, because cell-biological machinery are effectively hijacked by the earliest founders. At an engineering level, these results also suggested rabies virus titer-ranges for efficiently transducing starter cells with multiple viral barcodes.

Infecting and analyzing large numbers of starter cells in these control, no-spread experiments helped us to better understand many properties of rabies virus infections and barcoded rabies libraries. We compared the EnvA-RVdG-EGFP<sub>VBC</sub> library abundances of 1.29 million barcodes to their 28,755 founder infections distributed across 17,283 starter cells (Fig. 2g). Critically for later inferences, the abundance of a barcode in the infecting library predicted the number of cells it would infect (Fig. 2g). (A few barcodes that appeared to over-perform this expectation were flagged for computational removal from future analyses, Fig. 2g). In addition, some 6% of infections involved viral barcodes that we had not detected by sequencing the library, presumably because they were present at very low abundance (Fig. 2g) (any of these that infected multiple cells in this “no-spread” experiment were also flagged for removal from future analyses). Intriguingly, a small number of barcode pairs consistently appeared together in the same starter cells, even in distinct experiments (Supplementary Fig. 3f). Because rabies particles do not have strict genome size limitations<sup>9,17</sup>, we reasoned that barcode interdependence might result from concatenated genomes. (These pairs were also flagged and removed from future analyses; Methods). These analyses suggest the

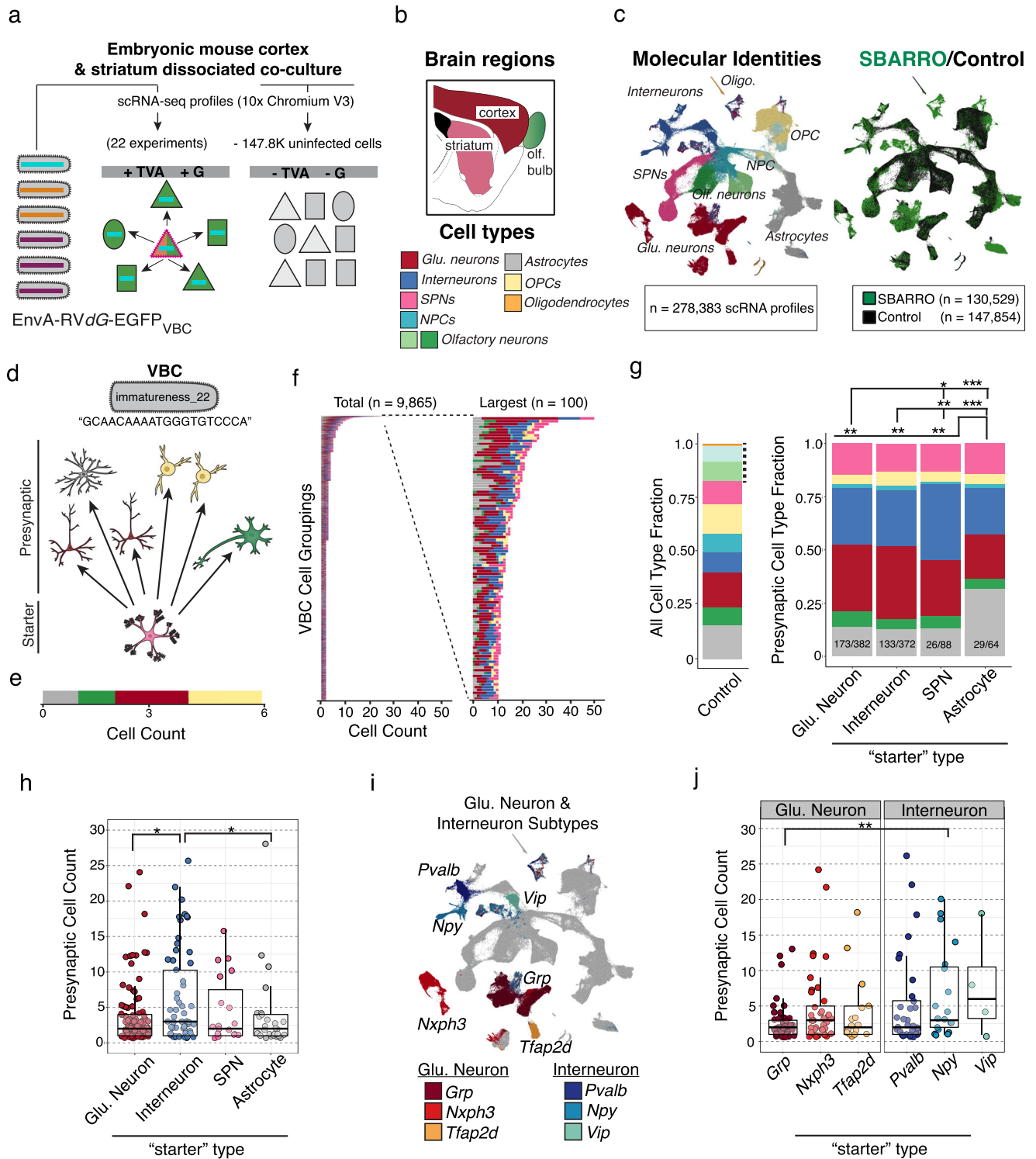
abundance of barcodes in library genomes has considerable predictive power in estimating the number of starter cell founder infections, but also highlight examples in which individual barcodes or barcode pairs defy expectations. Thus, each SBARRO infecting library should be carefully evaluated in a large number of starter cells, as we describe further below.

### High-throughput reconstruction of monosynaptic relationships

We next sought to describe cell-type-specific synaptic wiring of an in vitro culture. We focused on in vitro experiments because such systems (1) have been used extensively to screen for genes and molecules involved in synapse development; (2) can retain features of cell-type-specific connectivity; and (3) have more easily recoverable cells. To increase cell type diversity, we co-cultured cells dissociated from embryonic cortex, striatum and caudal olfactory areas. We sparsely seeded potential starter cells in each culture well by using the Syn1-EBFP-Cre rAAV at low titer to recombine rAAVs encoding TVA (CAG-Flex-TVA-mCherry) and the rabies glycoprotein (G; CAG-Flex-B19G) at high-titer (Methods), thus enabling efficient EnvA-mediated rabies founder infections and G-dependent presynaptic spread. (Sparsity minimizes the opportunity for starter cells to become secondarily infected as presynaptic cells, which could in principle support polysynaptic spread). Lastly, after 12 days in vitro, during a period of prolific synaptogenesis<sup>18</sup>, starter cells were transduced with EnvA-RVdG-EGFP<sub>VBC</sub> (MOI, -1.5) (Fig. 3a and Methods). After another 72 h, EGFP fluorescence was observed in putative presynaptic cells, many of which were spatially clustered around each dual-labeled starter cell and were interspersed with large numbers of uninfected cells (Supplementary Fig. 5a, b). This spatial pattern of rabies spread was consistent with the idea that the probability of neuronal connectivity scales roughly with spatial proximity and suggested that presynaptic networks innervating distinct starter cells were largely non-overlapping. No transduction (EGFP<sup>+</sup> cells) was observed without TVA receptor expression, suggesting that all infections entered experiments through starter cells.

We identified cell types in thousands of reconstructed monosynaptic networks, profiling RNA from EGFP<sup>+</sup> cells from 22 distinct culture wells ( $n = 3$  cell culture replicates; Supplementary Fig. 5d and Supplementary Data 2). To learn the molecular identities of each cell and determine in what ways the infected population might be different from the total ensemble of cultured cells, we co-clustered scRNA profiles from SBARRO experiments ( $n = 130,529$  cells; mean UMIs, 17,622) and uninfected control cells ( $n = 147,854$  cells; mean UMIs, 18,117) based on shared host-cell RNA signatures (Methods). Cultured cell RNA profiles were from diverse and developmentally dynamic cell populations (Fig. 3b, c). We identified four populations of glutamatergic neurons; polydendrocytes; oligodendrocytes; and neural precursor cells (NPCs) developing into astrocytes and several mature GABAergic lineages (including three major interneuron populations, two olfactory-related neuron types, and spiny projection neurons (SPNs); Supplementary Fig. 6a-c). Neuronal identity assignments were confirmed by an integrated analysis with scRNA profiles from adult mouse neocortex<sup>19</sup> (Supplementary Fig. 6d). Compared to the relative abundance of control cells, rabies-infected cells were enriched among mature interneurons ( $\log_2(\text{rv}/\text{control}) = 1.21$ ), SPNs (0.94), glutamatergic neurons (0.83) and astrocytes (0.32), and depleted from developmentally immature cells (NPCs = -2.45; immature neurons = -2.44), mature GABAergic olfactory types (-1.67), oligodendrocytes (-0.94) and polydendrocytes (-0.91) (Supplementary Fig. 6e).

We detected putative synaptic networks as clonal expansion of viral barcodes observed across cells (Supplementary Fig. 7a). Paired anatomical/SBARRO datasets suggests that roughly 10% of infected cells entered our single-cell analyses; the missing cells were likely lost or destroyed during physical dissociation and FACS-enrichment or remained unsampled after microfluidics-based RNA barcoding. Thus,



synaptic networks are detectable yet contain only a small subset of the cells associated with each network (Supplementary Fig. 5c–e).

We identified starter cells by their TVA expression (Supplementary Fig. 8a–d and Methods). Presynaptic and starter cells were composed of similar cell types, but starter cells expressed a larger number of unique barcodes (mean ± s.e.m.: starter, 4.3 ± 1; presynaptic, 2.8 ± 0.007,  $p < 2.2e-16$ , Kolmogorov–Smirnov Test; Supplementary Fig. 8e,f), which is expected as some barcodes may fail to transit and infect other cells in the analysis. Comparing FACS-based counts of fluorescently labeled starter or presynaptic cells suggested that starter cells failed to enter our analyses more frequently than presynaptic cells

did (2.5% vs 25%; Supplementary Fig. 5c–e); this could reflect increased fragility and loss due to prolonged infection, or insufficient ascertainment of recombined TVA mRNAs. Thus, we expect many of our identified synaptic networks to be “orphaned” from their starter cell.

We developed a statistical framework to filter one or more co-expressed viral barcodes based on the (1) estimated number of founder infections and the (2) barcode abundance in the infecting library (Methods). We also excluded barcodes ( $n = 551$ ) or barcode pairs ( $n = 689$ ) that (in the control experiments) infected or co-infected more cells than expected based on their library abundance (Fig. 2g; Supplementary Fig. 7b; Methods). For example, in an experiment

**Fig. 3 | Massively parallel inference of cell-type-specific synaptic connectivity using SBARRO.** **a** Experimental schematic. The EnvA-RVdG-EGFP<sub>VBC</sub> library transduced starter cells (+TVA/+G) from which individual virion clonally replicate and undergo monosynaptic spread into presynaptic cells. scRNA-seq libraries were prepared from either (1) SBARRO EGFP+ cells ( $n = 22$  culture wells from  $n = 3$  mouse preparations;  $n = 130.5$  K scRNA profiles) or (2) preparation-matched control cells ( $n = 147.6$  K scRNA profiles). **b** Sagittal brain schematic color-coded by region from which cells were co-cultured. **c** UMAP embedding of scRNA profiles color-coded and labeled by coarse molecular identity (left, Supplementary Fig. 6a) or SBARRO/control status (right) following LIGER analysis. **d** An example SBARRO network inferred through shared expression of the VBC assigned the name “immatureness\_22”, a barcoded genome with <1% chance infecting >1 starter cell. The “immatureness\_22” network consists of an SPN starter cell and heterogeneous collection of putative presynaptic cells (color-coded as in **c**). **e** Horizontal bar plot representing the cell types of the “immatureness\_22” presynaptic network. **f** Horizontal bar plots for  $n = 9865$  SBARRO networks with  $\geq 2$  cells (left) and the

largest 100 networks (right). Of all networks,  $n = 365$  networks (3.7%) included starter cell assignments. **g–i** Properties of presynaptic networks. **g** Fractional cell-type compositions from RV uninfected control wells (left) and presynaptic networks stratified by starter cell type (right). A subset of cell types observed in controls (dotted vertical line) were never or rarely observed in presynaptic networks and excluded from network analysis. Presynaptic cell type composition varies by starter cell type ( $*p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.001$ , Chi-Square Test). The number of aggregated networks and total presynaptic cells are shown (networks/total cells). Network data include  $n = 365$  networks identified across  $n = 22$  culture wells derived  $n = 3$  primary cell culture replicates. **h** Inferred presynaptic network sizes differ by starter cell type ( $**p < 0.05$ , two-sided Wilcoxon Test). **i** UMAP embedding color-coded and labeled by glutamatergic neuron and interneuron subtype (Supplementary Fig. 6a, c). **j** Inferred presynaptic network sizes by starter cell subtype ( $**p < 0.01$ ; two-sided Wilcoxon Test). Boxes define the interquartile range and whiskers delineate 1.5 times this range. Source data are provided as a Source Data file.

estimated to contain 2484 total founder infections, we observed one example barcode in an SPN starter cell and six diverse presynaptic cells (Fig. 3d and Supplementary Fig. 7c). Based on the low abundance of the barcode in the infecting library (frequency =  $3.5 \times 10^{-6}$ ), we estimate that this barcode had a <1% chance of participating in more than one founder infection in this experiment. Thus this barcode passed our threshold (of <10%) and defined a uCIP (Methods).

We retained  $n = 1810$  of 5142 total viral barcodes, which alone (27%) or as a pair (73%), enabled  $n = 9865$  non-redundant uCIP inferences of synaptic networks with  $\geq 2$  cells ( $n = 21,458$  EGFP + scRNA profiles; Fig. 3f). Inferred networks contained 2–52 cells (mean = 3, median = 2), consistent with ~10% ascertainment of rabies-infected cells in culture (Supplementary Fig. 5c, d). Inferred networks contained predominantly neurons (79%) of diverse types, with smaller contributions from astrocytes (15%) and polydendrocytes (5%). A minority of these networks contained a clear starter cell ( $n = 365$  or 3.7%), as identified by a rabies infected cell with statistically significant enrichment of Cre-recombined TVA transcripts (Supplementary Fig. 8). In each of the remaining networks ( $n = 9500$  or 96.3%), a starter cell was not identified because starter cells were (1) destroyed during cell dissociation and FACS; (2) not sampled during scRNA-seq library preparation; or (3) lacked sufficient evidence for Cre-recombined TVA receptor expression.

To determine whether the cell-type composition or the size of presynaptic networks varied across postsynaptic cell types, we focused on the 3.7% of networks with an identified starter cell ( $n = 365$  of 9865 total networks). Presynaptic cell types differed according to postsynaptic cell type, as pair-wise comparisons suggested quantitative differences in presynaptic cell-type proportions between astrocytes versus neurons ( $p = 0.007$ – $0.001$ , Chi-Square Test) and glutamatergic neurons and interneurons versus SPNs ( $p = 0.01$  and  $0.0002$ ), but not between glutamatergic neurons and interneurons ( $p > 0.05$ , Chi-Square Test; Fig. 3g).

The sizes of inferred presynaptic networks exhibited variance that was partially explained by postsynaptic cell type ( $p = 0.048$ , Kruskal–Wallis Test). Pair-wise comparisons revealed that glutamatergic neurons and astrocytes tended to have smaller presynaptic networks than interneurons did ( $p = 0.018$ – $0.022$ , Wilcoxon Test), while SPN presynaptic networks did not detectably differ from those of other cell types ( $n = 109$  glutamatergic neurons, mean  $\pm$  s.e.m. presynaptic cells =  $3.6 \pm 0.4$ ;  $n = 29$  astrocytes,  $3.7 \pm 1.02$ ;  $n = 61$  interneurons,  $6.7 \pm 1.0$ ;  $n = 12$  SPNs,  $4.6 \pm 1.1$ ; Fig. 3h). Differences in neuronal presynaptic network size appeared to be driven in part by cell subtypes (Fig. 3i, j and Supplementary Fig. 7d, e); while neuronal subtype categories did not, as a whole, rise to predictive significance in explaining variance in presynaptic network size ( $p = 0.08$ , Kruskal–Wallis test), paired comparisons revealed that *Grp* + glutamatergic neurons and *Npy* + interneurons tended to have small

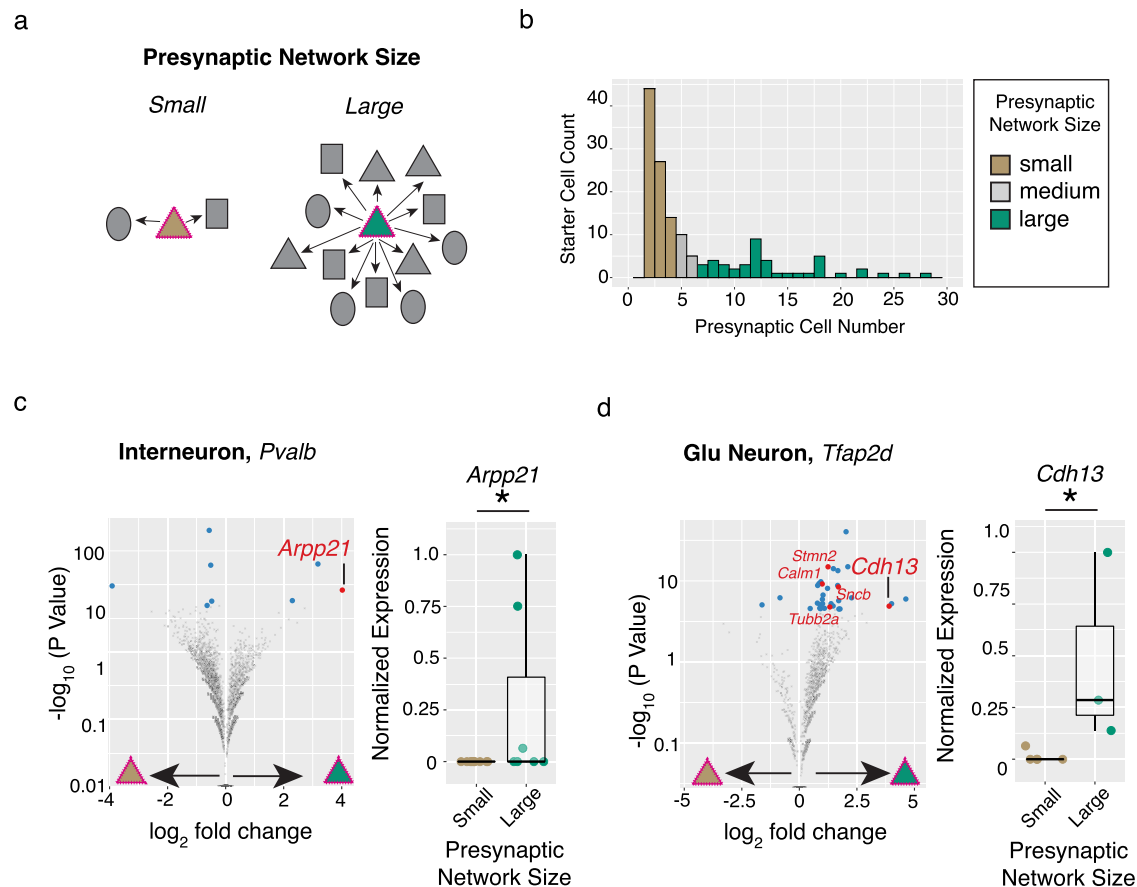
( $n = 50$ , mean  $\pm$  s.e.m. =  $2.7 \pm 0.37$  cells) and large ( $n = 20$ , mean =  $8 \pm 1.95$  cells) presynaptic networks ( $p = 0.004$ , Wilcoxon Test), respectively. These results indicate that the number and molecular composition of putative presynaptic cells in an inferred network are qualitatively similar across postsynaptic cell types at an early, promiscuous stage of synaptogenesis in vitro, but highlight important exceptions in which postsynaptic cell type biases the number and classes of putative presynaptic partner cells. Differences in the number of presynaptic partner cells might relate to dendritic size differences in vivo. For example, compared to adult mouse neocortex, *Grp* + glutamatergic neurons are most similar to L2/3 IT and L4/5 IT subtypes found in superficial cortical layers, which tend to have small dendritic arbors, while *Nxph3* + glutamatergic neurons are most similar to L6b, L5 NP and L6 CT subtypes found in deeper cortical layers, which tend to have larger dendritic arbors<sup>19,20</sup> (Supplementary Fig. 6d).

### Postsynaptic RNAs associated with presynaptic networks

The formation and selective stabilization of synapses is shaped by competitive processes driven by molecular variation within<sup>21</sup> and across cell populations<sup>22</sup>, yet many of the molecules remain unknown and incompletely understood. We sought to use the data from these experiments—in which synaptic connectivity inferences and molecular properties were measured in the same cells—to analyze how molecular variation associated with the properties of cell-type-specific networks.

We first asked whether presynaptic network size was explained by infection magnitude or innate immune response in starter cells, since these properties of infection could skew the results (Fig. 4a). We separated starter cells into two groups based on presynaptic network size, ranging from networks of 2–4 ascertained cells (“small”) or 7–52 ascertained cells (“large”) and four groups based on starter cell type (Fig. 4b). We compared both the viral load (the fraction of total cellular mRNAs from the rabies virus genome) and an aggregate innate immunity expression score ( $n = 564$  genes<sup>23</sup>) across these groups ( $n = 144$  scRNA profiles; Supplementary Fig. 9a). We found that, while both infection metrics varied by starter cell type (viral load,  $p = 0.05$ ; innate immune expression score,  $p = 5.2 \times 10^{-12}$ , two-way ANOVA Test), they did not associate with the presynaptic network size (viral load,  $p = 0.10$ ; innate immune expression score,  $p = 0.16$ ).

We next sought to find genes whose expression levels in starter cells associated with the number of inferred presynaptic partner cells (Fig. 3j). We hypothesized that differences in the expression of genes promoting or restricting synaptogenesis or dendrite growth influence the number of presynaptic cells innervating each starter cell. We focused our comparisons at the most granular cell subtype level and narrowed our testing to those genes sufficiently expressed and skewed in aggregate across presynaptic network size groups (Methods). We



**Fig. 4 | Postsynaptic RNA levels associated with rabies-based inferences of presynaptic network size.** **a** Schematic of postsynaptic starter cells with small (brown triangle) or large (green triangle) numbers of presynaptic partner cells. **b** Histogram of inferred presynaptic network sizes for  $n = 144$  starter cell RNA profiles belonging to one of four major cell types (glutamatergic neurons, interneurons, SPNs or astrocytes). **c, d** Differential expression testing identifies *Arpp21* upregulated in *Pvalb* interneurons ( $n = 9$  small versus  $n = 7$  large RNA profiles) and *Cdh13* as *Tfp2d* Glutamatergic neurons ( $n = 5$  small versus  $n = 3$  large RNA profiles) starter cells with large presynaptic networks. Left, Volcano plots illustrating results

from differential expression testing of starter cell subtypes in which UMI counts were aggregated by inferred presynaptic network size category (Fisher's Exact Test; Methods). Genes passing corrected  $p$  value thresholds ( $p$  corrected  $< 0.05$ , blue dots) were further tested for differences in single-cell scaled expression (two-sided Wilcoxon Test; Methods) and those that pass this additional test ( $p$  uncorrected  $< 0.05$ , red dots) are labeled. Right, normalized expression levels ( $p < 0.05$ , *Arpp21*,  $p = 0.044$ ; *Cdh13*,  $p = 0.026$ . Two-sided Wilcoxon Test). Boxes define the interquartile range and whiskers delineate 1.5 times this range. Source data are provided as a Source Data file.

used permutation to create negative-control distributions in which each starter cell RNA profile was replaced by a randomly selected presynaptic RNA profile of the same type.

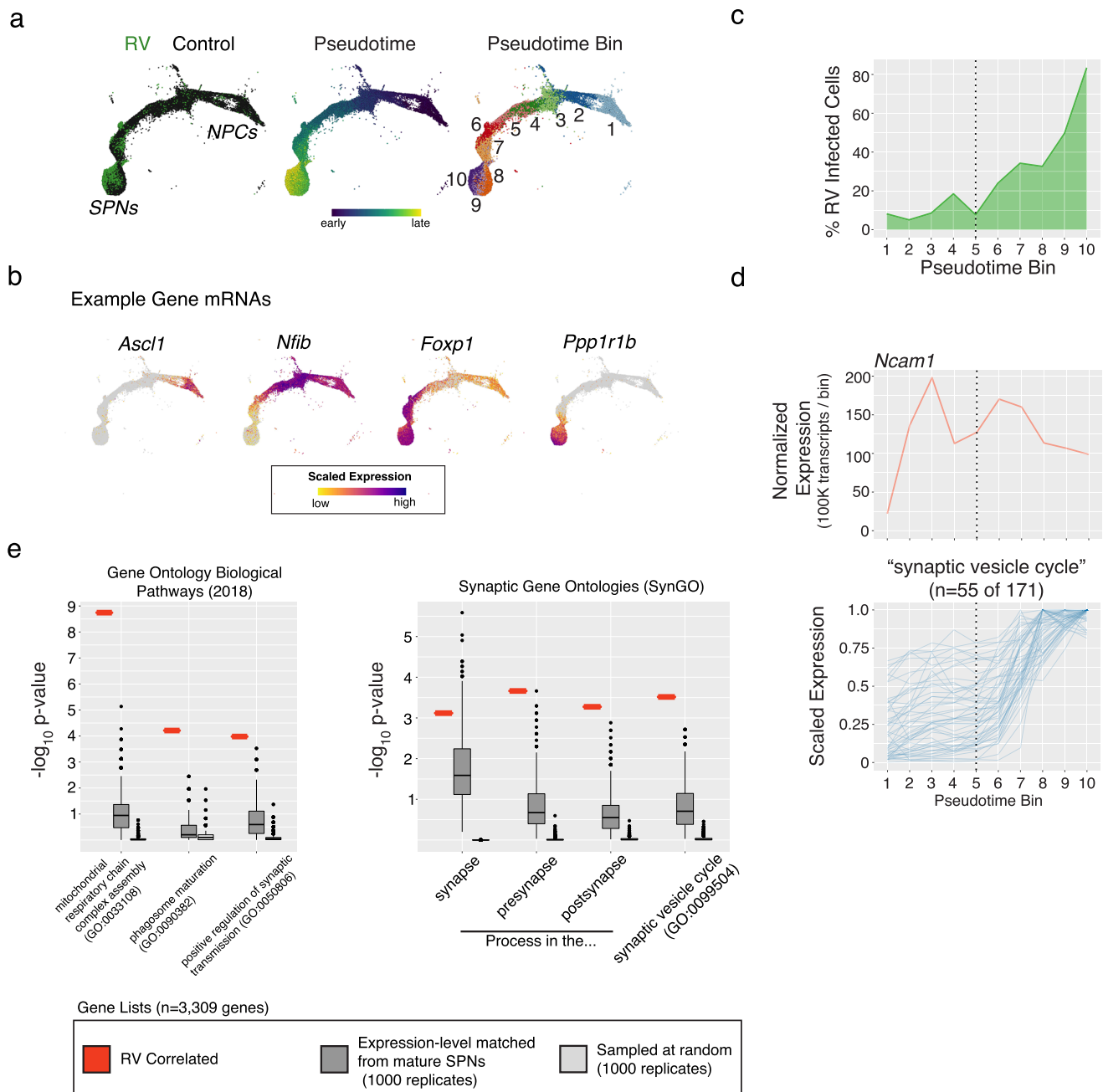
Across eight starter cell subtypes, 13 genes exhibited differential expression across presynaptic network size categories ( $p < 0.05$ , Wilcoxon Test; Fig. 4c; Supplementary Fig. 9b, c). Though this did not exceed the number of genes nominated in permuted analyses (mean  $\pm$  s.e.m. =  $15.6 \pm 1.3$ ), independent biological evidence strongly supported roles for two of the most strongly differentially expressed genes, both of which were more highly expressed in starter cells with large networks (relative to cells with small networks) and have described roles in promoting dendritic growth or synapse formation through developmental loss-of-function or ectopic overexpression experiments. *Arpp21*—which was upregulated in postsynaptic *Pvalb* + interneurons with large presynaptic networks—encodes an RNA binding protein that promotes dendritic growth by activating translation of target RNAs and whose cell-to-cell dynamic range might be extended due to intronic-encoded inhibitory microRNA<sup>24</sup>. *Cdh13*—which was upregulated in postsynaptic *Tfp2d* + glutamatergic neurons with large presynaptic networks—encodes an atypical protocadherin, one of four genes previously identified as driving synaptogenesis in a large-scale neuronal RNAi screen<sup>13</sup> (Fig. 4c). Interestingly, both *Arpp21* and *Cdh13* RNAs increase in expression as SPNs (Supplementary Fig. 9d) and

cortical neurons<sup>25</sup> mature, suggesting that elevated expression in “starter” neurons with large presynaptic networks might be a proxy for a more-advanced developmental state. Thus, while these results are suggestive, future analyses with cell-type-specific functional manipulations would be needed to determine whether these genes have causal roles.

### Rabies transmission correlates with synaptic maturation

Accurate interpretation of how rabies-inferred synaptic networks relate to actual synaptic connectivity and function is critically limited by our incomplete understanding of the molecules and cellular processes through which rabies enters, exits, and interacts with diverse host brain cell types. To determine which RNAs and biological pathways contribute to rabies transmission, we leveraged asynchronous development and variable rabies transmission in cultured cells to identify gene expression patterns that correlated with increased infectivity along the developmental trajectory stretching from neural precursor cells into mature SPNs (Fig. 5a, b). We strictly ordered each of the 32,503 scRNA profiles in pseudotime using Monocle<sup>3</sup><sup>26</sup> and confirmed the expected developmental processes through Gene Ontology Biological Pathway (GOBP)<sup>27,28</sup> enrichment analysis of co-regulated genes ( $\sim 25\%$  of the coding genome;  $n = 7844$  genes; Supplementary Fig. 10a).





**Fig. 5 | The developmental emergence of rabies transmission co-occurs with the maturation of synaptic function.** **a** UMAP embedding of scRNA profiles ( $n = 32,503$ ) along a trajectory of development from immature neural precursor cells (NPCs) to mature spiny projection neurons (SPNs). Left, color-coded by rabies virus infected (i.e. SBARRO;  $n = 8837$  profiles) or uninfected control cells from paired cultures ( $n = 23,666$  RNA profiles); Middle, pseudotime; Right, pseudotime bins ( $n = 10$ ). **b** Example expression plots for four developmentally regulated genes. **c** For each pseudotime bin, the percentage of scRNA profiles corresponding to rabies virus infected SBARRO cells over the total number of all cells. **d** RNA levels across pseudotime bins. Of four described rabies receptors<sup>8</sup>, *Ncam1* (top) has the only appreciable expression. *Ncam1* expression precedes the major developmental increase in rabies transmission. The subset of genes ( $n = 55$  of 171) in the “synaptic vesicle” SynGO category (bottom) whose RNA levels correlate with rabies virus infectivity ( $n = 3309$  genes total). **e** Gene Ontology Biological Pathways (GOBP) and

Synaptic Gene Ontology (SynGO) analyses for rabies-infectivity correlated genes ( $n = 3309$ ; Methods). Left, analyses were conducted on the correlated gene list (red) as well as two sets of control genes (each with 1000 replicates of  $n = 3309$  genes). In the “Expression-matched” set (dark gray), genes were selected from the mature SPN metacell (pseudotime bin = 10) in a manner that matched expression levels of the correlated genes. In the “Random” set (light gray), genes were selected at random from those for which expressed RNA was detected. Statistical testing was performed using one-sided Fisher’s exact test with false discovery rate correction using default parameters for GOBP and SynGO. Categories with adjusted  $p$ -values  $< 0.05$  were considered enriched. Uncorrected  $p$ -value distributions for  $n = 3$  GO-BP categories (middle) and  $n = 4$  SynGO categories (right) for which the rabies-infectivity correlated gene set was statistically enriched. Source data are provided as a Source Data file.

Using just the uninfected cells, we identified  $n = 3309$  genes with RNA levels that correlated ( $r > 0.75$ ) with developmental increases in rabies transmission (Fig. 5c–e and Methods). Interestingly, *Ncam1* mRNA was the only one of four described rabies receptors<sup>7,29,30</sup> with

appreciable expression in these experiments and appeared in cells before high rates of infectivity, suggesting NCAM1 protein alone is not sufficient for rabies transmission (Fig. 5d). To discover which cellular processes might endow infectivity, we performed GOBP with the gene

set we identified and compared the results to control gene sets sampled at random or from expression-matched mature SPN profiles. We identified selective enrichments in (1) “mitochondrial respiratory chain complex assembly”; (2) “phagosome maturation”; and (3) “positive regulation of synaptic transmission”, which were absent from control gene sets (Fig. 5e and Supplementary Fig. 10b). To refine which synaptic processes were implicated in infectivity, we queried synaptic gene ontologies<sup>31</sup> and identified selective enrichment for “synaptic vesicle cycle” (for which  $n=55$  of 171 genes were correlated with infectivity). This analysis suggests that, in addition to the expression of viral entry receptors, operational synaptic transmission is critical for inter-cell rabies transmission and nominates specific genes implicated in the onset of synaptic transmission and rabies entry.

## Discussion

Our understanding of how synaptic networks emerge during development and how they are regulated by genetic and biological programs will benefit from measurements of synaptic connections that are systematic, quantitative, and connected to detailed molecular profiles of individual cells. Comprehensive characterization of the synaptic organization of neural circuits is challenging with current electrophysiological and anatomical methods, due to the small sizes of synapses, expansive geometries of axons and dendrites, and lack of knowledge of the cell subtypes involved. Such limitations have tended to separate synaptic-network biology from other subfields of neuroscience that are adopting highly parallel approaches for characterizing molecular repertoires<sup>32–35</sup> or neural activity in many individual cells<sup>36</sup>.

Here we demonstrate that many individual synaptic networks can be reconstructed from scRNA-seq data, thus allowing direct connectivity relationships to be inferred across thousands of individual cells for which genome-wide RNA expression has also been ascertained. Our data suggest that, during synaptogenesis in vitro, connectivity is shaped by cell type in a quantitative rather than qualitative way. Individual starter cells had considerable variance in their number of presynaptic partners, which appeared unrelated to the degree of infection or innate immune response, but was partially explained by neuron type and gene expression patterns. We found that *Arpp21* and *Cdh13* had higher expression within starter cells with more presynaptic partners. Interestingly, *Cdh13*—an atypical transmembrane protein of the Cadherin superfamily—was previously identified as a key postsynaptic gene driving both excitatory and inhibitory synaptogenesis through a systematic RNAi screen<sup>13</sup>. In addition, *Arpp21* overexpression or knock-out bidirectionally controls the size and complexity of pyramidal neuron dendrites during postnatal development, likely by potentiating the translation of bound mRNA species that promote dendritogenesis<sup>24</sup>. These proof-of-concept observations suggest that extant molecular heterogeneity may associate with different properties of a given cell’s presynaptic network and that SBARRO analyses are a means to access and quantify such relationships.

We designed SBARRO to be adaptable to emerging single-cell genomic technologies. For example, methods enabling single-cell spatial transcriptomics<sup>37–39</sup> or in situ sequencing<sup>40,41</sup> will allow the locations and anatomical properties of SBARRO cells to be mapped in vivo without cell loss. Moreover, long-read RNA isoform sequencing<sup>42,43</sup> could address long-standing hypotheses for how alternative splicing helps generate an extracellular adhesion code between synaptically connected cells within and across cell types<sup>44</sup>.

Unknown features of rabies cell biology represent a current limitation in the interpretation of SBARRO datasets. A detailed understanding of how our inferred, digital monosynaptic relationships relate to extraordinarily diverse and highly dynamic synaptic structures requires a comprehensive description of how rabies interacts with and transits between host brain cells of different types. On one hand, previous studies provide direct and circumstantial evidence that

suggest, at least to a first approximation, that rabies transmission events are selective for synapses made directly onto infected neurons<sup>45,46</sup>. Among postsynaptic neurons of the same class with spatially intermixed dendrites, presynaptic labeling respects synapse-selective motor arcs in the spinal cord<sup>47</sup>. Similarly, in primary visual cortex, intermingled layer 2/3 glutamatergic neurons distinguished by firing properties to visual cues, appear to inherit those selective properties from presynaptic cells labeled by rabies infection<sup>48,49</sup>. On the other hand, the efficiency of rabies transmission can be very low for certain cell-type-specific axons<sup>50</sup> and appears to be modulated by presynaptic firing rate<sup>51,52</sup>. Our correlative molecular data suggest developmentally mature presynaptic function is critical for rabies uptake in neurons (Fig. 5). However, the extent to which rabies egress and entry exclusively use synapse-associated processes; occur through direct synaptic contacts; and are affected by neural activity across diverse brain circuits, all remain to be firmly established<sup>52</sup>. Moreover, it will be necessary to continue to study the ways in which infection alters host cells’ molecular programs, as these alterations could affect synapse-associated processes<sup>53–56</sup>.

Rabies infection of non-neuronal cell classes, such as astrocytes and polydendrocytes, is a minor yet clear feature of our in vitro experiments and is also observed in vivo<sup>57,58</sup>. While both cell types interact intimately with synapses, especially during development, more experiments are necessary to understand the molecular mechanisms underlying rabies transmission across non-neuronal classes. Single-cell, single-virion inferences of these interactions may offer valuable insight: RABID-seq<sup>5</sup> analyses suggest rabies can be transmitted from infected astrocytes into presumed physically adjacent microglia and that specialized host cell signatures are associated and detectable with these interactions. What role synapses play in these glial interactions is not clear (Microglia were not present in our experiments in vitro).

SBARRO complements and extends other virus-based systems for studying cell-type-specific axonal projections and cell population-based synaptic connectivity. For example, MAPseq<sup>40</sup> and VECTORseq<sup>59</sup> enable anterograde and retrograde labeling of axonal projections, respectively, in combination with transcriptional analyses of individual cells. Herpes virus-based Connect-seq<sup>60</sup> enables putative presynaptic networks (which are not strictly monosynaptic) to be identified upstream of individual Cre-expression postsynaptic cell populations (but not individual cells). Similarly, Trans-seq<sup>61</sup> enables a non-viral anterograde labeling of cell populations in direct synaptic partnership that can be used in conjunction with scRNA-seq. By combining rabies barcoding with scRNA-seq and custom computational analyses, SBARRO enables many individual monosynaptic networks to be sequenced in parallel from individual experiments and associated with the transcriptional signatures of each cell in the network.

Mammalian synaptogenesis is particularly challenging to study with traditional methods due to the many cell types and molecules involved, its protracted nature in space and time, and intrinsic noise that arises from being a competitive, cell-to-cell process. By facilitating connectivity inferences and RNA sampling from the same individual cells, we hope that fast, scalable, all-molecular approaches such as SBARRO—which may be eventually deployed in non-destructive ways<sup>62</sup>—can complement established connectomic technologies based on super-resolution imaging of synaptic anatomy and lead into new insights into the molecular specification and maintenance of selective synaptic relationships.

## Methods

### Barcoding rabies virus plasmids and RNA genomes

Rabies virus rescue encapsidates RNA genomes from DNA templates. Generating rabies virus particle libraries with millions of unique and similarly abundant genomic barcodes presents a two-part challenge not encountered when rescuing a single genomic species. First, a

plasmid library carrying hyper-diverse barcoded DNA genomes is created *ab initio*. Second, barcode loss and abundance skews must be minimized during plasmid amplification (in bacteria) then in rabies virus rescue and replication (in mammalian cell culture). To address these challenges, custom protocols were developed to (1) introduce barcode sequences into DNA plasmids using PCR (achieving near-theoretical levels of plasmid-to-plasmid barcode diversity; Fig. 1g and Supplementary Fig. 2a); (2) more uniformly amplify plasmid DNA through optimized bacterial transformation and plate-based growth conditions (Supplementary Fig. 2c); and (3) rescue rabies virus (with native or pseudotyped coat proteins) in ways that mitigate distortions in barcode representation, initially created by the very low-probability of individual rescue events<sup>9</sup> and then exacerbated by biases in viral replication. Our 7–9 day protocol is three-fold faster and achieves titers equivalent or higher than the field-standard published protocol ( $1 \times 10^{8-9}$  IU/mL; Supplementary Fig. 2d)<sup>16,47</sup>. Details for each of the three protocol steps are found in the sections below. Barcodes present in DNA plasmids and RNA genomes were quantified through sequencing-based approaches in which oligonucleotide probes containing unique molecular identifier (UMI) sequences were hybridized to barcode-adjacent sequences and then polymerase-extended through the barcode region (Supplementary Fig. 1c and Supplementary Table 1); the resulting paired UMI-barcode sequences were used to count individual molecules. Inflation of barcode sequences and UMI counts due to mutations arising during library amplification and Illumina sequencing were accounted for (see description in Results) using a custom algorithm for post-hoc mutation correction. See the “Quantifying barcodes from plasmids and rabies virus genome” section below for details.

**PCR-based plasmid barcoding.** To generate plasmid libraries in which individual circular plasmids encode unique barcode sequences, we developed a PCR-based molecular workflow in which a bipartite barcode cassette can be targeted to arbitrary regions of a non-barcoded plasmid template (Supplementary Fig. 1b). We applied our system to the SAD-B19 genome plasmid in which the G gene has been replaced by *EGFP* (cSPBN-4GFP, Addgene #52487<sup>16</sup>), targeting the barcode cassette to the 3' UTR of *EGFP* adjacent to the viral polyadenylation sequence<sup>63</sup>. To introduce each half of the barcode cassette, whole-plasmid PCR was performed with forward and reverse primers targeting the desired region. Each primer contains 3' plasmid-complementary sequence followed by a 5' tail with 10 bps of random nucleotides further flanked by a restriction cassette which includes the *PluI* (“GGCGCC”) restriction site (pSPBN-GFP Barcoding, Forward Primer: B19\_barcode\_F, Reverse Primer: B19\_barcode\_R; Supplementary Table 1). During PCR (See “Barcoding PCR” protocol), each round of primer hybridization and extension introduces a unique barcode, resulting in a linear, double-stranded amplicon collection in which unique 10 bp barcodes have been introduced into the 5' terminus of each DNA strand. The desired ~14.5 kb amplicons were size-selected using standard low-gel agarose (Sigma-Aldrich, A9414) electrophoresis and cleaned (Zymo Research, Gel DNA Recovery Kit #D4001), then cleaned and concentrated to >200 ng/μl (Zymo Research, DNA Clean & Concentrator-25 #D4033). To efficiently circularize the amplicons using the barcode restriction cassette and to remove remaining template plasmid and unwanted linear products, we developed a series of enzymatic reactions that consecutively performed in the same tube, saving time and avoiding DNA damage and loss due to repeated purification (See “Plasmid Circularization Protocol”). Briefly, *DpnI* digest removes remaining methylated plasmid DNA; *PluI* restriction and T4 ligation circularize the amplicons thus covalently bonding each of two 10 bp barcodes into a 36 bp barcode cassette; and *RecBCD* selectively degrades linear DNA over circularized plasmid containing non-complementary barcode sequences, typically enriching the percentage of circularized product ~3.5 fold (from  $-20 \pm 0.7\%$  to  $71 \pm 4\%$ ,  $n = 6$

experiments,  $\pm$  denotes s.e.m.). “Barcoding PCR Protocol” (25 μl total): (1) Q5 High-Fidelity 2X Master Mix (NEB, #M0515), 12.5 μl; (2) Forward primer (B19\_barcode\_F), 1.25 μl; (3) Reverse primer (B19\_barcode\_R), 1.25 μl; (4) Template plasmid (cSPBN-4GFP, 0.4 ng/μl), 5; (5) Water (to 25 μl). Cycle Conditions: (1) 98 °C for 30 s; (2) 98 °C for 10 s; (3) 72 °C for 30 s; (4) 72 °C for 480 s; (5) Go to Step 2–4, 35 times; (6) 72 °C for 60 s. “Plasmid Circularization Protocol”, three steps. Step 1 = Digest (50 μl reaction): (1) *PluI* (10,000 U/mL, NEB R0713S), 0.5 μl; (2) *DpnI* (20,000 U/mL, NEB R0176S), 0.25 μl; (3) CutSmart Buffer (10x, NEB B7204S), 5 μl; (4) Linear Barcoded DNA, 500 ng; (5) Water (to 50 μl). Step 1 Cycle Conditions: (1) 37 °C for 1 h; (2) 80 °C for 20 min (heat inactivation). Step 2 = Ligation (spike-in, +5.1 μl): (1) T4 DNA Ligase (10,000 U/ml, NEB M0202T), 0.1 μl (200 U); ATP (10 mM, NEB P0756L), 5 μl. Step 2 Cycle Conditions: (1) 4 °C for 2 h; 2. 65 °C for 20 min (heat inactivation). Step 3 = Circular plasmid enrichment (spike-in, +8.8 μl): (1) *Exonuclease V* (*RecBCD*, 10,000 U/ml, NEB M0345L), 1 μl; (2) NEBuffer 4 (10x, NEB B7004S), 1.4 μl; (3) ATP (10 mM, NEB P0756L), 6.4 μl. Step 3 Cycle Conditions: 37 °C for 1 h; 2. 70 °C for 30 min (heat inactivation).

**Amplifying barcoded DNA plasmids.** Rabies virus rescue requires tens of micrograms of supercoiled rabies virus genome plasmid for cell transfection. To amplify and supercoil DNA plasmid libraries carrying hyper-diverse barcodes in a manner that minimizes loss and skew of barcoded plasmid representation, we transformed eight vials of chemically competent One Shot OmniMAX 2 T1<sup>8</sup> cells (ThermoFisher Scientific, C854003) in parallel each with 200 ng of DNA from the “Plasmid Circularization Protocol”. After 1 h of recovery growth, cultures were combined and 2 ml of the cell mixture was spread over  $n = 8$  large plates (24.5 × 24.5 cm; Corning, CLS431111<sup>64</sup>) containing LB Agar (Sigma-Aldrich, L2897) and Ampicillin (100 μg/mL; Sigma-Aldrich, A5354) and grown overnight at 37 °C. Colonies were scraped from each plate with 15 ml LB and pelleted through centrifugation (6000 × *g* for 15 min at 4 °C). Plasmids were isolated from cell pellets using the EndoFree Plasmid Maxi Kit (0.45 g cells/column; Qiagen, 12362). Sequencing-based barcode quantification (see below “Quantifying barcodes from plasmids and rabies virus genomes”) comparing plasmids prepared from pooled transformants grown on plates (as described) versus liquid culture (250 ml) demonstrated that plate-based growth dramatically reduced overrepresentation of plasmid barcodes (Supplementary Fig. 2c), presumably by homogenizing clonal growth rates.

**Rescuing barcoded rabies virus libraries.** De novo rescue of negative-stranded RNA viruses requires transfection-based encapsidation of positive-stranded RNA genomes with N, P and L proteins; the minimal replication-competent nucleocapsid<sup>65</sup>; Genomes lacking the G gene additionally require G protein such that replicating particles can spread cell-to-cell<sup>16</sup>. Two properties of rescue create challenges for generating particle libraries with millions of unique and uniformly abundant genomes. First, cells in which rescue events occur are rare (<1:10,000 transfected cells<sup>9</sup>), creating a limited number of cellular environments in which encapsidation can occur. Second, state-of-the-art packaging protocols serially infect fresh cultured cells to increase viral titer; increasing, with each passage, the opportunities for individual clones gain a replication advantage. To develop a rabies virus rescue protocol for barcoded genomes, we first systematically characterized how barcode abundances behaved after transfection and each passage stage of a widely used protocol<sup>16</sup>. We observed that (1) minimally and on average, hundreds of unique rabies virus genomes were encapsidated per encapsidation-competent cell and that (2) viral passages tended to reduce the number of unique barcoded genomes and distort their relative abundances (Supplementary Fig. 2d). Therefore, we increased the total number of encapsidation-competent cells by optimizing large-scale rabies virus transfection and created a one-

step rescue protocol capable of generating rabies virus libraries with millions of unique genomes at similarly high titers ( $\sim 2.5 \times 10^9$  IU/ml) but threefold faster (7–9 days) than published protocols<sup>16,47</sup>. Pseudotyping with non-native coat proteins requires an additional 6 days. Specifically, poly-L-lysine (Sigma-Aldrich, P4707) coated T-225 flasks containing 85–95% confluent HEK-293T/17 cells (ATCC, CRL-11268) were each transfected (Xfect, Takara #631318) with a DNA cocktail containing (1) the barcoded rabies virus plasmid library (131.36  $\mu$ g) and CAG-promoter driven plasmids for T7 polymerase (23.66 mg, Addgene 59926) and SAD-B19 helper proteins (N, 52.11  $\mu$ g, Addgene 59924; P, 30.15  $\mu$ g, Addgene 59925; L, 23.70  $\mu$ g, Addgene 59922; G, 20.26  $\mu$ g, Addgene 59921). Cells were maintained with DMEM with GlutaMAX supplement, pyruvate, high glucose media (Thermo Fisher Scientific, 10569010) supplemented with 5% fetal bovine serum (Thermo Fisher Scientific, 10082147) and 1x antibiotic-antimycotic (Thermo Fisher Scientific, 15240062) and incubated at 35 °C with 5% CO<sub>2</sub>. Five days post-transfection, culture media was collected for either (1) unpseudotyped rabies virus recovery or (2) EnvA pseudotyping. For EnvA pseudotyping, BHK-EnvA cells (Columbia Univ. Zuckerman Virus Core), initially grown in 15 cm dishes (Corning, 08-772-24) to 85–95% confluence, were infected with filtered media (0.22  $\mu$ m PES; Corning, 431097) from the transfected T-225 plates; cells are then rinsed, pelleted and re-plated first in 15 cm plates then again in T-225 flasks. Specifically, following 6 h of incubation with particle-containing media, cells from each plate are rinsed with two rounds of cold DPBS (+Ca, +Mg), trypsinized with 5 mL of trypsin-EDTA (Thermo Fisher Scientific, 25300-054) for 30 s at 35 °C, pelleted with centrifugation (300  $\times$  g, 4 min) in DMEM + 10% FBS and then re-plated in 15 cm plates and allowed to incubate overnight (–16–24 h) before being re-plated in T-225 flasks with DMEM + 5% FBS. T-225 plate media is supplemented with 3–5 mL of DMEM + 5% FBS each day for 4 days before being collected and concentrated. Specifically, collected media is incubated with benzonase nuclease (1:1000 dilution; Millipore Sigma, 70664) for 30 min at 37 °C and filtered (0.22  $\mu$ m PES). For ultracentrifugation (Beckman Coulter, SW32Ti rotor), 2 mL of 20% (w/v) sucrose in DPBS (–Ca, –Mg) is prepared in ultracentrifuge tubes (Beckman Coulter, 344058) to which the divided EnvA-pseudotyped viral media is added before pelleting (20,000 RPM for 2 h at 4 °C). Residual media is removed and viral pellets are each resuspended in 15  $\mu$ L of DPBS (–Ca, –Mg) on ice before orbital shaking at 4 °C for 8 h. Volumes are then combined, aliquoted, and stored at –80 °C. Titers were established by quantifying infected HEK-TVA and HEK-293T/17 cells in 12-well plates (80% confluence) using serial dilutions. To ensure EnvA-pseudotyping was complete, <2 HEK-293T/17 cells per well were tolerated following infection with 1  $\mu$ L with full-strength sample.

#### Quantifying barcodes from plasmids and rabies virus genomes.

UMI-based counting of barcodes from DNA plasmids and RNA genomes was accomplished with similar molecular (See “UMI-based counting of genome and plasmid barcodes” and Supplementary Fig. 2) and informatic workflows. RNA genomes were extracted using the ZR Viral RNA kit (Zymo Research, R1041) from particles ascertained from (1) end stage high-titer viral aliquots or (2) from cell culture media used for rabies virus rescue after PEG-based precipitation (Abcam, ab102538) and quantified using the High Sensitivity RNA ScreenTape assay (Agilent, 5067–5579). To count barcode abundances of individual RNA genomes or DNA plasmids, an oligonucleotide (B19\_UMI\_F) containing a SMRT PCR handle, 12 bp UMI and 33 bps of barcode-adjacent homologous sequence were hybridized then polymerase-extended through the barcode region. Remaining RNA genomes were selectively digested using RNase H (New England Biolabs, M0297S) and reactions were cleaned with Agencourt AMPure XP beads (1:1 volume; Beckman Coulter, A63881) retaining first-strand cDNA. The UMI-tagged genomic cDNA or plasmid DNA strands were then selectively amplified (14–18 PCR cycles; 16 median) using primers which

introduce the Illumina P5 (P5-TSO\_Hybrid) and indexed P7 (P7i-L5UTR\_seq) sequences. Amplicon libraries were sequenced on an Illumina MiSeq or NextSeq550 using a custom primer (Read1-CustomSeqB) to seed 110 Read 1 cycles. Base pairs (bp) 1–12 were assigned as the UMI. The two 10 bp viral barcodes were informatically extracted from the barcode cassette using a custom algorithms based on local sequence alignment algorithm and (“TagReadWithRabiesBarcodes” & “FilterValidRabiesBarcodes”). To account for artifactual barcode sequences created by mutations acquired during the library amplification and sequencing, we developed an algorithm to identify and collapse “families” of barcodes with similar sequences likely related through acquired mutations (“CollapseTagWithContext, MUTATIONAL\_COLLAPSE = true”). Specifically, after ordering barcodes most to least abundant, we considered each barcode as a “parent” and identified “siblings” sequences within Hamming distance of 1 of the “parent” barcode. The process was then iterated for each new “sibling” until no new “siblings” were discovered. The entire barcode family was assigned the sequence of the “parent.” UMI-parent barcode sequence pairs were then used to count each “parent” barcode in the library (after collapsing UMI-barcode sequences in which the UMIs associated with the same “parent” were Hamming distance  $\leq 1$ ). This approach drastically reduced the inflation of barcode sequences and counts due to library preparation and sequencing (Supplementary Fig. 2b). UMI-based counting of RNA genome and DNA plasmid barcodes. RNA Genomes, four steps. Step 1: RNA Genomes - UMI Hybridization (24  $\mu$ l reaction): (1) Rabies ssRNA genomes or dsDNA plasmid, 5–25 ng; (2) UMI oligo (10  $\mu$ M; B19\_UMI\_F), 4  $\mu$ l. (3) dNTPs (10 mM), 4  $\mu$ l; (4) Water (to 24  $\mu$ l). Cycle Conditions: (1) 72 °C for 4 min; (2) 4 °C. Step 2: RNA Genomes - Reverse Transcription (spike-in, +16  $\mu$ l): (1) 20% Ficoll PM 400 (Sigma Aldrich, 26873–85-8), 4  $\mu$ l; (2) NxGen RNase Inhibitor (Lucigen, 30281-2), 0.5  $\mu$ l; (3) Maxima H-RT (Thermo Scientific, EP0751), 2  $\mu$ l; 5x Maximal H-RT Buffer, 6  $\mu$ l; Water, 3.5  $\mu$ l. Cycle Conditions: 1. 42 °C for 90 min; 2. 85 °C 5 min (inactivation). Step 3: RNA Genomes - RNase H Treatment (spike-in, +2  $\mu$ l): RNase H (5000 U/mL, NEB M0297S), 2  $\mu$ l. Cycle Conditions: (1) 37 °C for 30 min; (2) 65 °C 20 min (inactivation). DNA Plasmids, one step: Step 1: DNA Plasmids - UMI Hybridization & polymerization (50  $\mu$ l reaction): (1) dsDNA plasmid (5–25 ng); (2) 2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602), 25  $\mu$ l; (3) UMI oligo (10  $\mu$ M; B19\_UMI\_F), 2  $\mu$ l; (4) Water (to 50  $\mu$ l). Cycle Conditions: (1) 98 °C for 3 min; (2) 68 °C 30 s; 3. 72 °C 20 s. Illumina Adaptor PCR (50  $\mu$ l reaction): (1) UMI-tagged DNA, 5  $\mu$ l; (2) P5 Primer (10  $\mu$ M; P5-TSO\_Hybrid), 1  $\mu$ l; (3) P7 Primer (10  $\mu$ M; P7i-L5UTR\_seq), 1  $\mu$ l; (4) 2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602), 25  $\mu$ l; Water (to 50  $\mu$ l). Cycle Conditions: (1) 98 °C for 30 s; (2) 98 °C for 10 s; (3) 72 °C for 30 s; (4) 72 °C for 60 s; (5) Go to Step 2–4, 13–17 times; (6) 72 °C for 120 s.

#### Synaptic cell culture

Cells were dissociated from the cortex or striatum of embryonic day 16 (E16) C57Blk6/N mouse brains and maintained for 14 days in vitro (DIV) following protocols approved by the Harvard Medical School (IACUC protocol number 05155). rAAVs were transduced on DIV 5 to functionalize “starter” cells. On DIV 12, EnvA-RVdG-EGFP<sub>VBC</sub> libraries were transduced and infection was allowed to proceed for 72–96 h before scRNA-seq libraries were generated. Pregnant C57Blk6/N dams (Charles River Laboratories) were housed in a barrier mouse facility (12 h light/dark cycle; room temperature with 40–60% humidity) before being heavily anesthetized by isoflurane inhalation, decapitated and the brains of embryonic male and female pups (litter size = 4–9) removed in ice-cold 1X Dissociation Media (“DM”; containing (in mM): 10.52 MgCl<sub>2</sub> (Sigma-Aldrich, M2393); 10.53 HEPES (Sigma-Aldrich, H3375); 1.32 Kynurenic Acid (Sigma-Aldrich, K3375) in HBSS (Thermo Fisher Scientific, 14175079)) in which cortex or striatum were dissected from each brain, pooled and incubated in sterile-filtered (0.22  $\mu$ m; Corning, 431097) DM + Papain/L-Cysteine (3.4 units

Papain and 0.172 mM L-Cysteine; Worthington Biochemical, LK003178) for 3–5 min at 37 °C. Brain tissue is then washed twice with 2–3 mL sterile-filtered DM + Trypsin Inhibitor (1 mg/mL; Sigma Aldrich, T9253) and incubated in the third wash for 3–5 min at 37 °C. DM + Trypsin Inhibitor is replaced with 6 mL of sterile-filtered “Plating Media” (“PM”; containing: DMEM (ATCC, 30-2002) and 10% FBS (ATCC, 30-2020)) in which digested brain volumes are titrated into single cells with a pipetteman equipped with a 5 mL pipet tip. Cell concentration was measured by diluting cells 1 to 5 in PM, mixed with an equal volume of 0.4% Trypan Blue (Thermo Fisher Scientific, 15250-061), and quantified using the Countess II Automated Cell Counter (Life Technologies). Each well of six-well cell culture treated plates coated with 0.1% Poly-L-ornithine (3 µg/mL; Sigma Aldrich, P4957) were seeded with ~750 K cells and maintained with sterile-filtered neurobasal medium (Thermo Fisher Scientific, 21103049), supplemented with serum-free B-27 (Thermo Fisher Scientific, 17504044), GlutaMAX (Invitrogen, 35050061) and Penicillin:Streptomycin (VWR, 45000-652). For imaging experiments, cells were seeded on glass coverslips (Fisher Scientific, 12-546) coated with 0.1% Poly-L-ornithine and Laminin (5 µg/mL; Thermo Fisher Scientific, 23017015). On DIV 5, a cocktail of three rAAVs was used to functionalize starter cells. Our starter cell strategy was designed to deliver consistent, high MOIs of rAAV per starter cell while flexibly controlling the number of starter cells in each culture through Cre delivery. Specifically, 1 µl CAG-Flex-TVA-mCherry (“TCB”; serotype, 2-9; titer,  $2.2 \times 10^{13}$  genomes/mL; MOI,  $\sim 2.9e^4$ ; UNC Vector Core) and 1 µl CAG-Flex-B19G (serotype, 2-9; titer,  $1.6 \times 10^{13}$  genomes/mL; MOI,  $\sim 2.1e^4$ ; UNC Vector Core) were added to each well along with 1 µl of Syn1-EBFP-Cre (serotype, 2-1; titer,  $6 \times 10^{12}$  genomes/mL; MOI,  $\sim 8-0.08$ ; Addgene, 51507-AAV1) delivered at full strength or diluted  $1:10^3$ – $10^4$ . At DIV 12, 1 µl of EnvA-RVdG-EGFP<sub>VBC</sub> library (titer,  $0.19$ – $1.1 \times 10^{10}$  IU/mL; Total cell MOI, 2.5–14.8) was added to each well. Epifluorescence imaging was used to monitor the progress of infections, including starter cell locations and morphology (based on TVA-mCherry fluorescence) as well as rabies virus transduction and spread from starter cells (based EGFP fluorescence). EnvA-RVdG-EGFP<sub>VBC</sub> transduction was completely dependent on the TVA receptor, since no EGFP fluorescence was observed in equivalent experiments in which the CAG-Flex-TVA-mCherry rAAV was excluded. To prepare cultures grown on coverslips, neurobasal medium was removed and each culture well was rinsed three times with 1× PBS and then fixed with fresh 4% paraformaldehyde at room temperature for 30 min, followed by three rinses with 1× PBS. For fluorescence imaging, coverslips were slide-mounted and nuclei counterstained using ProLong Gold Antifade (Thermo Fisher Scientific, P36934).

### Sequencing single-cell mRNAs: host cell, rabies virus barcodes and recombined rAAV

scRNA-seq libraries were generated using the Chromium Single Cell v2 or v3 Chemistry platform (10x Genomics), prepared following kit guidelines, and sequenced to a depth of ~45K reads per cell (Illumina NovaSeq 6000). Sequences were aligned using STAR v2.4.0a<sup>66</sup> against a composite genome consisting of GRCm38.81, barcoded cSPBN-4GFP and rAAV accessory sequences (including the 3' UTR and TVA-mCherry, rabies G and Cre coding sequences) using a workflow similar to that described for Drop-seq<sup>67</sup>. To create input cell suspensions, a protocol developed for the adult mouse brain<sup>32</sup> was adapted for in vitro synaptic cultures. Culture wells were first incubated for ~20 min at 37 °C with 1.8 mL of Dissociation Media (DM) containing Papain and Protease 23 until detachment of the cell monolayer. Cultures were gently swirled and incubated for an additional 5 min. Each well was then supplemented with 1 mL of DM before transfer into a 5 mL eppendorf tube in which cells were pelleted through centrifugation ( $300 \times g$  for 5 min). The supernatant was removed and replaced with 1 mL of ice-cold DM in which cells were titrated by successively smaller bore polished glass Pasteur pipets. The cells were then re-pelleted and

resuspended in 0.5 mL of DM before being filtered through a pre-wet 40 µm cell strainer (Corning, 352340). For SBARRO experiments, rabies infected cells were enriched from total cell suspensions through fluorescent activated cell sorting (FACS) using the MoFlo Astrios EQ cell sorter (Beckman Coulter; 70 µm nozzle) into 25 µl of DM. RV-derived EGFP fluorescence was used to gate for SBARRO cells. For experiments in which distinct scRNA-seq libraries were created for starter and presynaptic cells, mCherry fluorescence (driven by cre-recombined rAAV genomes encoding TVA-mCherry) was used as an additional gate to sort starter (GFP+/ TVA-mCherry+) or presynaptic (EGFP+/ TVA-mCherry-) populations. Post hoc FACS analysis was performed with FloJo software (BD Biosciences). For scRNA-seq libraries downstream of FACS, 1.7K–33K cells (based on FACS counts) were loaded per single-cell RNA capture reaction. To generate scRNA-seq libraries for which total cell suspensions were used as input, cell concentrations were quantified using the Countess II Automated Cell Counter (Life Technologies) and 10K–16K cells were loaded per single-cell RNA capture reaction. Rabies virus barcoded EGFP and cre-recombined TVA-mCherry rAAV 3' mRNAs were independently amplified (See “Selective mRNA Adaptor PCR” protocol; Rabies EGFP: ~255 bp amplicon, 10–14 cycles; rAAV TVA-mCherry: ~1125 bp amplicon, 34 cycles) from single-cell cDNA using primers that introduced the Illumina P5 site (P5-TSO\_Hybrid) and indexed P7 site onto bar-coded 3' EGFP (BC\_Seq\_P7i'x'\_GFP\_v4c) or recombined TVA-mCherry (P7'x'\_TCB\_CreOn\_v4). Barcoded EGFP and recombined TVA-mCherry libraries were multiplexed and sequenced separately on an Illumina NextSeq500 using a High Output 150 cycle Kit (Stock Read1 primer; Library concentrations: EGFP, 1.8 pM with 20% PhiX; TVA-mCherry, 0.4 pM with 50% PhiX; Cycle distributions: Read1 = 28, Read2 = 98, Index = 8; Reads per library: EGFP, 43M–90M; TVA-mCherry, 121K–8.6M). To generate integer counts of recombined TVA-mCherry transcripts per cell, sequences generated from recombined TVA-mCherry library were aligned using STAR v2.4.0a<sup>66</sup> against a composite genome consisting of GRCm38.81, the RVdG-EGFP<sub>VBC</sub> genome and rAAV accessory sequences (including the 3' UTR and TVA-mCherry, rabies G and Cre coding sequences). The sequences of UMIs associated with each gene and cell barcode were collapsed within an edit distance of 2. To quantify the number of TVA-mCherry mRNAs derived from cre-recombined rAAV genomes per cell, UMI counts mapping to the TVA-mCherry coding sequence or the 3' UTR were summed. To discover and quantify the RV-derived VBCs in the 3' UTR of EGFP mRNA, raw VBC sequences were informatically extracted from each read (as described above for plasmids and viral genome sequences). To accurately reconstruct and count VBCs in each single-cell, we leveraged the single-cell nature of the data to informatically account for two types of artifacts: (1) the inflation of barcode sequences generated by mutations during library amplification and sequencing and (2) swapping of non-adjacent VBC and cell barcode/UMI sequences due to strand displacement during PCR amplification (“CollapseTagWithContext, ADAPTIVE\_EDIT\_DISTANCE = true” & “BipartiteRabiesVirusCollapse”). To account for mutations, we assumed that in individual cells, closely related barcode sequences were likely to originate from mutations introduced during library preparation or sequencing rather than independent infections of rabies virus particles with similar 20 bp genomic barcodes. Thus we evaluated Hamming edit distance relationships across all sufficiently abundant VBCs (Inclusion Threshold:  $\geq 3$  (“No RG” experiments) or 5 (“SCC” Experiments) UMIs) found within each cell. From these edit distance distributions, many low-abundance “sibling” VBCs with sequences similar to a single, more numerous “parent” VBC were assigned the VBC of the “parent”; collapsing these mutationally-related VBC “families” corrected the strong artifactual correlation present in the raw data in which cells with more VBC UMIs also tended to have more unique VBCs (Supplementary Fig. 3a) reduced the number of included VBC-VBC counts by 82.4%. In the single-cell cDNA, cell barcode/UMI sequences are separated from the

VBC cassette by >20 bps—including tracts of A/T homopolymers—providing an opportunity for mispairing of critical barcode sequences during PCR through strand-displacement or mispriming. To account for mispairing events, in cells with multiple VBCs, we developed a collapse algorithm based on fraction of shared UMI sequences (within edit distance 2) shared across each pair of VBCs. For pairs with >50% UMI sharing, the “sibling” VBC with fewer UMIs was assigned the VBC of the more abundant “parent”, enforcing that CBC-UMI barcodes should not be used by more than a single VBC. The ratio of within-cell VBC collapse events due to UMI sharing versus total CBC-VBC counts averaged  $0.13 \pm 0.02$  (s.e.m.) across experiments; a correction which reduced the number CBC-VBCs counts by an additional 1%. Taken together, these two VBC collapse steps reduced included CBC-VBC counts by 83.4% as compared to the raw data—thus drastically altering the inferred groupings of single cells into networks—and also shaped within-cell VBC quantification, altering the UMI counts for ~15% of VBCs (Change in VBC UMIs: mean, 5.7; median, 2). Selective mRNA Adaptor PCR (50  $\mu$ l reaction): (1) 10x cDNA (-7–12 ng/ $\mu$ l), 1; (2) P5 Primer (10 mM; P5–10x\_Hybrid), 1; P7 Primer (10  $\mu$ M; P7i'x'-GFP or P7i'x'-TCB\_CreOn), 1; Q5 High-Fidelity 2x Master Mix (New England Biolabs, M0492L), 25; Water, 22. Cycle Conditions: (1) 98 °C for 30 s; (2) 98 °C for 10 s; (3) 72 °C for 30 s; (4) 72 °C for 60 s; (5) Go to Step 2–4, (see above for rabies virus or rAAV cycle number ranges); (6) 72 °C for 120 s

### Identification, clustering, and analysis of host cell scRNA profiles

To discover the molecular identities of SBARRO cells, we first distinguished single-cell RNA libraries from background by leveraging properties of both single-cell RNA and VBC data from individual experiments. Specifically, using total single-cell RNA data, we identified cell profiles (1) exclusively associated with cell barcodes provided by 10x Genomics (corresponding to v2 or v3 chemistry) and exhibiting (2) large UMI counts and low fractions of mitochondrial and ribosomal transcripts<sup>32</sup>. In parallel, we used the mutation-collapsed VBC data (see above) to filter and retain those cell profiles with at least a single VBC ascertained with  $\geq 3$  (v2 chemistry) or  $\geq 5$  (v3 chemistry) UMI counts. We used the union of cell barcodes identified by RNA-based and VBC-based methods to generate digital gene expression matrices (DGEs) for each experiment. DGEs were input into a two-staged analysis pipeline based on independent components analysis (ICA)<sup>32</sup>, a semi-supervised approach for grouping scRNA profiles into clusters then subclusters. scRNA profiles corresponding to cell-cell doublets and cell outliers were identified, flagged and excluded from downstream analyses using an Independent Components Analysis-based method<sup>32</sup>. The identities of clusters and subclusters were systematically annotated based on molecular marker expression<sup>32,68</sup>. Prior to ICA analysis, DGEs were pruned of (1) genes present on the rabies virus or mitochondrial genomes and (2) small scRNA profiles (profiles with  $\geq 500$  UMIs (SCC Experiment) or  $\geq 50$  genes (noRG Experiment) were retained) to promote high-quality clustering based on host cell nuclear gene expression. Additional DGEs (subject to the same gene and cell filtering criteria) were also generated while including rabies virus genes to aid in the downstream analyses of rabies virus expression (Supplementary Data 2).

To enhance molecular identification of cells in the SCC experiments, an integrated analysis of SBARRO and control cell libraries was performed using LIGER<sup>69</sup>. Control cells were sampled from total cell suspension not subject to FACS and derived from physically adjacent culture wells seeded with the same cell suspensions as SBARRO experiments. Input DGEs for LIGER analysis lacked rabies virus and mitochondrial genes and were filtered to remove “cell-cell doublet” or “outlier” RNA profiles as identified by upstream ICA-based analysis. LIGER alignment and clustering (factorization,  $k=40$ ,  $\lambda=3$ ; quantile alignment, resolution=0.4,  $\text{knn}_k=20$ ) results were visualized with UMAP embedding and systematically annotated using marker gene expression. Of the 130.5 K SBARRO scRNA profiles, 28.4 K

were grouped into three clusters (cluster 1, 8, and 13) which contained cells of multiple classes and were defined by expression signatures related to GO biological processes such as “cytokine-mediated signaling pathway (GO:0019221)” (cluster 13; adjusted  $p$  value  $< 2.1 \times 10^{-14}$ ) or “PERK-mediated unfolded protein response (GO:0036499)” (cluster 8; adjusted  $p$  value  $< 0.001$ ). To clarify the molecular identities of these 28.4 K cells, we re-aligned these SBARRO libraries to the control cells from SCC7 and SCC8 using LIGER (factorization,  $k=60$ ,  $\lambda=3$ ; quantile alignment, resolution=0.4,  $\text{knn}_k=20$ ). The resulting analysis split SBARRO libraries across 32 (of 37 total) clusters which exhibited molecular marker expression consistent with known cell populations. Clusters were then systematically annotated in a manner consistent with the initial LIGER analysis guided by the original and re-aligned cluster identities of control cells. In total, we identified  $n=20$  “granular” cell populations which could be grouped into  $n=11$  “coarse” populations. To identify and visualize genes differentially expressed across granular populations, we used the FindMarkers() and DotPlot() functions from Seurat<sup>70,71</sup>. To evaluate which cell populations were sensitive or recalcitrant to rabies virus infection, for each population, we performed a chi-square test comparing the number of SBARRO and control scRNA profiles to the dataset totals and used the resulting residuals as a metric of enrichment or depletion. To determine the relationship between the SBARRO molecular identities and cortical neuron types from the adult mouse cortex, we used LIGER to jointly analyze scRNA profiles from 78.8 K uninfected SBARRO control cells and 57 K cells from various neocortical regions ascertained by the Allen Institute<sup>19</sup>. We conducted separate analyses for glutamatergic and GABAergic neurons using the same LIGER parameters (factorization,  $k=10$ ,  $\lambda=30$ ; quantile alignment, resolution=0.2,  $\text{knn}_k=400$ ).

### Transcriptional identification of starter and presynaptic cells

Starter cells are functionalized after Cre-mediated recombination inverts TVA-mCherry and rabies virus B19G transgenes within the FLEX rAAV genome into the sense orientation with respect to the CAG promoter<sup>11,72</sup>. RNA-based identification of starter cells in a direct and qualitative manner is complicated using these vectors in the context of 3' scRNA-seq since (1) low expression caused Cre mRNAs not to be captured with high-probability in Cre+ cells (caused by the limiting MOI of the Syn1-EBFP-Cre rAAV ( $\sim 8 \times 10^{-3}$  –  $8 \times 10^{-5}$ ); mild RNA Polymerase II recruitment with the *Synapsin1* promoter; and lack of 3' motifs to promote mRNA stability) and (2) the vast majority of scRNA-seq reads that align to FLEX rAAV genome do so in the 3' UTR, a region unaffected by recombination. To overcome these limitations and identify starter cells from scRNA profiles alone, we developed a protocol to selectively amplify and sequence only TVA-mCherry mRNAs transcribed from Cre-recombined rAAV genomes (see “Selective mRNA Adaptor PCR” above) and a downstream informatic approach using these data to identify rare, candidate starter cells. Specifically, we developed a binomial test with experiment-specific success rate parameter (updated using an expectation-maximization-like approach) to identify cells in which both recombined TVA-mCherry UMIs (relative to total rAAV UMIs) and total rAAV UMIs (relative to host cell RNA UMIs) were enriched in a manner unlikely to be due to chance (Bonferroni-corrected  $p < 0.01$ ). Properties of RNA expression that distinguish identified starter cells from presynaptic cells in SBARRO experiments—such as the ratio of recombined TVA-mCherry UMIs/total rAAV UMIs—were observed in scRNA-seq libraries in which starter cells were physically separated from presynaptic cells using FACS.

### Cell-type-specific synaptic network inference using rabies virus barcodes

To facilitate interactive analysis and discovery of cell-type-specific SBARRO networks, we created an R Shiny program (“Terminal E”) which allows dynamic filtering and plotting of VBCs and VBC-based synaptic networks. Terminal E integrates data from VBC libraries and

individual SBARRO experiments (organized into “collections” that enable cross-experiment meta-analyses). Library-level data include VBC abundances and user-defined, library-specific lists of single VBCs or VBC pairs to exclude from network inference. SBARRO-level data center around the properties of each cell in each experiment, which, at minimum, include (1) VBC UMI counts; (2) molecular cell type identities; and (3) starter or presynaptic assignments. Inferred synaptic networks are collections of cells which share one or more VBC which are statistically likely to have entered those cells through clonal replication and spread from a single starter cell infection. Terminal E supports the inference of such networks through two stages of VBC filtering. During the first stage, filters completely exclude VBCs from network consideration. We only considered VBCs with  $\geq 7$  UMIs (SCC experiments, v3 10x chemistry) or  $\geq 3$  UMIs (No RG Experiments, v2 10x chemistry). We additionally removed those VBCs with high abundances in genomes of the infecting library (FI trust score  $\geq 5$ ; see below). We further identified specific VBCs for exclusion by (1) comparing their presence and abundance in rabies virus libraries to behavior transducing 17 K starter cell scRNA profiles (containing 28.8 K founder infections) or (2) by comparing across  $n = 22$  independent SBARRO experiments. We excluded the named categories of VBCs below based on the following criteria (Supplementary Fig. 7b):

1. “Felony” VBCs ( $n = 234$ ). Absent from library but observed in  $>1$  of 5015 starter cells (Fig. 2g).
2. “Misdemeanor” VBCs ( $n = 78$ ). Present in library, but observed in  $>2$  starter scRNA profiles with library frequency  $<10^{-6}$  or observed in  $>8$  starter scRNA profiles with library frequency  $<10^{-5.5}$  (Fig. 2f, g).
3. “Cross Experiment” VBCs ( $n = 239$ ) Absent from library but observed in  $>1$  of 22 independent SBARRO experiments.

In the second stage, VBCs or VBC sets are filtered such that those retained were suitably rare enough to enter the experiment through a single starter cell. Critical to this stage of experiment-specific VBC filtering is an estimate of the total number of experiment-specific founder infections starter cells; a subset of which lead to cell-to-cell spread (spreading founder infections). To estimate the number of spreading founder infections for each experiment, we developed an analytical approach designed to mimic founder infections by drawing samples of VBCs from the viral library. We created distributions describing the number of library draws ( $n = 10$  replicates, with VBC replacement) required to match the number of unique VBCs observed in  $\geq 2$  cells present in each experiment; median values set the experiment-specific founder infection estimates. To evaluate whether each VBC or VBC set was suitably rare enough to be included for network inference, we calculated and assigned an “founder infection trust” (FI trust score). The “FI trust” score equates to the number of spreading founder infections that could in theory occur for that VBC or VBC set before a second founder infection was expected (at a given probability) by leveraging library VBC frequencies:

$$\text{FI trust}_p = \log_{10}(p) / \log_{10}(1 - f)$$

Where  $f$  is the frequency of an individual VBC in the library (or, for the VBC set, the inferred frequency calculated by multiplying the frequencies of each VBC member) and  $p$  is the probability of avoiding a second occurrence of the VBC or VBC set. For a given experiment, VBC or VBC sets that were retained to infer synaptic networks were those for which  $\text{FI trust}_p >$  experiment estimates for founder infections. For SCC experiments,  $p = 0.9$ . Inferred networks containing a single scRNA profile assigned as a starter cell were then split into postsynaptic (i.e. the starter) and presynaptic cells; networks lacking a starter cell were starter-orphaned and all scRNA profiles were presumed presynaptic. Terminal E facilitates the comparisons of presynaptic network size (the number of presynaptic cells) and cell type composition across

postsynaptic starter cells of different types. After first stage filtering, VBCs absent from the library were presumed to be rare and assigned the lowest library frequency value to match library VBCs counted with a single UMI. After stratifying presynaptic networks by starter cell type, presynaptic cell type compositions were compared using a Chi-Square Test. Presynaptic network sizes were assigned to each scRNA profile using described rules (Supplementary Fig. 7d) and presynaptic network sizes were compared across starter cell types using a Wilcoxon Rank-Sum test.

### Postsynaptic RNAs associated with the number of presynaptic partner cells

Postsynaptic RNA profiles from the SCC experiments identified by as one of four abundant starter cell types (glutamatergic neurons, interneurons, SPNs or astrocytes;  $n = 144$ ) were binned based on their inferred presynaptic network size: small (2–4 cells); medium (5–6 cells) or large (7+ cells). These bins approximate inflections in the total distribution of presynaptic network sizes (Fig. 4b). To determine if viral load was associated with presynaptic network size category, the fraction of total cellular mRNA derived from all five rabies virus genes was compared using a Wilcoxon Test. A similar strategy was used to test for differences in innate immune responses, using an aggregate expression score derived from the 564 genes (a subset of 646 total genes for which we detected a transcript) curated as part of the mouse innate immune response<sup>23</sup>. To identify candidate gene expression differences associated with presynaptic network size, we compared starter cell RNA profiles associated with “small” and “large” presynaptic across cell subtypes rather than types, to avoid confounds due to differences in subtype compositions. Because these comparisons involved small numbers scRNA profiles (range: 2–28 scRNA profiles; mean, 6.3; median, 4) we used the following strategy to identify genes and contextualize how likely such differences were likely to arise by chance. First, we used Fisher’s Exact Test to evaluate differences in UMI counts across scRNA profiles aggregated by small or large presynaptic size category. We considered only those genes with  $\geq 25$  UMIs thus lessening the burden of multiple hypothesis testing and corrected our  $p$  value cut  $<0.05$  by the number of tests completed within each subtype comparison. Second, for the genes which passed threshold, we used a Wilcoxon Test to determine whether normalized RNA levels (removing contributions from rabies virus mRNAs before normalization) differed across the population of individual cells associated with each presynaptic network size category, using  $p < 0.05$  as our cutoff. To help determine which of the genes we identified were likely due to chance, we repeated this testing procedure for 100 permuted comparisons in which each starter cell RNA profile was replaced by a presynaptic cell RNA profile of the same subtype choose at random. Genes identified by multiple permuted replicates in the same cell subtype were flagged as potentially spurious and not considered further. Log fold expression changes were calculated after normalizing the number of gene-specific UMIs by total UMIs for large and small presynaptic network categories and then scaling the data to 100,000 transcripts after the addition of a pseudocount.

### Identifying mRNAs correlated with rabies virus transmission across SPN development

Monocle3<sup>26,73</sup> was used to calculate pseudotime scores for scRNA profiles of developing SPNs, from neural precursor cells to mature neurons, after preprocessing (method = PCA; number of dimensions = 10) and alignment SBARRO and control libraries (alignment  $k = 3000$ ). To identify modules of genes with similar expression levels over pseudotime, all genes were first tested for pseudotime-associated expression using the `graph_test()` function. Genes with  $q$  values = 0 were retained and modules identified using the `find_gene_modules()` function with resolution = 0.001. Pseudotime-ordered cells were grouped into 10 bins. For each bin, (1) the fraction of total cells that

were SBARRO (rather than control) in origin were calculated and (2) a meta-control cell was created by summing control cell UMIs and normalizing such that mRNA expression values summed to 100,000. Pearson correlations were calculated for each detected gene by comparing SBARRO fractions and normalized gene expression values across the pseudotime bins. SBARRO-correlated genes ( $n = 3309$ ) were defined as those genes in which  $r \geq 0.75$ . SBARRO-correlated genes were tested for gene set enrichment using PantherGO<sup>27,28,74</sup> via EnrichR<sup>75</sup> and SynGO<sup>31</sup> and compared to two sets ( $n = 1000$  replicate gene selections/set) of  $n = 3309$  control genes, selected either (1) from expression-matched deciles built from the developmentally mature SPN meta-control cell (bin = 10) or from (2) at random from genes with expressed RNA.

### Reporting summary

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

### Data availability

The raw and processed sequencing data reported in this paper are available from the GEO repository (Accession Code: [GSE214386](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/)). The GRCm38.81 mouse genome build is publicly available ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001635.20/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/)). PantherDB(17.0; <http://pantherdb.org/>) and SynGO(release 20180731; <https://www.syngoportal.org/>) source data are provided as a Source Data file for all main (Figs. 1–5) and Supplemental Figures (Supplemental Data Figs. 1–10). Source data are provided with this paper.

### Code availability

Software and core computational analysis to align and process scRNA-seq reads are freely available: <https://github.com/broadinstitute/Drop-seq/releases>. Other custom analysis code is available by request because the code contains additional analyses not included in this publication.

### References

- Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
- Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **40**, 181–185 (2018).
- Biddy, B. A. et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
- Kim, D.-W. et al. Multimodal analysis of cell types in a hypothalamic node controlling social behavior. *Cell* **179**, 713–728.e17 (2019).
- Clark, I. C. et al. Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372**, eabf1230 (2021).
- Ugolini, G. Advances in viral transneuronal tracing. *J. Neurosci. Methods* **194**, 2–20 (2010).
- Lafon, M. Rabies virus receptors. *J. NeuroVirol.* **11**, 82–87 (2005).
- Kelly, R. M. & Strick, P. L. Rabies as a transneuronal tracer of circuits in the central nervous system. *J. Neurosci. Methods* **103**, 63–71 (2000).
- Ghanem, A. & Conzelmann, K.-K. G gene-deficient single-round rabies viruses for neuronal circuit analysis. *Virus Res.* **216**, 41–54 (2016).
- Wickersham, I. R. et al. Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
- Miyamichi, K. et al. Dissecting local circuits: parvalbumin interneurons underlie broad feedback control of olfactory bulb output. *Neuron* **80**, 1232–1245 (2013).
- Mebatsion, T., König, M. & Conzelmann, K. K. Budding of rabies virus particles in the absence of the spike glycoprotein. *Cell* **84**, 941–951 (1996).
- Paradis, S. et al. An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron* **53**, 217–232 (2007).
- Sharma, K. et al. High-throughput genetic screen for synaptogenic factors: identification of LRP6 as critical for excitatory synapse development. *Cell Rep.* **5**, 1330–1341 (2013).
- Williams, M. E. et al. Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. *Neuron* **71**, 640–655 (2011).
- Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nat. Protoc.* **5**, 595–606 (2010).
- Huang, A. S. & Baltimore, D. Defective viral particles and viral disease processes. *Nature* **226**, 325–327 (1970).
- Ichikawa, M., Muramoto, K., Kobayashi, K., Kawahara, M. & Kuroda, Y. Formation and maturation of synapses in primary cultures of rat cerebral cortical cells: an electron microscopic study. *Neurosci. Res.* **16**, 95–103 (1993).
- Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021).
- Kanari, L. et al. Objective morphological classification of neocortical pyramidal cells. *Cereb. Cortex* **29**, 1719–1735 (2019).
- Kwon, H.-B. et al. Neuroligin-1-dependent competition regulates cortical synaptogenesis and synapse number. *Nat. Neuro* **15**, 1–10 (2012).
- Berns, D. S., DeNardo, L. A., Pederick, D. T. & Luo, L. Teneurin-3 controls topographic circuit assembly in the hippocampus. *Nature* **554**, 328–332 (2018).
- Breuer, K. et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228–D1233 (2013).
- Rehfeld, F. et al. The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat. Commun.* **9**, 1–13 (2018).
- Matho, K. S. et al. Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598**, 182–187 (2021).
- Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 1–31 (2019).
- Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
- The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still going strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
- Thoulouze, M. I. et al. The neural cell adhesion molecule is a receptor for rabies virus. *J. Virol.* **72**, 7181–7190 (1998).
- Wang, J. et al. Metabotropic glutamate receptor subtype 2 is a cellular receptor for rabies virus. *PLoS Pathog.* **14**, e1007189 (2018).
- Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234.e4 (2019).
- Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018).
- Krienen, F. M. et al. Innovations present in the primate interneuron repertoire. *Nature* **586**, 1–29 (2020).
- Gouwens, N. W. et al. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 935–953.e19 (2020).
- Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **18**, 1–34 (2020).
- Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife* <https://doi.org/10.7554/elife.14472.001> (2016).



37. Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
38. Stickels, R. R. et al. Sensitive spatial genome wide expression profiling at cellular resolution. **92**, 1168–13 (2020).
39. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **11**, 1–24 (2019).
40. Chen, X. et al. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772–786.e19 (2019).
41. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, 5691–5618 (2018).
42. Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
43. Ray, T. A. et al. Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nat. Commun.* **11**, 1–20 (2020).
44. Südhof, T. C. Synaptic neurexin complexes: a molecular code for the logic of neural circuits. *Cell* **171**, 745–769 (2017).
45. Wickersham, I. R., Finke, S., Conzelmann, K.-K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat. Methods* **4**, 47–49 (2006).
46. Ren, S.-Q., Li, Z., Lin, S., Bergami, M. & Shi, S.-H. Precise long-range microcircuit-to-microcircuit communication connects the frontal and sensory cortices in the mammalian brain. *Neuron* **104**, 1–17 (2019).
47. Reardon, T. R. et al. Rabies virus CVS-N2c deltaG strain enhances retrograde synaptic transfer and neuronal viability. *Neuron* **89**, 711–724 (2016).
48. Rossi, L. F., Harris, K. D. & Carandini, M. Spatial connectivity matches direction selectivity in visual cortex. *Nature* **588**, 648–652 (2020).
49. Wertz, A. et al. Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349**, 70–74 (2015).
50. Albisetti, G. W. et al. Identification of two classes of somatosensory neurons that display resistance to retrograde infection by rabies virus. *J. Neurosci.* **37**, 10358–10371 (2017).
51. Beier, K. T. et al. Rabies screen reveals GPe control of cocaine-triggered plasticity. *Nature* **549**, 345–350 (2017).
52. Rogers, A. & Beier, K. T. Can transsynaptic viral strategies be used to reveal functional aspects of neural circuitry? *J. Neurosci. Methods* **348**, 109005 (2020).
53. Prośniak, M., Hooper, D. C., Dietzschold, B. & Koprowski, H. Effect of rabies virus infection on gene expression in mouse brain. *Proc. Natl Acad. Sci. USA* **98**, 2758–2763 (2001).
54. Kim, S. et al. Early transcriptional changes in rabies virus-infected neurons and their impact on neuronal functions. *Front. Microbiol.* **12**, 730892 (2021).
55. Huang, K. W. & Sabatini, B. L. Single-cell analysis of neuroinflammatory responses following intracranial injection of G-deleted rabies viruses. *Front. Cell Neurosci.* **14**, 65 (2020).
56. Patiño, M. et al. Single-cell transcriptomic classification of rabies-infected cortical neurons. *Proc. Natl Acad. Sci. USA* **119**, e2203677119 (2022).
57. Mount, C. W., Yalçın, B., Cunliffe-Koehler, K., Sundaresh, S. & Monje, M. Monosynaptic tracing maps brain-wide afferent oligodendrocyte precursor cell connectivity. *eLife* <https://doi.org/10.7554/elife.49291.001> (2019).
58. Mori, T. & Morimoto, K. Rabies virus glycoprotein variants display different patterns in rabies monosynaptic tracing. *Front. Neuroanat.* **7**, 47 (2014).
59. Cheung, V. et al. Virally encoded connectivity transgenic overlay RNA sequencing (VECTORseq) defines projection neurons involved in sensorimotor integration. *Cell Rep.* **37**, 110131–110131 (2021).
60. Hanchate, N. K. et al. Connect-seq to superimpose molecular on anatomical neural circuit maps. *Proc. Natl Acad. Sci. USA* **117**, 4375–4384 (2020).
61. Tsai, N. Y. et al. Trans-Seq maps a selective mammalian retinotectal synapse instructed by Nephronectin. *Nat. Neurosci.* **25**, 659–674 (2022).
62. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **337**, 1–14 (2017).
63. Barr, J. N., Whelan, S. P. & Wertz, G. W. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J. Virol.* **71**, 8718–8725 (1997).
64. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
65. Schnell, M. J., Mebatsion, T. & Conzelmann, K. K. Infectious rabies viruses from cloned cDNA. *EMBO J.* **13**, 4195–4203 (1994).
66. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
67. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
68. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 1–18 (2019).
69. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
70. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
71. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
72. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
73. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
74. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2018).
75. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma.* **14**, 128 (2013).

## Acknowledgements

This work was supported by the Broad Institute's Stanley Center for Psychiatric Research and a Helen Hay Whitney Postdoctoral Fellowship to A.S. The authors thank Frank Koopmans for SynGO analysis scripts, Dr. Fenna Krienen, Dr. Marta Florio, Dr. Christina Usher and other members of the McCarroll lab for helpful advice.

## Author contributions

A.S. conceived the idea, designed and supervised experiments, analyzed the data and drafted the manuscript. Plasmid barcoding, C.V. and A.S. Rabies virus packaging, K.W.H., A.P., C.V., A.S., and B.L.S. Cell culture, K.S. and C.S. scRNA-seq experiments/analysis and molecular protocols, C.S., K.S., H.S., and A.S. Algorithm development, S.A.M., J.N., and A.S. Analysis software, S.K. Computational support, A.W. Manuscript preparation, A.S. and S.A.M. with input from other authors.

## Competing interests

A.S. and S.A.M. are listed as inventors on a patent application related to the work. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-34334-1>.

**Correspondence** and requests for materials should be addressed to Arpiar Saunders or Steven A. McCarroll.

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

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

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

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

© The Author(s) 2022