1 **TITLE**

2 Ascertaining cells' synaptic connections and RNA expression simultaneously with

3 massively barcoded rabies virus libraries

4

5 **AUTHORS**

6 Arpiar Saunders[1,2,4]*, Kee Wui Huang[3], Cassandra Vondrak[1,2], Christina Hughes[1,2],

7 Karina Smolyar[1,2], Harsha Sen[1,2], Adrienne C. Philson[3], James Nemesh[1,2], Alec

8 Wysoker[1,2], Seva Kashin[1,2], Bernardo L. Sabatini[3] and Steven A. McCarroll[1,2]*

9

10 [1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

11 [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge,

12 MA 02142, USA

13 [3]Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical

14 School, Boston, MA 02115, USA

15 [4]Vollum Institute, Oregon Health & Science University, Portland, OR, 97239, USA

16

17 * Correspondence:

18 saundear@ohsu.edu

19 mccarroll@genetics.med.harvard.edu

20

21

22

23

24

ABSTRACT

Brain function depends on forming and maintaining connections between neurons of specific types, ensuring neural function while allowing the plasticity necessary for cellular and behavioral dynamics. However, systematic descriptions of how brain cell types organize into synaptic networks and which molecules instruct these relationships 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[1]. Thousands of these partner relationships can be ascertained in a single experiment, alongside genome-wide RNA profiles – and thus cell identities and molecular states – of each host cell. We used SBARRO to describe synaptic networks formed by diverse mouse brain cell types *in vitro*, leveraging a system similar to those used to identify synaptogenic molecules. We found that the molecular identity (cell type/subtype) of the starter cell predicted the number and types of cells that had synapsed onto it. Rabies transmission tended to occur into cells with RNA-expression signatures related to developmental maturation and synaptic transmission. The estimated size of a cell's presynaptic network, relative to that of other cells of the same type, associated with increased expression of *Arpp21* and *Cdh13*. By tracking individual virions and their clonal progeny as they travel among host cells, single-cell, single-virion genomic technologies offer new opportunities to map the synaptic organization of neural circuits in health and disease.

MAIN

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[2-5]. Together with engineered proteins and viruses, additional cell features such as protein expression[2], developmental origin[3,4], axonal projection patterns[5] and physical interactions[6] 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[7]. Prevailing models suggest such transmission events occur at synapses, likely due to the presence of viral entry receptors[8] and high rates of membrane turnover. While the synaptic phenomenology of rabies virus transmission has been used for decades to discover neural pathways[9], inefficient conversion of plasmid DNA into infective RNA-containing particles has largely precluded using rabies and other *lyssaviruses* in genomic applications[10].

Here, we introduce SBARRO (Synaptic Barcode Analysis by Retrograde Rabies ReadOut), which combines monosynaptic rabies virus tracing, 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 monosynaptic networks to be reconstructed in parallel by tracking paths of clonal infections which

3

71   originate in postsynaptic starter cells and spread to those cells' presynaptic partners. By

72   replacing the endogenous glycoprotein gene (*G*) – necessary for viral spread – with

73   *EGFP* in the viral genome, rabies virus transmission is restricted to cells that are directly

74   presynaptic[1,11]. In sampling cellular RNAs alongside VBCs, our approach reveals: 1)

75   postsynaptic vs presynaptic cell identities; 2) cell types and molecular states (via host

76   cell RNAs); and synaptic networks (via shared VBCs; **Fig. 1a**). In genomics, rabies virus

77   has been previously used in RABID-seq[6], in which rabies virus spread was used to infer

78   putative direct physical contacts between glia, a previously unknown and

79   uncharacterized type of rabies virus transmission.

80

81   Here we present a comprehensive experimental and analytical framework for using

82   SBARRO to discover structural and molecular properties of cell-type-specific synaptic

83   connectivity. We leverage dissociated mouse brain cells grown into synaptic networks *in*

84   *vitro*, similar to systems used to identify synaptogenic molecules[12,13] and for which

85   features of *in vivo* connectivity can remain[14]. We discover that presynaptic network

86   properties, such as cell type composition and size, are conferred in part by the

87   postsynaptic cell type. Finally, we discover that rabies virus spread associates with a

88   molecular signature of synaptic maturation, suggesting that functioning synapses, in

89   addition to entry receptors, are critical for uptake of rabies virus.

90

91

92   **Barcoding millions of rabies viral genomes**

93   Synaptic tracing with barcoded rabies requires libraries of barcoded rabies virus particles

94   that have high numbers of unique barcodes which are as uniform as possible in

95   abundance, such that individual viral barcodes are introduced into no more than one
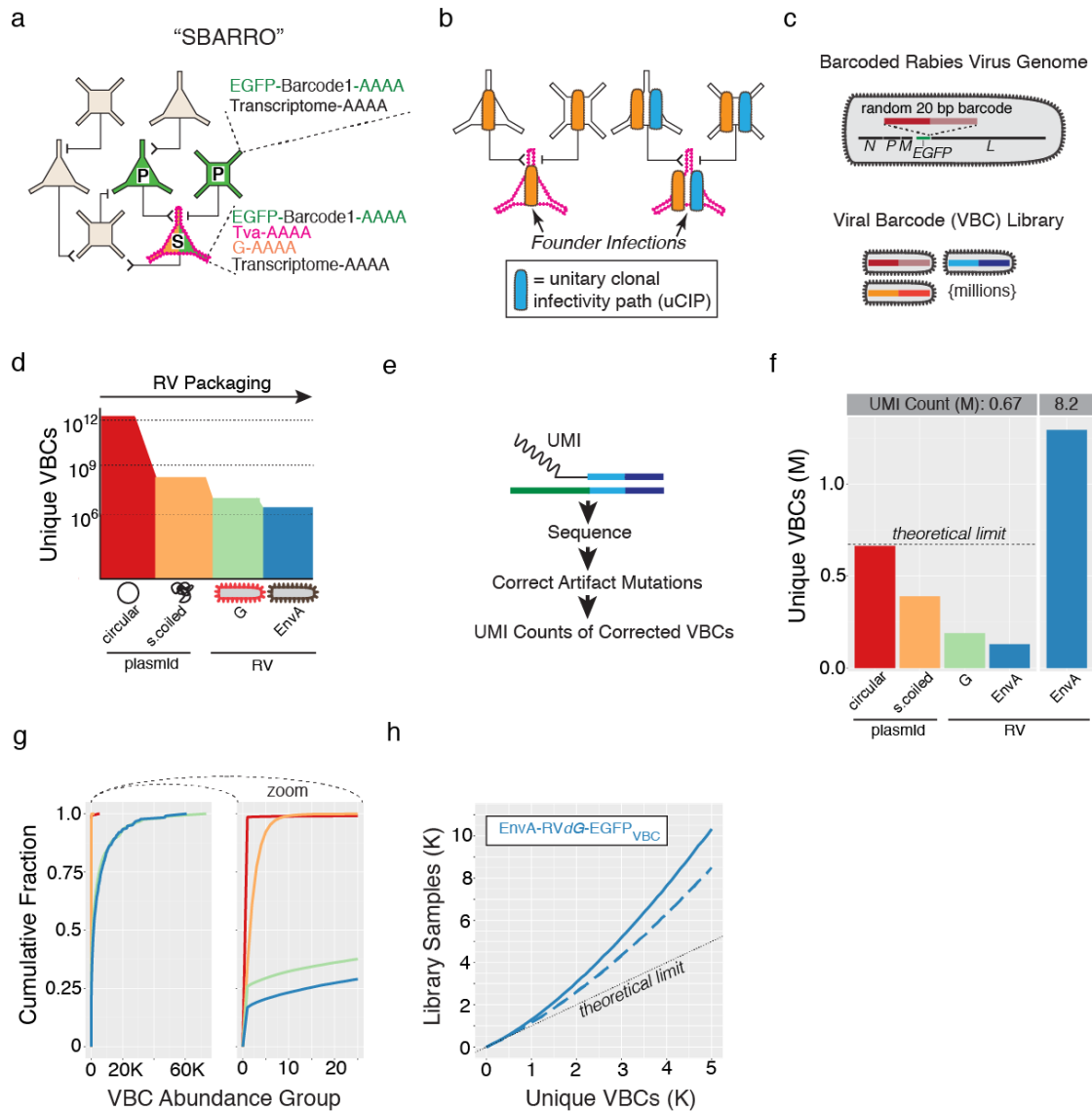
Figure 1

Figure 1. **Single-virion RNA tracking enabled by libraries of rabies virus particles encapsidating millions of uniquely barcoded genomes. a.** Monosynaptic SBARRO schematic. TVA-expressing starter cells ("S") complemented in *trans* with rabies virus glycoprotein (G) are selectively transduced by EnvA-pseudotyped rabies virus in which *G* has been replaced barcoded *EGFP* (EnvA-RV*dG*-EGFP$_{VBC}$). G-complemented clonal particles spread a single retrograde synapse into presynaptic partner cells ("P"). Single-

cell RNA profiles inform 1) synaptic groupings (from *EGFP* based viral barcode (VBC) sharing); 2) Starter or presynaptic status (from *TVA* mRNAs) and 3) host cell type (by capturing thousands of cellular mRNAs). **b.** Monosynaptic relationships are inferred through unitary clonal infectivity paths (uCIPs), defined by a subset of VBCs carried by those rabies virus particles sufficiently rare enough in the infecting library to seed single founder infections in starter cells. **c.** Rabies virus particles are distinguished from each other by a 20 bp bipartite VBC in the 3' UTR of *EGFP*. VBC libraries contain millions of unique particles. **d.** Schematic of VBC diversity during barcoded rabies virus packaging. **e.** Schematic of sequencing-based genomic VBC quantification using unique molecular identifiers (UMIs; **Extended Data Fig. 1c**). Post-cellular polymerase mutations incurred during library amplification and sequencing were corrected informatically (**Extended Data Fig. 2b** and **Methods**). **f-h.** VBC diversity metrics (color-coded as in panel **d**). **f.** Unique VBCs identified by 0.67 million (M) UMI counts across each packaging stage (*left*) or with ~12.2 fold more counts (8.2 M) after EnvA pseudotyping (*right*). **g.** Cumulative distribution of VBCs binned by "VBC abundance group" (AG) across packaging stages (0.67 M counts / stage). Total counts for all VBCs sampled once belong to AG = 1; sampled twice belong AG = 2, etc. **h.** The relationship between the number of unique VBCs identified after a given number of *in silico* samples drawn from the EnvA-RV*dG-EGFP*VBC library (8.2 M counts; blue line) and after removing the 88 most abundant VBCs (dashed blue line). The dotted line shows maximum theoretical diversity (in which every drawn VBC is unique).

96   starter cell (in an experiment) and can be used to define "unitary" clonal infectivity paths

97   (uCIPs; **Fig. 1b**). Inefficiencies in creating negative-stranded RNA viruses from DNA

98   have historically precluded generating complex rabies libraries[10]. To generate libraries

99   encoding millions of barcodes, we developed molecular and computational methods to

100   introduce, retain, and quantify barcodes in DNA plasmids and in rescued RNA genomes

101   (**Fig. 1c,d**; **Extended Data Fig. 1** and **Extended Data Fig. 2**).

102

103   We first developed a PCR-based strategy to flexibly engineer bipartite barcodes,

104   generated through combinatorial diversity, into circular DNA plasmids (**Extended Data**

105   **Fig. 1b**), followed by transformation and plate-based growth conditions optimized to

106   retain DNA plasmid barcode diversity (**Extended Data Fig. 2b,c**). We also created a

107   rabies rescue system – achieving equivalent viral titers 3-fold faster than the current

108   protocols[15] – that minimized barcode loss and disproportionate amplification during viral

109   replication (**Extended Data Fig. 2d**). To assess viral barcode diversity and distribution,

110   we used single-molecule sequencing (**Fig. 1e** and **Extended Data Fig. 1c**), for which we

111   developed analysis methods to identify and correct for PCR and sequencing mutations

112   (**Extended Data Fig. 2b** and **Methods**).

113

114   We used this approach to generate an EnvA-pseudotyped rabies library with a 20 bp

115   randomer encoded in the 3' UTR of *EGFP* of SAD-dG-B19 (EnvA-RV*dG-EGFP*$_{\text{VBC}}$;

116   **Extended Data Fig. 1a**). We compared the total number of unique barcodes and their

117   relative abundances across each production stage (**Fig. 1d-g** and **Extended Data Fig.**

118   **1a**). After PCR and circularization, nearly every sequenced plasmid contained a unique

119   barcode. This diversity was reduced by bacterial amplification, though without

120   substantially distorting representation of the retained barcodes. Rabies rescue induced

121 barcode loss and abundance distortions and was mildly exacerbated by EnvA-

122 pseudotyping. Deeper sequencing of the final EnvA-RV*dG-EGFP*$_{VBC}$ genomes (6.4

123 unique molecular identifiers (UMI) per viral barcode on average) quantified the relative

124 abundances of 1.29 million unique, error-corrected barcodes.

125

126 To estimate the fraction of EnvA-RV*dG-EGFP*$_{VBC}$ founder infections that would be from

127 viral particles with unique barcodes, we performed *in silico* mock infections by randomly

128 sampling barcodes from the sequenced genomes of the infecting library and calculated

129 the resulting number of unique barcodes (**Fig. 1h**). For these analyses, we used 50%

130 unique barcodes as our benchmark, though the actual number of unique founder

131 infections depends on properties of the infecting library and the number of founder

132 infections in the experiment. Sampling up to ~8,900 library genomes resulted in >50%

133 unique barcodes; this could be increased to 15,500 library genomes by filtering out the

134 88 most abundant barcodes in the library, and to 83,600 library genomes by mixing 9

135 equivalent libraries *in silico* (**Extended Data Fig. 2e**). These analyses suggested that

136 our optimized protocols have helped overcome inefficiencies that previously limited

137 rabies applications in scalable genomics research and suggested uCIPs can be

138 efficiently generated from thousands of founder infections.

139

140

141 **Characterizing the barcoded rabies library with 28,000 founder infections**

142 To directly determine the relationship between our barcoded EnvA-RV*dG-EGFP*$_{VBC}$

143 particles and the brain cells they infect, we infected three replicate cell cultures derived

144 from embryonic mouse cortex. Infections were targeted to cells by recombinant adeno-

145 associated virus (rAAV) expression of TVA; the host cells lacked the G protein
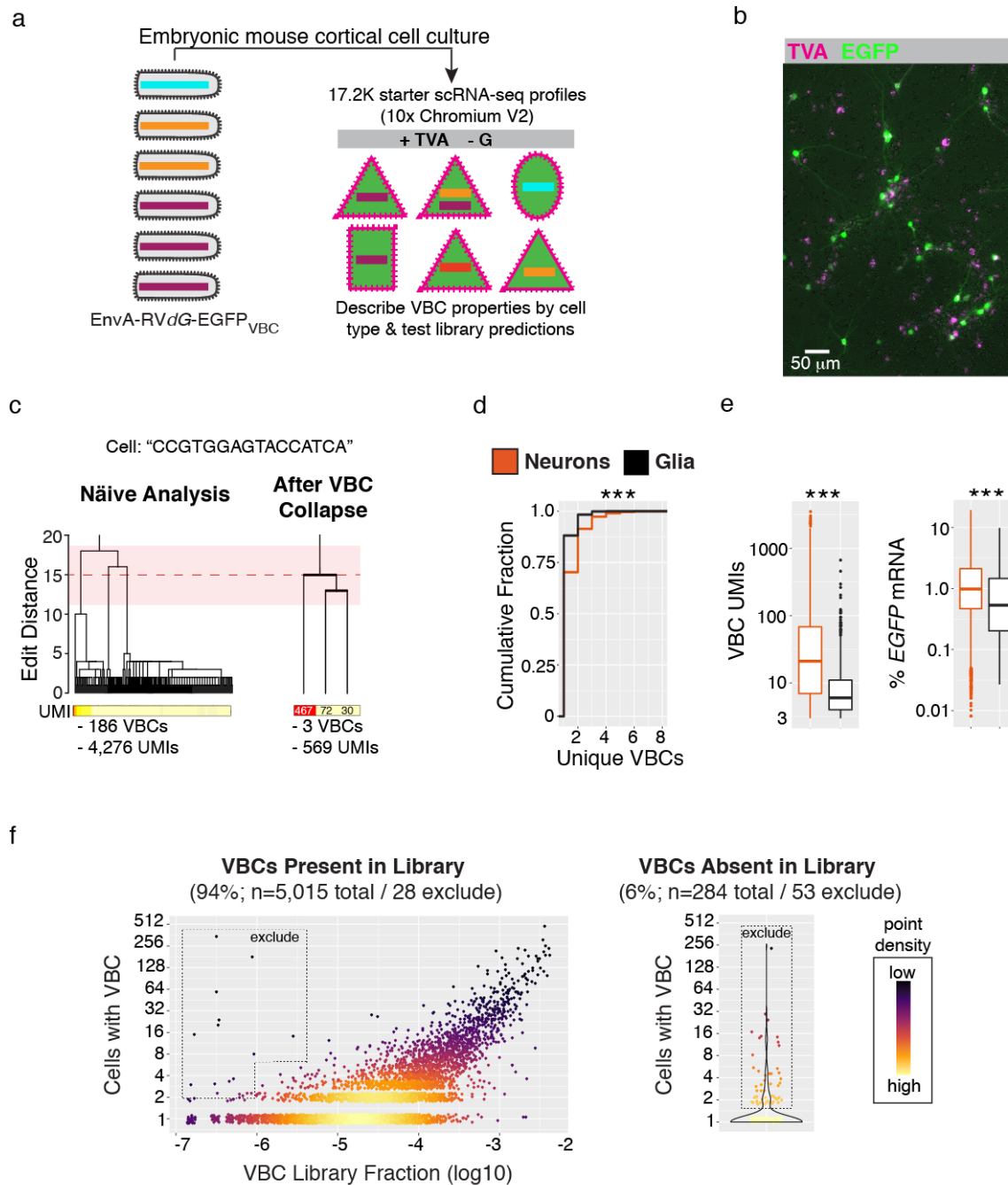
Figure 2

a

Embryonic mouse cortical cell culture

17.2K starter scRNA-seq profiles
(10x Chromium V2)

+ TVA   - G

Describe VBC properties by cell
type & test library predictions

EnvA-RV*dG*-EGFP$_{VBC}$

b

TVA EGFP

50 μm

c

Cell: "CCGTGGAGTACCATCA"

Näive Analysis   **After VBC Collapse**

Edit Distance

UMI

- 186 VBCs   - 3 VBCs
- 4,276 UMIs   - 569 UMIs

467 72 30

d

Neurons ■**Glia**

***

Cumulative Fraction

Unique VBCs

e

***   ***

VBC UMIs   % *EGFP* mRNA

f

VBCs Present in Library
(94%; n=5,015 total / 28 exclude)

Cells with VBC

exclude

VBC Library Fraction (log10)

VBCs Absent in Library
(6%; n=284 total / 53 exclude)

Cells with VBC

exclude

point density

low

high

Figure 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. The EnvA-RV*dG-EGFP*$_{VBC}$ library transduced starter cells (+TVA) from which

the rabies virus could replicate but not spread to other cells (-G). RNA profiles were captured (n=60.8K cells), including from cells infected by rabies virus which function as a corpus of starter cells (n=17.2K). **b.** Representative image of dissociated mouse brain cell cultures (14 days *in vitro*) expressing TVA (magenta) and EGFP (green). 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) from distribution of edit distances among random barcode sequences. **d-e**. Comparison of single cell VBC properties ascertained from RNA profiles of neurons (n= 4,222) or glia (n= 914; *** = $p < 2.2e-16$, Kolmogorov–Smirnov Test). Only data from 1:10 EnvA-RV*dG-EGFP*$_{VBC}$ dilution are shown (**Methods** and **Extended Data Fig. 3f**). **d**. Cumulative distribution of unique VBCs. **e**. Total VBC UMIs (*left*) or % EGFP mRNA (*right*). **f.** Critically evaluating the performance of the EnvA-RV*dG-EGFP*$_{VBC}$ through a corpus of 17.2K 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**).

146 necessary for rabies spread (**Fig. 2a,b** and **Methods)**. After 72 hours, we collected

147 60,816 transcriptomes (n=6 scRNA-seq libraries each from a single culture well) that

148 captured both the cellular RNAs and the barcoded region of *EGFP* mRNAs (**Extended**

149 **Data Fig. 3** and **Methods)**.

150

151 Naïve analysis of barcode sequences from individual cells initially suggested large

152 barcode "families" with many highly similar sequences (**Fig. 2c**). Reasoning that such

153 relationships were largely created by PCR or sequencing errors, we developed an

154 algorithm to collapse families of highly similar barcode sequences into the single

155 barcode responsible for the putative founder infection (**Methods**). After collapse,

156 barcode sequences associated with different inferred founder infections in the same

157 cells had the same distribution of similarity relationships (edit distances) as random

158 barcodes did. Furthermore, there were (1) similar numbers of UMIs covering the barcode

159 and *EGFP* transcript within the same cells and (2) independence in the number of

160 unique barcodes and barcode-associated RNA counts (**Extended Data Fig. 3a,b**). All

161 data presented hereafter have been computationally collapsed in this way.

162

163 Cells in which we detected at least one viral barcode also tended to have devoted a

164 substantial fraction of their transcription to rabies genes (% of total UMIs mean±sem,

165 15±0.14), compared to cells that did not have a barcode (0.3±0.007%, a rate

166 consistent with background due to ambient cell-free RNA). This suggests that barcode

167 ascertainment was sensitive, selective, and distinguished infected starter cells

168 (n=17,283) from neighboring uninfected (n= 43,533) cells (**Extended Data Fig. 3c**).

169 Putatively infected cells (those cells for which >1% of total UMIs came from rabies

170 genes) for which we failed to detect a barcode tended to have very small RNA profiles (<

171 500 UMIs). Additionally, less than 2% of cells with a barcode and a large RNA profile

172 (>10,000 UMIs) lacked viral loads indicative of infection (<1% rabies RNA), suggesting

173 spurious viral barcode associations were rare.

174

175 Using the above data, we investigated whether the properties of infection differed among

176 starter cell types. We found that infection mainly occurred in glutamatergic neurons,

177 interneurons and astrocytes, and was less frequently observed in other glia types

178 (polydendrocytes and oligodendrocytes), neural precursor cells, and cells undergoing

179 mitosis (**Extended Data Fig. 3e** and **Methods)**. Even among infected cells, analysis

180 revealed clear differences in infection properties: relative to infected glia, infected

181 neurons tended to have more founder infections (unique barcodes mean$\pm$sem: Neurons,

182 1.4$\pm$0.01; Glia, 1.1$\pm$0.01), far more barcoded rabies transcripts detected per RNA profile

183 (Neurons, 97.7$\pm$4.0; Glia, 13.8$\pm$ 1.1), and higher percentages of *EGFP* per RNA profile

184 (Neurons, 1.5$\pm$0.03; Glia, 1.1$\pm$0.05)(**Fig. 2d,e**), revealing previously unknown cell-type-

185 specific properties of rabies virus infection.

186

187 In principle, the combination of multiple founder infections in the same starter cell could

188 help define uCIPs through coupled presynaptic spread, but in practice, cell biological

189 constraints might limit the number of founder infections. To evaluate this, we leveraged

190 the viral barcodes to quantify the multiplicity of infection (MOI) at single-cell resolution

191 and to relate this to the titer of the infecting library (**Extended Data Fig. 3f**). At the

192 lowest titer we tried (MOI, ~0.15), more than 97% of neuron and astrocyte RNA profiles

193 were associated with a single VBC (unique VBCs mean$\pm$sem for neurons/astrocytes:

194 MOI ~0.15, 1.07±0.009/1.03±0.01). Infections with 10-fold higher titer resulted in more

195 multiply infected cells (with two or more viral barcodes)(MOI ~1.5, 1.42±0.01/1.14±0.01).

196 However, we saw only minimal further increases at 100-fold higher titer (MOI ~15,

197 1.6±0.01 / 1.2±0.02). At a biological level, these data suggest an intrinsic asymptote in

198 the number of founder infections individual cells will meaningfully sustain – perhaps, for

199 example, because cell-biological machinery are effectively hijacked by the earliest

200 founders. At an engineering level, these results also suggested rabies virus titer-ranges

201 for efficiently transducing starter cells with multiple viral barcodes.

202

203 Infecting and analyzing large numbers of starter cells in these control, no-spread

204 experiments helped us to better understand many properties of rabies virus infections

205 and barcoded rabies libraries. We compared the EnvA-RV*dG-EGFP*$_{VBC}$ library

206 abundances of 1.29 million barcodes to their 28,755 founder infections distributed across

207 17,283 starter cells (**Fig. 2f**). Critically for later inferences, the abundance of a barcode

208 in the infecting library predicted the number of cells it would infect (**Fig. 2f**). (A few

209 barcodes that appeared to overperform this expectation were flagged for computational

210 removal from future analyses, **Fig. 2f**). In addition, some 6% of infections involved viral

211 barcodes that we had not detected by sequencing the library, presumably because they

212 were present at very low abundance (**Fig. 2f**) (any of these that infected multiple cells in

213 this "no-spread" experiment were also flagged for removal from future analyses).

214 Intriguingly, a small number of barcode pairs consistently appeared together in the same

215 starter cells, even in distinct experiments. Because rabies particles do not have strict

216 genome size limitations[10,16], we reasoned that barcode interdependence might result

217 from concatenated genomes. (These pairs were also flagged and removed from future

218   analyses; **Methods**). These analyses suggest the abundance of barcodes in library

219   genomes has considerable predictive power in estimating the number of starter cell

220   founder infections, but also highlight examples in which individual barcodes or barcode

221   pairs defy expectations. Thus, each SBARRO infecting library should be carefully

222   evaluated in a large number of starter cells, as we describe further below.

223

224

225   **Massively parallel inference of monosynaptic relationships between cells**

226   We next sought to describe cell-type-specific synaptic wiring of an *in vitro* culture. We

227   focused on *in vitro* experiments because such systems 1) have been used extensively to

228   screen for genes and molecules involved in synapse development; 2) can retain features

229   of cell-type-specific connectivity, and 3) have more easily recoverable cells; in our

230   hands, recovery of rabies infected neurons after *in vivo* experiments was inefficient. To

231   increase cell type diversity, we co-cultured cells dissociated from embryonic cortex,

232   striatum and caudal olfactory areas. We sparsely seeded potential starter cells in each

233   culture well by using rAAVs to express TVA and the rabies glycoprotein (G), thus

234   enabling EnvA-mediated rabies founder infections and G-dependent presynaptic spread.

235   (Sparsity minimizes the opportunity for starter cells to become secondarily infected as

236   presynaptic cells, which could in principle support polysynaptic spread). Lastly, after 12

237   days *in vitro*, during a period of prolific synaptogenesis[17], starter cells were transduced

238   with EnvA-RV*dG-EGFP*$_{VBC}$ (MOI, ~1.5)(**Fig. 3a** and **Methods**). After another 72 hours,

239   EGFP fluorescence was observed in putative presynaptic cells, many of which were

240   spatially clustered around each dual-labeled starter cell and were interspersed with large

241   numbers of uninfected cells (**Extended Data Fig. 4a,b**). This spatial pattern of rabies

242   spread was consistent with the idea that the probability of neuronal connectivity scales
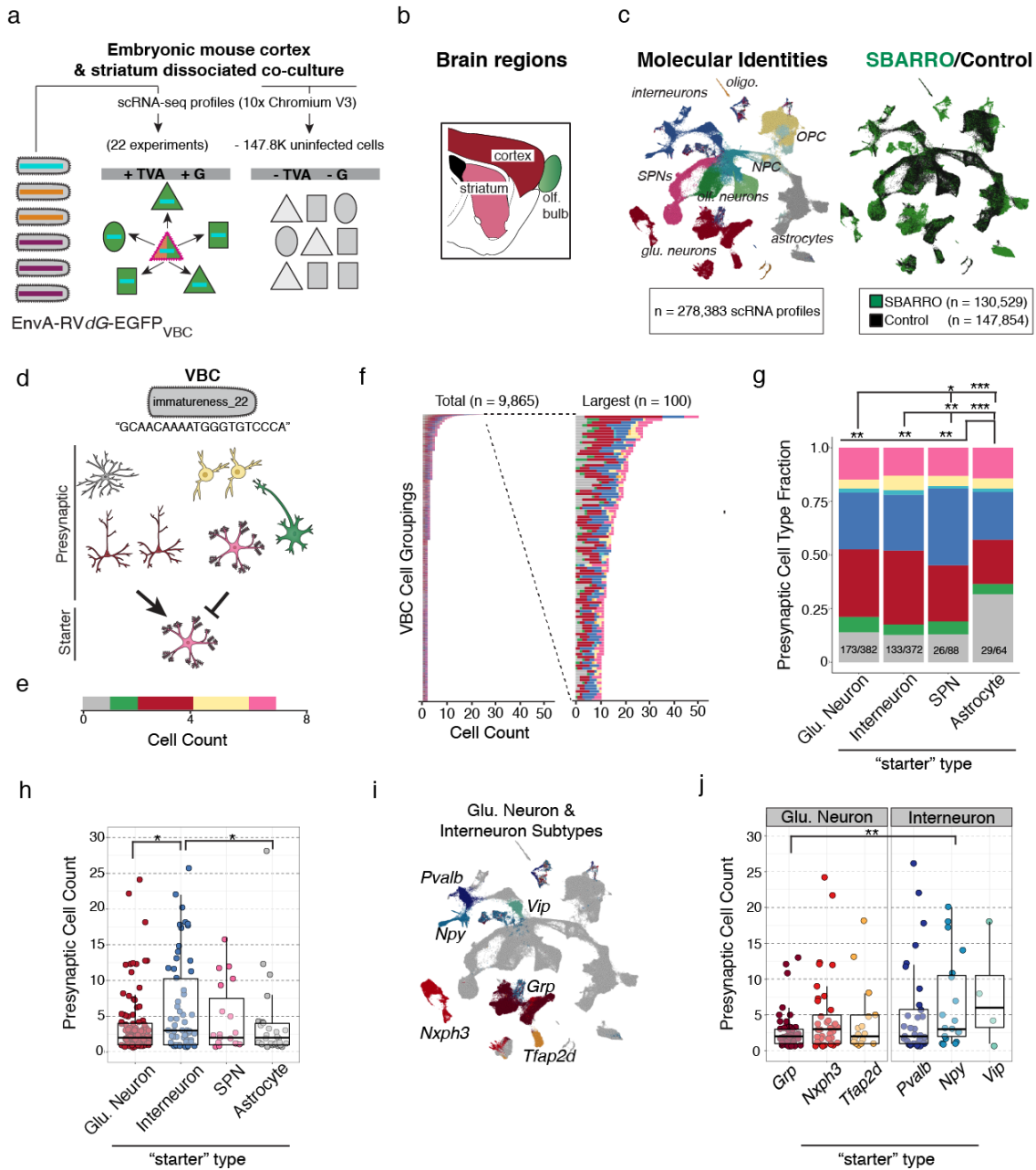
Figure 3

Figure 3. Massively parallel inference of cell-type-specific synaptic connectivity using SBARRO. a. Experimental schematic. The EnvA-RV*dG-EGFP*VBC library transduced starter cells (+TVA/+G) from which individual virion clonally replicate and undergo retrograde, monosynaptic spread into presynaptic cells. scRNA-seq libraries

were prepared from either 1) SBARRO EGFP+ cells (n=23 culture wells from n=3 mouse preparations; n=130.5K scRNA profiles) or 2) preparation-matched control cells (n=147.6K scRNA profiles). **b.** Sagittal mouse 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,* **Extended Data Fig. 5a**) or SBARRO/control status (*right*) following LIGER analysis (**Methods**). **d.** An example SBARRO network inferred through shared expression of VBC assigned the named "immatureness_22". (Names were assigned to each VBC to better track VBC identities; **Methods**). Rabies particles encapsidating the "immatureness_22" genome are rare enough such that transduction of more than one starter cell founder infection is estimated to have a < 1% chance of occurring (**Methods**). The "immatureness_22" network consists of a 7 scRNA profiles with associated molecular identities, including an SPN starter cell and heterogenous collection of putative presynaptic cells (color-coded as in **c**). **e.** Cell-type composition of the inferred "immatureness_22" presynaptic network represented as a horizontal bar plot. **f.** Horizontal bar plots for n=9,865 inferred SBARRO networks with >= 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 inferred presynaptic networks stratified by starter cell type. **g.** Fractional cell-type compositions from inferred presynaptic networks exhibit quantitative differences across starter cell types (* = 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). **h.** Inferred presynaptic network sizes differ by starter cell type. (** = p < 0.05, Wilcoxon Test). **i.** UMAP embedding color-coded and labeled by glutamatergic neuron and interneuron subtype (**Extended Data Fig. 5a,b**). **j.** Inferred presynaptic network sizes by starter cell subtype (** = p < 0.01; Wilcoxon Test).

243   roughly with spatial proximity and suggested that presynaptic networks innervating

244   distinct starter cells were largely non-overlapping. No transduction (EGFP+ cells) was

245   observed without TVA receptor expression, suggesting that all infections entered

246   experiments through starter cells (data not shown).

247

248   We identified cell types in thousands of reconstructed monosynaptic networks, profiling

249   RNA from EGFP+ cells from 23 distinct culture wells (n=3 cell culture replicates). To

250   learn the molecular identities of each cell and determine in what ways the infected

251   population might be different from the total ensemble of cultured cells, we co-clustered

252   scRNA profiles from SBARRO experiments (n=130,529 cells; mean UMIs, 17,622) and

253   uninfected control cells (n=147,854 cells; mean UMIs, 18,117) based on shared host-cell

254   RNA signatures (**Methods**). Cultured cell RNA profiles were from diverse and

255   developmentally dynamic cell populations (**Fig. 3b,c**). We identified four populations of

256   glutamatergic neurons; polydendrocytes; oligodendrocytes; and neural precursor cells

257   (NPCs) developing into astrocytes and several mature GABAergic lineages (including

258   three major interneuron populations, two olfactory-related neuron types, and spiny

259   projection neurons (SPNs); **Extended Data Fig. 5a,b**). Neuronal identity assignments

260   were confirmed by an integrated analysis with scRNA profiles from adult mouse

261   neocortex[18] (**Extended Data Fig. 5c**). Compared to the relative abundance of control

262   cells, rabies-infected cells were enriched among mature interneurons (log2(rv/control) =

263   1.21), SPNs (0.94), glutamatergic neurons (0.83) and astrocytes (0.32), and depleted

264   from developmentally immature cells (NPCs = -2.45; immature neurons = -2.44), mature

265   GABAergic olfactory types (-1.67), oligodendrocytes (-0.94) and polydendrocytes (-

266   0.91)(**Extended Data Fig. 5d**).

267

268 We detected putative synaptic networks as clonal expansion of viral barcodes observed

269 across cells (**Extended Data Fig. 6a**). Paired anatomical/SBARRO datasets suggests

270 that roughly 10% of infected cells entered our single-cell analyses; the missing cells

271 were likely lost or destroyed during physical dissociation and FACS-enrichment or

272 remained unsampled after microfluidics-based RNA barcoding. Thus, synaptic networks

273 are detectable yet contain only a small subset of the cells associated with each network

274 (**Extended Data Fig. 4c,d**).

275

276 We identified starter cells by their TVA expression (**Extended Data Fig. 7a-d** and

277 **Methods**). Presynaptic and starter cells were composed of similar cell types, but starter

278 cells expressed a larger number of unique barcodes (mean±sem: starter, 4.3±1;

279 presynaptic, 2.8±0.007, p < 2.2e-16, Kolmogorov–Smirnov Test; **Extended Data Fig.**

280 **7e,f**), which is expected as some barcodes may fail to transit and infect other cells in the

281 analysis. Comparing FACS-based counts of fluorescently labelled starter or presynaptic

282 cells suggested that starter cells failed to enter our analyses more frequently than

283 presynaptic cells did (2.5 vs 25%; **Extended Data Fig. 4c,d**); this could reflect increased

284 fragility and loss due to prolonged infection, or insufficient ascertainment of recombined

285 TVA mRNAs. Thus, we expect many of our identified synaptic networks to be "orphaned"

286 from their starter cell.

287

288 We developed a statistical framework to filter one or more co-expressed viral barcodes

289 based on the 1) estimated number of founder infections and the 2) barcode abundance

290 in the infecting library (**Methods**). We also excluded barcodes (n=551) or barcode pairs

291 (n=689) that (in the control experiments) infected or co-infected more cells than

292   expected based on their library abundance (**Fig. 2f**; **Extended Data Fig. 6b**; **Methods**).

293   For example, in an experiment estimated to contain 2,484 total founder infections, we

294   observed one example barcode in an SPN starter cell and seven diverse presynaptic

295   cells (**Fig. 3d**). Based on the low abundance of the barcode in the infecting library

296   (frequency = 3.5×10^{-6}), we estimate that this barcode had a <1% chance of participating

297   in more than one founder infection in this experiment. Thus this barcode passed our

298   threshold (of <10%) and defined a uCIP (**Methods**).

299

300   We retained n=1,810 of 5,142 total viral barcodes, which alone or in combination,

301   enabled n=9,865 non-redundant uCIP inferences of synaptic networks with >= 2 cells (n

302   = 21,458 scRNA profiles; **Fig. 3f**). Inferred networks contained 2-52 cells (mean = 3.1,

303   median = 2), consistent with ~10% ascertainment of rabies-infected cells in culture

304   (**Extended Data Fig. 4c,d**). Inferred networks contained predominantly neurons (79%)

305   of diverse types, with smaller contributions from astrocytes (15%) and polydendrocytes

306   (5%).

307

308   To determine whether the cell-type composition or the size of presynaptic networks

309   varied across postsynaptic cell types, we focused on the 3.7% of networks with an

310   identified starter cell (n=365 of 9,865 total networks). Presynaptic cell types differed

311   according to postsynaptic cell type, as pair-wise comparisons suggested quantitative

312   differences in presynaptic cell-type proportions between astrocytes versus neurons (p =

313   0.007 - 0.001, Chi-Square Test) and glutamatergic neurons and interneurons versus

314   SPNs (p = 0.01 and 0.0002), but not between glutamatergic neurons and interneurons (p

315   > 0.05, Chi-Square Test; **Fig. 3g**).

316   The sizes of inferred presynaptic networks exhibited variance that was partially

317   explained by postsynaptic cell type (p = 0.048, Kruskal-Wallis Test). Pair-wise

318   comparisons revealed that glutamatergic neurons and astrocytes tended to have smaller

319   presynaptic networks than interneurons did (p = 0.018 – 0.022, Wilcoxon Test), while

320   SPN presynaptic networks did not detectably differ from those of other cell types (n=109

321   glutamatergic neurons, mean±sem presynaptic cells= 3.6±0.4; n=29 astrocytes,

322   3.7±1.02; n=61 interneurons, 6.7±1.0; n=12 SPNs, 4.6±1.1; **Fig. 3h**). Differences in

323   neuronal presynaptic network size appeared to be driven in part by cell subtypes (**Fig.**

324   **3i,j** and **Extended Data Fig. 6c,d**); while neuronal subtype categories did not, as a

325   whole, rise to predictive significance in explaining variance in presynaptic network size

326   (p = 0.08, Kruskal-Wallis test), paired comparisons revealed that *Grp+* glutamatergic

327   neurons and *Npy+* interneurons tended to have small (n=50, mean±sem = 2.7±0.37

328   cells) and large (n=20, mean = 8±1.95 cells) presynaptic networks (p = 0.004, Wilcoxon

329   Test), respectively. These results indicate that the number and molecular composition of

330   putative presynaptic cells in an inferred network are qualitatively similar across

331   postsynaptic cell types at an early, promiscuous stage of synaptogenesis *in vitro*, but

332   highlight important exceptions in which postsynaptic cell type biases the number and

333   classes of putative presynaptic partner cells. Differences in the number of presynaptic

334   partner cells might relate to dendritic size differences *in vivo*. For example, compared to

335   adult mouse neocortex, *Grp+* glutamatergic neurons are most similar to L2/3 IT and L4/5

336   IT subtypes found in superficial cortical layers, which tend to have small dendritic arbors,

337   while *Nxph3+* glutamatergic neurons are most similar to L6b, L5 NP and L6 CT subtypes

338   found in deeper cortical layers, which tend to have larger dendritic arbors[18,19] (**Extended**

339   **Data Fig. 5c**).

340

341

342 **Postsynaptic RNAs associated with presynaptic network properties**

343 The formation and selective stabilization of synapses is shaped by competitive

344 processes driven by molecular variation within[20] and across cell populations[21], yet many

345 of the molecules remain unknown and incompletely understood. We sought to use the

346 data from these experiments – in which synaptic connectivity inferences and molecular

347 properties were measured in the same cells – to analyze how molecular variation

348 associated with the properties of cell-type-specific networks.

349

350 We first asked whether presynaptic network size was explained by infection magnitude

351 or innate immune response in starter cells, since these properties of infection could skew

352 the results (**Fig. 4a**). We separated starter cells into two groups based on presynaptic

353 network size, ranging from networks of 2–4 ascertained cells ("small") or 7–52

354 ascertained cells ("large") and four groups based on starter cell type (**Fig. 4b**). We

355 compared both the viral load (the fraction of total cellular mRNAs from the rabies virus

356 genome) and an aggregate innate immunity expression score (n=564 genes[22]) across

357 these groups (n=144 scRNA profiles; **Extended Data Fig. 8a**). We found that, while both

358 infection metrics varied by starter cell type (viral load, $p = 0.05$; innate immune

359 expression score, $p = 5.2 \times 10^{-12}$, Two-way ANOVA Test), they did not associate with the

360 presynaptic network size (viral load, $p = 0.10$; innate immune expression score, $p =$

361 0.16).

362

363 We next sought to find genes whose expression levels in starter cells associated with

364 the number of inferred presynaptic partner cells (**Fig. 3j)**. We hypothesized that
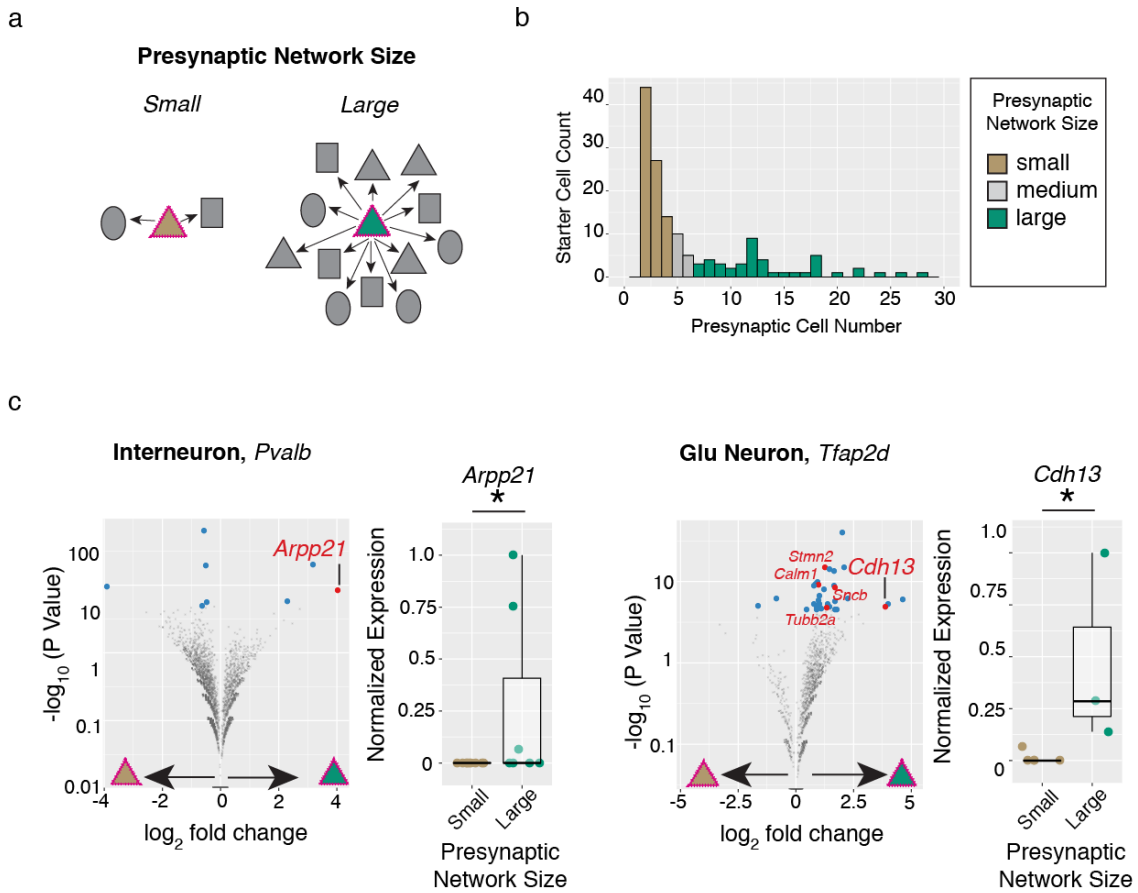
Figure 4

Figure 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 *Tfap2d* 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 ($p < 0.05$, blue dots) were further tested for differences in single-cell scaled expression (Wilcoxon Test; **Methods**) and those that pass this additional test ($p < 0.05$, red dots) are labeled. *Right*, normalized expression levels.

365   differences in the expression of genes promoting or restricting synaptogenesis or

366   dendrite growth influence the number of presynaptic cells innervating each starter cell.

367   We focused our comparisons at the most granular cell subtype level and narrowed our

368   testing to those genes sufficiently expressed and skewed in aggregate across

369   presynaptic network size groups (**Methods**). We used permutation to create negative-

370   control distributions in which each starter cell RNA profile was replaced by a randomly

371   selected presynaptic RNA profile of the same type.

372

373   Across eight starter cell subtypes, 13 genes exhibited differential expression across

374   presynaptic network size categories ($p < 0.05$, Wilcoxon Test; **Fig. 4c**; **Extended Data**

375   **Fig. 8b,c**). Though this did not exceed the number of genes nominated in permuted

376   analyses (mean±sem = 15.6±1.3), independent biological evidence strongly supported

377   roles for two of the most strongly differentially expressed genes, both of which were

378   more highly expressed in starter cells with large networks (relative to cells with small

379   networks) and have described roles in promoting dendritic growth or synapse formation

380   through developmental loss-of-function or ectopic overexpression experiments. *Arpp21* –

381   which was upregulated ~16 fold in postsynaptic *Pvalb*+ interneurons with large

382   presynaptic networks – encodes an RNA binding protein that promotes dendritic growth

383   by activating translation of target RNAs and whose cell-to-cell dynamic range might be

384   extended due to intronic-encoded inhibitory microRNA[23]. *Cdh13* – which was

385   upregulated ~15 fold in postsynaptic *Tfap2d*+ glutamatergic neurons with large

386   presynaptic networks – encodes an atypical protocadherin, one of four genes previously

387   identified as driving synaptogenesis in a large-scale neuronal RNAi screen[12].

388

389

390 **RNAs correlated with rabies virus transmission implicate synaptic function**

391 Accurate interpretation of how rabies-inferred synaptic networks relate to actual synaptic

392 connectivity and function is critically limited by our incomplete understanding of the

393 molecules and cellular processes through which rabies enters, exits, and interacts with

394 diverse host brain cell types. To determine which RNAs and biological pathways

395 contribute to rabies transmission, we leveraged asynchronous development and variable

396 rabies transmission in cultured cells to identify gene expression patterns that correlated

397 with increased infectivity along the developmental trajectory stretching from neural

398 precursor cells into mature SPNs (**Fig. 5a,b**). We strictly ordered each of the 32,503

399 scRNA profiles in pseudotime using Monocle3[24] and confirmed the expected

400 developmental processes through Gene Ontology Biological Pathway (GOBP)[25,26]

401 enrichment analysis of co-regulated genes (~25% of the coding genome; n= 7,844

402 genes; **Extended Data Fig. 9a**).

403

404 We identified n=3,309 genes with RNA levels that correlated (r > 0.75) with increased

405 rabies transmission (**Fig. 5d,e** and **Methods**). Interestingly, *Ncam1* mRNA was the only

406 one of four described rabies receptors[8,27,28] with appreciable expression in these

407 experiments, and appeared in cells before high rates of infectivity, suggesting NCAM1

408 protein alone is not sufficient for rabies transmission (**Fig. 5d**). To discover which cellular

409 processes might endow infectivity, we performed GOBP with the gene set we identified

410 and compared the results to control gene sets sampled at random or from expression-

411 matched mature SPN profiles. We identified selective enrichments in 1) "mitochondrial

412 respiratory chain complex assembly"; 2) "phagosome maturation"; and 3) "positive

413 regulation of synaptic transmission", which were absent from control gene sets (**Fig. 5d**).

414 To refine which synaptic processes were implicated in infectivity, we queried synaptic

Figure 5
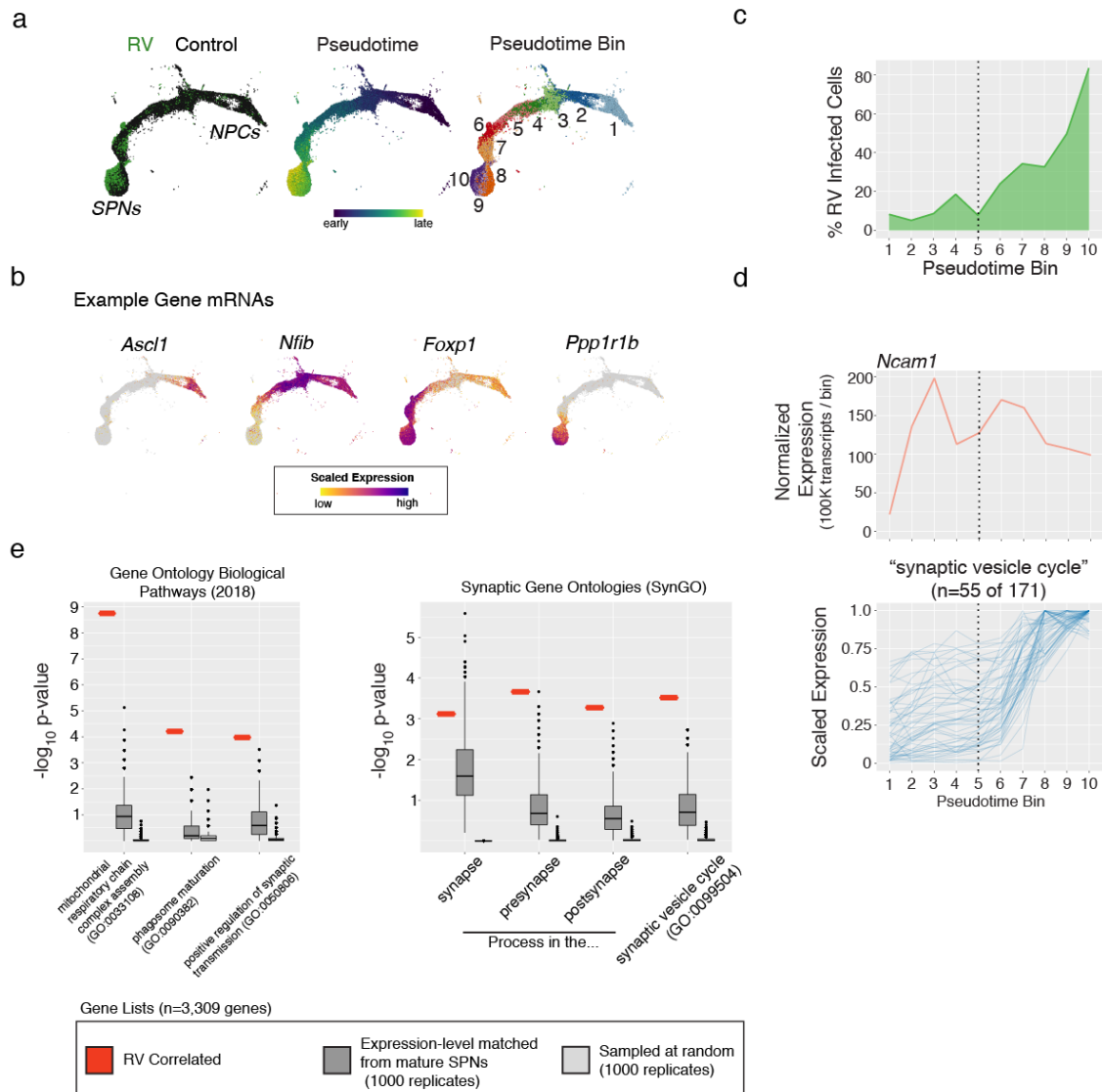
Figure 5. The developmental emergence of rabies transmission co-occurs with the maturation of synaptic function. a. UMAP embedding of scRNA profiles (n=32,503) along a trajectory of development from immature neural precursor cells (NPCs) to mature spiny projection neurons (SPNs). *Left*, color-coded by rabies virus infected (ie SBARRO; n=8,837 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 [8], *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=3,309 genes total). **e.** Gene Ontology Biological Pathways (GOBP) and Synaptic Gene Ontology (SynGO) analyses for rabies-infectivity correlated genes (n=3,309; **Methods**). *Left*, analyses were conducted on the correlated gene list (red) as well as two sets of control genes (each with 1,000 replicates of n=3,309 genes). In the "Expression-matched" set (dark grey), 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 grey), genes were selected at random from those for which expressed RNA was detected. 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.

415 gene ontologies[29] and identified selective enrichment for "synaptic vesicle cycle" (for

416 which n=55 of 171 genes were correlated with infectivity). This analysis suggests that, in

417 addition to the expression of viral entry receptors, operational synaptic transmission is

418 critical for inter-cell rabies transmission and nominates specific genes implicated in the

419 onset of synaptic transmission and rabies entry (**Extended Data Fig. 9b**).

420

421

422 **DISCUSSION**

423 Our understanding of how synaptic networks emerge during development and how they

424 are regulated by genetic and biological programs will benefit from measurements of

425 synaptic connections that are systematic, quantitative, and connected to detailed

426 molecular profiles of individual cells. Comprehensive characterization of the synaptic

427 organization of neural circuits is challenging with current electrophysiological and

428 anatomical methods, due to the small sizes of synapses, expansive geometries of axons

429 and dendrites, and lack of knowledge of the cell subtypes involved. Such limitations

430 have tended to separate synaptic-network biology from other subfields of neuroscience

431 that are adopting highly parallel approaches for characterizing molecular repertoires[30-33]

432 or neural activity in many individual cells[34].

433

434 Here we demonstrate that synaptic networks can be reconstructed from scRNA-seq

435 data, thus allowing direct connectivity relationships to be inferred across thousands of

436 individual cells for which genome-wide RNA expression has also been ascertained. Our

437 data suggest that, during synaptogenesis *in vitro*, connectivity is shaped by cell type in a

438 quantitative rather than qualitative way. Individual starter cells had considerable variance

439 in their number of presynaptic partners, which appeared unrelated to the degree of

440   infection or innate immune response, but partially explained by neuron type and gene

441   expression patterns. We found that *Arpp21* and *Cdh13* had higher expression within

442   starter cells with more presynaptic partners. Interestingly, *Cdh13* – an atypical

443   transmembrane protein of the Cadherin superfamily – was previously identified as a key

444   postsynaptic gene driving both excitatory and inhibitory synaptogenesis through a

445   systematic RNAi screen[12]. In addition, *Arpp21* overexpression or knock-out

446   bidirectionally controls the size and complexity of pyramidal neuron dendrites during

447   postnatal development, likely by potentiating the translation of bound mRNA species that

448   promote dendritogenesis[23]. These proof-of-concept observations suggest that extant

449   molecular heterogeneity may associate with different properties of a given cell's

450   presynaptic network and that SBARRO analyses are a means to access and quantify

451   such relationships.

452

453   We designed SBARRO to be adaptable to emerging single-cell genomic technologies.

454   For example, methods enabling single-cell spatial transcriptomics[35-37] or *in situ*

455   sequencing[38,39] will allow the locations and anatomical properties of SBARRO cells to be

456   mapped *in vivo* without cell loss. Moreover, long-read RNA isoform sequencing[40,41] could

457   address long-standing hypotheses for how alternative splicing helps generate an

458   extracellular adhesion code between synaptically connected cells within and across cell

459   types[42].

460

461   Unknown features of rabies cell biology represent a current limitation in the interpretation

462   of SBARRO datasets. A detailed understanding of how our inferred, digital monosynaptic

463   relationships relate to extraordinarily diverse and highly dynamic synaptic structures

464   requires a comprehensive description of how rabies interacts with and transits between

465   host brain cells of different types. On one hand, previous studies provide direct and

466   circumstantial evidence that suggest, at least to a first approximation, that rabies

467   transmission events are selective for synapses made directly onto infected neurons[43,44].

468   Among postsynaptic neurons of the same class with spatially intermixed dendrites,

469   presynaptic labeling respects synapse-selective motor arcs in the spinal cord. Similarly,

470   in primary visual cortex, intermingled layer 2/3 glutamatergic neurons distinguished by

471   firing properties to visual cues, appear to inherit those selective properties from

472   presynaptic cells labeled by rabies infection[45,46]. On the other hand, the efficiency of

473   rabies transmission can be very low for certain cell-type-specific axons[47] and appears to

474   be modulated by presynaptic firing rate[48,49]. Our correlative molecular data suggest

475   developmentally mature presynaptic function is critical for rabies uptake in neurons (**Fig.**

476   **5**). However, the extent to which rabies egress and entry exclusively use synapse-

477   associated processes; occur through direct synaptic contacts; and are affected by neural

478   activity across diverse brain circuits, all remain to be firmly established[48]. Moreover, it

479   will be necessary to study the ways in which infection alters host cells' molecular

480   programs, as these alterations could affect synapse-associated processes.

481

482   Rabies infection of non-neuronal cell classes, such as astrocytes and polydendrocytes,

483   is a minor yet clear feature of our *in vitro* experiments and is also observed *in vivo*

484   [10,52,53]. While both cell types interact intimately with synapses, especially during

485   development, more experiments are necessary to understand the molecular

486   mechanisms underlying rabies transmission across non-neuronal classes. Single-cell,

487   single-virion inferences of these interactions may offer valuable insight: RABID-seq[6]

488   analyses suggest rabies can be transmitted from infected astrocytes into presumed

489   physically-adjacent microglia and that specialized host cell signatures are associated

490   and detectable with these interactions. What role synapses play in these glial

491   interactions is not clear. (Microglia were not present in our experiments *in vitro*).

492

493   Mammalian synaptogenesis is particularly challenging to study with traditional methods

494   due to the many cell types and molecules involved, its protracted nature in space and

495   time, and intrinsic noise that arises from being a competitive, cell-to-cell process. By

496   facilitating connectivity inferences and RNA sampling from the same individual cells, we

497   hope that fast, scalable, all-molecular approaches such as SBARRO – which may be

498   eventually deployed in non-destructive ways[50] – can complement established

499   connectomic technologies based on super-resolution imaging of synaptic anatomy.

500

501

502   **DATA AVAILABILITY.** The sequencing data reported in this paper are in the process of

503   being uploaded to GEO. A GEO accession number will be provided upon completion.

504

505   **CODE AVAILABILITY**. Software and core computational analysis to align and process

506   scRNA-seq reads are freely available: https://github.com/broadinstitute/Drop-

507   seq/releases. Other custom code available by request.

508

514

515 **AUTHOR CONTRIBUTIONS**

516 AS conceived the idea, designed and supervised experiments, analyzed the data and

517 drafted the manuscript. Plasmid barcoding, CV and AS. Rabies virus packaging, KWH,

518 AP, CV, AS and BLS. Cell culture, KS and CS. scRNA-seq experiments/analysis and

519 molecular protocols, CS, KS, HS and AS. Algorithm development, SAM, JN and AS.

520 Analysis software, SK. Computational support, AW. Manuscript preparation, AS and

521 SAM with input from other authors.

522

523 **AUTHOR INFORMATION.** Correspondence and request for materials should be

524 addressed to AS or SAM.

525

528

529 **SUPPLEMENTARY INFORMATION**

530

531 **Supplementary Table 1.** Oligonucleotide Guide

532

533 **CORRESPONDENCE**

534 Arpiar Saunders (saundear@ohsu.edu); Steve McCarroll

535 (mccarroll@genetics.med.harvard.edu)

536

537

538

REFERENCES

1. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53,** 639–647 (2007).

2. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14,** 865–868 (2017).

3. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* **40,** 181–15 (2018).

4. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564,** 219–224 (2018).

5. Kim, D.-W. *et al.* Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. *Cell* **179,** 713–728.e17 (2019).

6. Clark, I. C. *et al.* Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372,** (2021).

7. Ugolini, G. Advances in viral transneuronal tracing. *J Neurosci Methods* **194,** 2–20 (2010).

8. Lafon, M. Rabies virus receptors. *J NeuroVirology* **11,** 82–87 (2005).

9. 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).

10. Ghanem, A. & Conzelmann, K.-K. G gene-deficient single-round rabies viruses for neuronal circuit analysis. *Virus Research* **216,** 41–54 (2016).

11. Mebatsion, T., Konig, M. & Conzelmann, K. K. Budding of rabies virus particles in the absence of the spike glycoprotein. *Cell* **84,** 941–951 (1996).

12. Paradis, S. *et al.* An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. **53,** 217–232 (2007).

564 13. Sharma, K. *et al.* High-Throughput Genetic Screen for Synaptogenic Factors:

565 Identification of LRP6 as Critical for Excitatory Synapse Development. *Cell*

566 *Reports* **5,** 1330–1341 (2013).

567 14. Williams, M. E. *et al.* Cadherin-9 Regulates Synapse-Specific Differentiation in the

568 Developing Hippocampus. *Neuron* **71,** 640–655 (2011).

569 15. Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-

570 deleted rabies viruses for monosynaptic tracing and high-level gene expression in

571 neurons. *Nat Protoc* **5,** 595–606 (2010).

572 16. Huang, A. S. & Baltimore, D. Defective viral particles and viral disease processes.

573 *Nature* **226,** 325–327 (1970).

574 17. Ichikawa, M., Muramoto, K., Kobayashi, K., Kawahara, M. & Kuroda, Y. Formation

575 and maturation of synapses in primary cultures of rat cerebral cortical cells: an

576 electron microscopic study. *Neurosci Res* **16,** 95–103 (1993).

577 18. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and

578 hippocampal formation. *Cell* **184,** 3222–3241.e26 (2021).

579 19. Kanari, L. *et al.* Objective Morphological Classification of Neocortical Pyramidal

580 Cells. *Cereb Cortex* **29,** 1719–1735 (2019).

581 20. Kwon, H.-B. *et al.* Neuroligin-1–dependent competition regulates cortical

582 synaptogenesis and synapse number. *Nat Neuro* **15,** 1667–1674 (2012).

583 21. Berns, D. S., DeNardo, L. A., Pederick, D. T. & Luo, L. Teneurin-3 controls

584 topographic circuit assembly in the hippocampus. *Nature* **554,** 328–332 (2018).

585 22. Breuer, K. *et al.* InnateDB: systems biology of innate immunity and beyond--

586 recent updates and continuing curation. *Nucleic Acids Res* **41,** D1228–33 (2013).

587   23.   Rehfeld, F. *et al.* The RNA-binding protein ARPP21 controls dendritic branching

588   by functionally opposing the miRNA it hosts. *Nature Communications* **9,** 1–13

589   (2018).

590   24.   Cao, J. *et al.* The single-cell transcriptional landscape of mammalian

591   organogenesis. *Nature* **566,** 1–31 (2019).

592   25.   Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene

593   Ontology Consortium. *Nat Genet* **25,** 25–29 (2000).

594   26.   The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still

595   GOing strong. *Nucleic Acids Res* **47,** D330–D338 (2019).

596   27.   Thoulouze, M. I. *et al.* The neural cell adhesion molecule is a receptor for rabies

597   virus. *J Virol* **72,** 7181–7190 (1998).

598   28.   Wang, J. *et al.* Metabotropic glutamate receptor subtype 2 is a cellular receptor for

599   rabies virus. *PLoS Pathog* **14,** e1007189 (2018).

600   29.   Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge

601   Base for the Synapse. *Neuron* **103,** 217–234.e4 (2019).

602   30.   Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the

603   Adult Mouse Brain. *Cell* **174,** 1015–1030.e16 (2018).

604   31.   Krienen, F. M. *et al.* Innovations present in the primate interneuron repertoire.

605   *Nature* **586,** 1–29 (2020).

606   32.   Gouwens, N. W. *et al.* Integrated Morphoelectric and Transcriptomic Classification

607   of Cortical GABAergic Cells. *Cell* **183,** 935–953.e19 (2020).

608   33.   Scala, F. *et al.* Phenotypic variation of transcriptomic cell types in mouse motor

609   cortex. *Nature* **18,** 1–34 (2020).

610 34. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-
611   photon mesoscope with subcellular resolution for in vivo imaging. *eLife*
612   doi:10.7554/eLife.14472.001

613 35. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-
614   wide expression at high spatial resolution. *Science* **363,** 1463–1467 (2019).

615 36. Stickels, R. R. *et al.* Sensitive spatial genome wide expression profiling at cellular
616   resolution. **92,** 1168–13 (2020).

617 37. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by
618   RNA seqFISH. *Nature* **11,** 1–24 (2019).

619 38. Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection
620   Using In Situ Sequencing. *Cell* **179,** 772–786.e19 (2019).

621 39. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell
622   transcriptional states. *Science* **361,** 5691–5618 (2018).

623 40. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in
624   thousands of cerebellar cells. *Nat Biotechnol* **36,** 1197–1202 (2018).

625 41. Ray, T. A. *et al.* Comprehensive identification of mRNA isoforms reveals the
626   diversity of neural cell-surface molecules with roles in retinal development and
627   disease. *Nature Communications* **11,** 1–20 (2020).

628 42. Südhof, T. C. Synaptic Neurexin Complexes: A Molecular Code for the Logic of
629   Neural Circuits. *Cell* **171,** 745–769 (2017).

630 43. Wickersham, I. R., Finke, S., Conzelmann, K.-K. & Callaway, E. M. Retrograde
631   neuronal tracing with a deletion-mutant rabies virus. *Nat Methods* **4,** 47–49
632   (2006).

633 44. Ren, S.-Q., Li, Z., Lin, S., Bergami, M. & Shi, S.-H. Precise Long-Range

634   Microcircuit-to-Microcircuit Communication Connects the Frontal and Sensory

635   Cortices in the Mammalian Brain. *Neuron* **104,** 1–17 (2019).

636 45. Rossi, L. F., Harris, K. D. & Carandini, M. Spatial connectivity matches direction

637   selectivity in visual cortex. *Nature* **588,** 648–652 (2020).

638 46. Wertz, A. *et al.* Single-cell-initiated monosynaptic tracing reveals layer-specific

639   cortical network modules. *Science* **349,** 70–74 (2015).

640 47. Albisetti, G. W. *et al.* Identification of Two Classes of Somatosensory Neurons

641   That Display Resistance to Retrograde Infection by Rabies Virus. *J Neurosci* **37,**

642   10358–10371 (2017).

643 48. Rogers, A. & Beier, K. T. Can transsynaptic viral strategies be used to reveal

644   functional aspects of neural circuitry? *J Neurosci Methods* **348,** 109005 (2020).

645 49. Beier, K. T. *et al.* Rabies screen reveals GPe control of cocaine-triggered

646   plasticity. *Nature* **549,** 345–350 (2017).

647 50. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR–Cas encoding

648   of a digital movie into the genomes of a population of living bacteria. *Nature* **337,**

649   1–14 (2017).

650 51. Schnell, M. J., Mebatsion, T. & Conzelmann, K. K. Infectious rabies viruses from

651   cloned cDNA. *EMBO J* **13,** 4195–4203 (1994).

652 52. Reardon, T. R. *et al.* Rabies Virus CVS-N2c deltaG Strain Enhances Retrograde

653   Synaptic Transfer and Neuronal Viability. *Neuron* **89,** 711–724 (2016).

654 53. Barr, J. N., Whelan, S. P. & Wertz, G. W. cis-Acting signals involved in termination

655   of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the

656   U7 signal for polyadenylation. *J Virol* **71,** 8718–8725 (1997).

657   54.   Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered

658   CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

659   55.   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–

660   21 (2012).

661   56.   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of

662   Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).

663   57.   Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the

664   human cerebral cortex. *Nature* **570,** 1–18 (2019).

665   58.   Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts

666   Features of Brain Cell Identity. *Cell* **177,** 1873–1887.e17 (2019).

667   59.   Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177,** 1888–

668   1902.e21 (2019).

669   60.   Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell

670   transcriptomic data across different conditions, technologies, and species. *Nat*

671   *Biotechnol* **36,** 411–420 (2018).

672   61.   Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX Switch Targets

673   Channelrhodopsin-2 to Multiple Cell Types for Imaging and Long-Range Circuit

674   Mapping. *J Neurosci* **28,** 7025–7030 (2008).

675   62.   Miyamichi, K. *et al.* Dissecting Local Circuits: Parvalbumin Interneurons Underlie

676   Broad Feedback Control of Olfactory Bulb Output. *Neuron* **80,** 1232–1245 (2013).

677   63.   Lynn, D. J. *et al.* InnateDB: facilitating systems-level analyses of the mammalian

678   innate immune response. *Molecular Systems Biology* **4,** 218 (2008).

679   64.   Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed

680   by pseudotemporal ordering of single cells. *Nat Biotechnol* **32,** 381–386 (2014).

681   65.   Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version

682   14: more genomes, a new PANTHER GO-slim and improvements in enrichment

683   analysis tools. *Nucleic Acids Res* **47,** D419–D426 (2018).

684   66.   Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list

685   enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).

686

687 **Supplementary Table 1.** Oligonucleotide Guide

688

publication name	sequence(5'->3')	description
B19_barcode_F	ACCTGTGGCGCCACTGCNNNNNNNNNNCTTTTCAGTCGAGAAAAAAACATTAGATCAGAAGAACAAC	Barcoding primers for pSPBN-GFP plasmid
B19_barcode_R	ACCTGTGGCGCCACCGGNNNNNNNNNNTAGCTTACTTGTACAGCTCGTCCATGCCGAGAGTGATC	Barcoding primers for pSPBN-GFP plasmid
B19_UMI_F	AAGCAGTGGTATCAACGCAGAGTACNNNNNNNNNNNNNATCACTCTCGGCATGGACGAGCTGTACAAGTAA	UMI hybridization probe for barcode counting from rabies plasmids and RNA genomes
P5-TSO_Hybrid	AATGATACGGCGACCACCGAGATCTACACGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGT*A*C	P5-containing primer for amplification/sequencing of RV plasmids/genomes
P7i1-L5UTR_seq	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTTGCCAGTTGTTCTTCTGATCTAATG	P7-containing primer for amplification/sequencing of RV plasmids/genomes
Read1CustomSeqB	GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC	Custom Illumina Read1 primer
P7i1-GFP	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCATGGACGAGCTGTACAAGTAAGCTA	P7-containing primer for amplification/sequencing of RV mRNA after 10x capture
P5-10x_Hybrid	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	P5-containing primer for amplification/sequencing of mRNA after 10x capture
P7i1-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i2-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i3-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i4-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i5-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i6-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i7-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i8-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i9-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA
P7i10-TCB_CreOn	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCACAACGAGGACTACACCATCGT	P7-containing primer for amplification/sequencing of recombined rAAV TVA-mCherry mRNA

689

690

691

EXTENDED DATA FIGURES

693   **Extended Data Figure 1**

a

b

c

694

695   **Extended Data Figure 1. Molecular workflows for PCR-based plasmid barcoding**

696   **and UMI-based VBC quantification across plasmids, anti-sense rabies virus**

697   **genomes and mRNAs. a.** Schematic showing the VBC cassette integrated into the

698   rabies genome. The 20 bp VBC consists of two 10 bp barcodes linked by a restriction

699   cassette and sits 13 bps away from the rabies polyadenylation signal. **b.** Diagram

700   describing the novel PCR-based strategy for barcoding DNA plasmids as applied to

701   the cDNA version of the SAD-B19 RV*dG*-EGFP genome (pSPBN-4GFP). Forward and

702   reverse primers containing 10 bp randomers - tailed by 17 bps of containing adaptor

703 sequence, the PluTI site and protective bases – are targeted to adjacent regions of the

704 template plasmid. Amplification results in a collection of linear dsDNA molecules each of

705 which contain a unique combination of terminal 10 bp barcodes, which are then

706 circularized and enriched (versus both template plasmid and remaining linear products)

707 through a single-pot reaction. **c.** Illumina sequencing-based strategy for quantifying

708 plasmid and anti-sense genomic barcodes (*left*) or barcoded *EGFP* mRNA (*right*) using

709 unique molecular identifiers (UMIs). UMI-containing oligonucleotides (UMI = 12 bp

710 randomer) with a shared PCR handle are hybridized adjacent to and then polymerized

711 through the barcode cassette on ssDNA or anti-sense RNA genomes. The UMI-tagged

712 ssDNA molecules are then selectively PCR amplified using primers that contain Illumina

713 P5 and P7 sites and sequenced on an Illumina flow cell such that 110 Read 1 cycles

714 cover the barcode cassette. Barcoded *EGFP* mRNA is selectively amplified from sc-

715 cDNA using P5 and P7 containing primers and sequenced on an Illumina flowcell such

716 that 26-28 Read 1 cycles that capture the 16 bp Cell Barcode and 10-12 bp UMI

717 introduced by 10x Chromium v2 or v3 chemistry and 98 Read 2 cycles which extend

718 through the barcode cassette (**Methods)**.

719

720

721

722

723

724

725

726

727
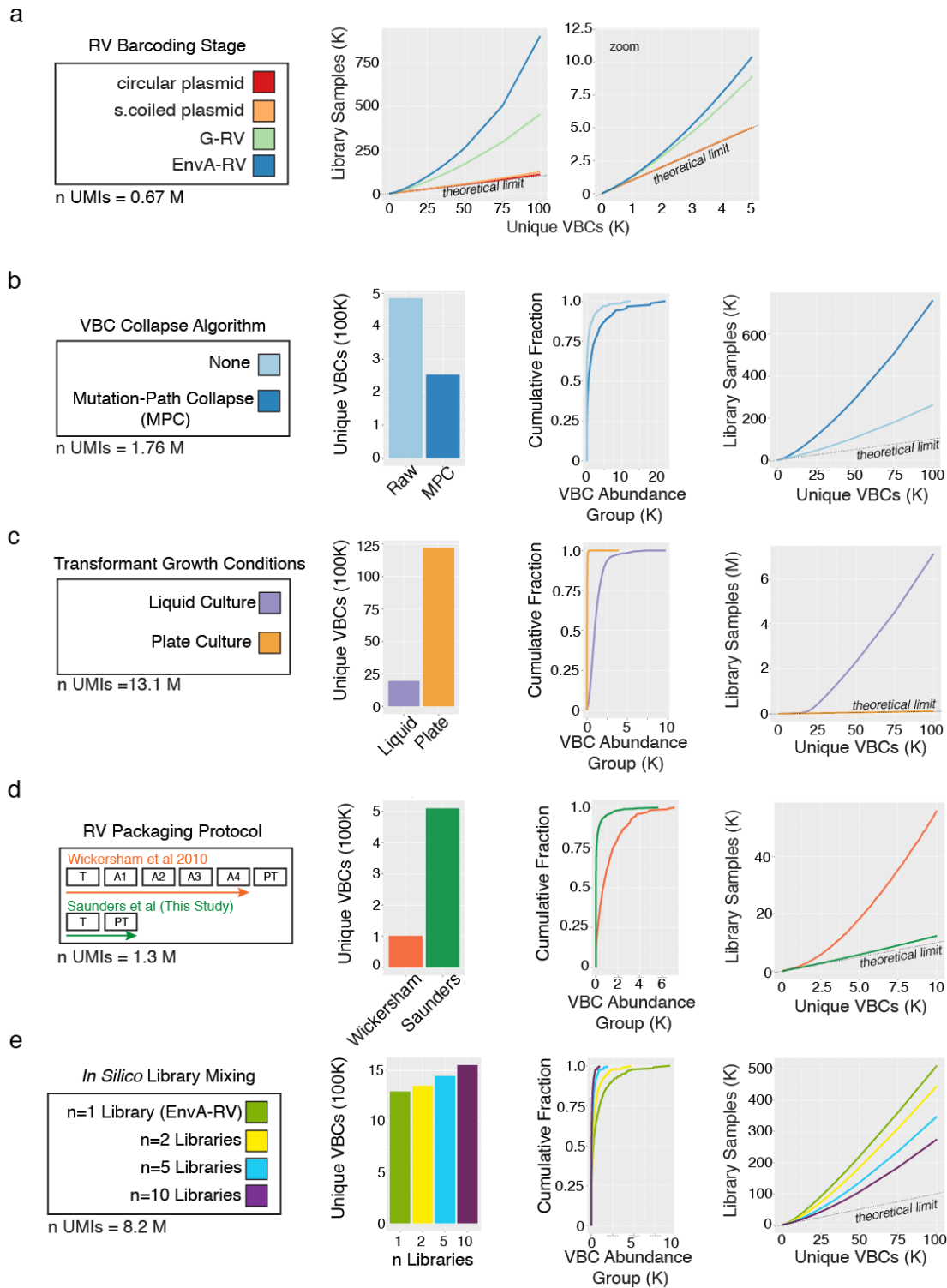
Extended Data Figure 2

730   **Extended Data Figure 2. Accurate and systematic quantification of VBC**

731   **abundances guides optimized protocols for plasmid barcoding and barcoded**

732 **rabies virus packaging. a.** Longitudinal assessment of VBC diversity across each

733 stage of rabies virus packaging protocol, as assayed through the library sampling

734 procedure in which total library samples are plotted against the number of unique

735 VBCs ascertained from each sample. (Companion data to **Fig. 1d,e**. Plot includes

736 data from **Fig. 1f** along with additional conditions). The dotted line shows maximum

737 theoretical diversity (in which every drawn VBC is unique). **b-e**. Quantification of VBC

738 abundance and diversity across various protocol conditions (sampled with equivalent

739 UMIs, *far left*) by plotting (*from left to right*) total unique VBCs; cumulative distribution

740 of UMIs by abundance group (as in **Fig. 1e**); and number of unique VBCs ascertained

741 from a given number of library samples (as in **a** above). **b.** The effect of "mutation-

742 path collapse" (MPC), an informatic approach implemented to help account for

743 artifactual inflation of barcodes driven by mutations to barcode sequences incurred

744 during library amplification or sequencing (**Methods**). **c.** The effect of *E.coli* growth

745 conditions (plated or liquid culture) after transformation with circular barcoded plasmid

746 library. Barcodes were sampled from super-coiled plasmid DNA. **d.** The effect of

747 rabies virus packaging protocols, comparing the widely used Wickersham et al. 2010

748 protocol versus the barcode diversity optimized protocol reported in this study

749 (Saunders et al. 2021). Barcodes were sampled from anti-sense genomes extracted

750 from EnvA-pseudotyped libraries. **e**. The effect of combining different numbers of

751 independent and equivalently diverse barcoded EnvA-pseudotyped libraries *in silico*

752 (**Methods**).

753

754

755

756

757

Extended Data Figure 3

759
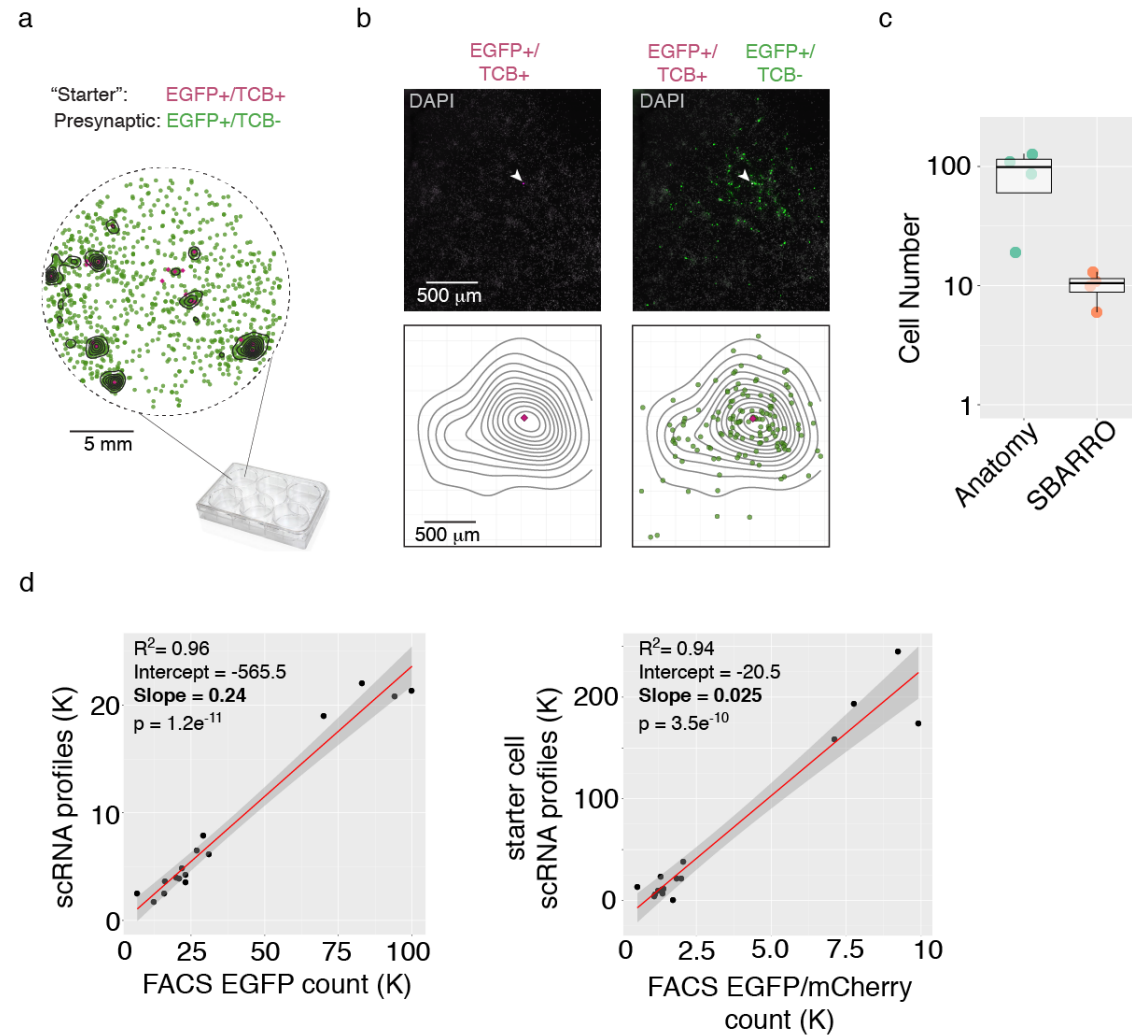
760
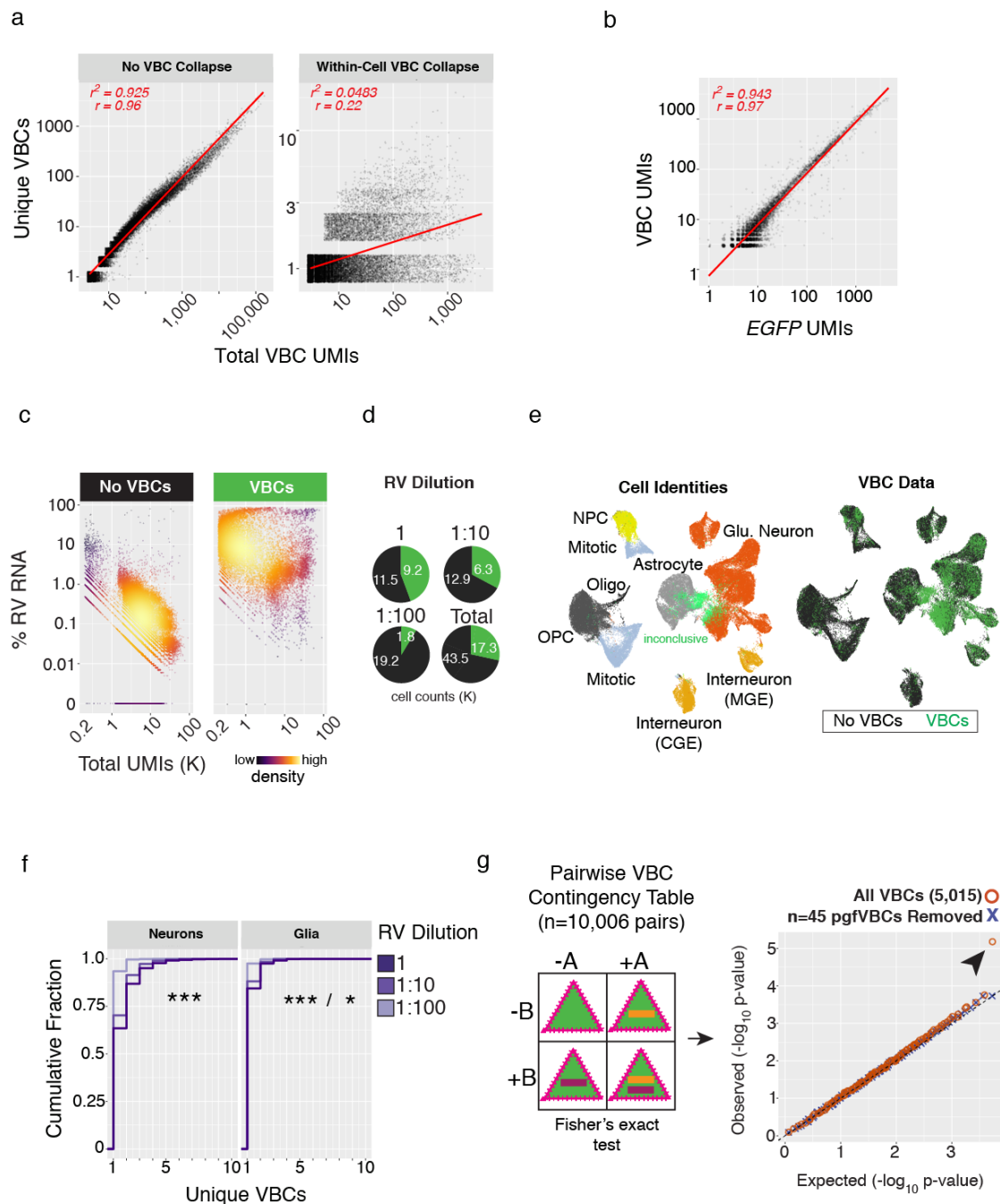
Extended Data Figure 3. Integrating host cell RNA and viral VBC data for thousands of individual starter cells relate how properties of barcoded rabies

763 **libraries behave across founder infections resolved by cell type. a.** An informatic

764 approach ("Within-Cell VBC Collapse", **Methods**) for reconstructing accurate VBC

765 sequences and UMI counts from scRNA data in light of amplification and sequencing

766 artifacts. *Left*, without correction, mutations in barcode sequences incurred during

767 PCR and Illumina sequencing inflate the number of unique VBC sequences observed

768 in each scRNA profile in proportion to the number of total VBC UMIs, leading to a

769 strong correlation (r = 0.96). *Right*, following "Within-Cell VBC Collapse," the

770 relationship between unique VBCs and total VBC UMIs becomes more independent (r

771 = 0.22). **b.** Single cell UMI counts for VBCs (inferred from 3' *EGFP* UTR sequencing)

772 and the *EGFP* mRNAs (inferred from host cell RNA sequencing) are highly correlated (r

773 = 0.97), indicating a strong correspondence across independent sequencing datasets

774 (data are from a single SBARRO experiment, n = 6,979 cells). **c-e.** Single-cell RNA

775 profiles ascertained from brain cells grown *in vitro* expressing *TVA* but not rabies *G*

776 were transduced with EnvA-RV*dG-EGFP*$_{VBC}$ at three different concentrations (no

777 dilution, "1" (MOI ~ 15); diluted one in ten "1:10" (MOI ~ 1.5); or one in a hundred

778 "1:100" (MOI ~ 0.15); **Methods**). **c.** VBC data are comprehensively ascertained from

779 infected ("VBCs"; n = 17,283) but not uninfected ("No VBCs"; n= 43,533) scRNA

780 profiles over a wide range of UMI counts and percentages of total viral RNA. **d.** Pie

781 charts illustrating percentages of scRNA profiles for which VBC data were ascertained

782 (green) or not ascertained (black) across rabies virus dilution conditions. Total cell

783 counts are listed. **e.** UMAP embedding of 60,816 scRNA profiles color-coded by

784 molecular identity (*left*) or VBC ascertainment status (*right*) following LIGER analysis

785 (**Methods**). A subset of infected scRNA profiles (n = 2,635) could not be definitively

786 identified (light green). **f.** Cumulative distribution of unique VBCs per cell, grouped by

787 neuron versus glia type and color-coded by rabies virus dilution. Increasing rabies virus

788 titer leads to more unique VBCs per cell, but does so in a sublinear manner with respect

789 to MOI, suggesting an intrinsic, cell-type-specific limit to the number of independent

790 founder infections (* = p < 0.05; *** = p < 0.001, Kolmogorov–Smirnov Test). **g.** Testing

791 VBC independence in the context of 17.2K starter cell founder infections with multiple

792 VBCs. *Left*, a schematic of the contingency table comparing the number of scRNA

793 profiles in which two VBCs ("A", purple; "B", orange) occur together (+A/+B),

794 independently (+A or +B) , or are not observed (-A/-B). VBC pairs which occurred

795 together more than chance (n=45 of 10,009 total pairs with Bonferroni-corrected p <

796 0.05, Fisher's Exact test) were considered putative genome fusion VBCs and flagged.

797 *Right*, Q-Q plot comparing observed vs expected p-values - expected p-values were

798 generated after VBCs were randomized across scRNA profiles - when all VBCs pairs

799 were considered (orange circles) and after n=45 pgfVBC pairs were removed (purple

800 "x"). Arrowhead highlights the inflation of observed p-values away from the

801 expectation of random driven by putative genome fusion VBCs (**Methods**).

802

803

804

805

806

807

808

809

810

811

812

813

Extended Data Figure 4

815

a

"Starter": EGFP+/TCB+
Presynaptic: EGFP+/TCB-

5 mm

b

EGFP+/
TCB+

EGFP+/   EGFP+/
TCB+   TCB-

DAPI

DAPI

500 μm

500 μm

c

Cell Number

100

10

1

Anatomy   SBARRO

d

R^2= 0.96
Intercept = -565.5
Slope = 0.24
p = 1.2e^{-11}

scRNA profiles (K)

20

10

0

0   25   50   75   100
FACS EGFP count (K)

R^2= 0.94
Intercept = -20.5
Slope = 0.025
p = 3.5e^{-10}

starter cell
scRNA profiles (K)

200

100

0

0   2.5   5.0   7.5   10
FACS EGFP/mCherry
count (K)

816

Extended Data Figure 4. Anatomy of monosynaptic rabies virus spread *in vitro*.

a-b. Monosynaptic cell-to-cell spread events of rabies virus in cell cultures derived

from dissociated embryonic mouse cortex exhibit stereotyped spatial patterning. In

each culture well, a small subset of potential starter cells was endowed using a rAAV

Cre-recombinase based strategy followed by transduction of EnvA-RV*dG-EGFP*$_{VBC}$.

Fluorescent scans of whole culture wells distinguish the locations of these spatially

823   sparse starter (EGFP+/TVA-mCherry+, magenta) and presynaptic cells (EGFP+/TVA-

824   mCherry-, green) (**Methods**). **a.** Locations of starter and presynaptic cells derived from

825   a scan of a representative culture well. Presynaptic cells tend to spatially cluster

826   around starter cells, but are also observed at in a distributed fashion at greater

827   distances from any individual starter. Contours illustrate areas density of rabies virus

828   infected cells. **b.** Higher magnification view showing the locations of a single starter

829   cell (*left*) and presynaptic cells in close proximity (*right*). *Top*, fluorescent images.

830   *Bottom*, plot of extracted cell locations. **c.** A comparison of inferred presynaptic

831   network sizes based on anatomical imaging (the number of clustered presynaptic

832   cells) or SBARRO sequencing (the number of presynaptic scRNA profiles based on

833   uCIPs). Experiments were performed in parallel from neighboring culture wells grown

834   from the same cell suspension. The largest four inferred networks from each modality

835   are shown. **d.** A comparison of FACS-based cell counts and scRNA profiles for all

836   SBARRO cells (*left*) or just starter cells (*right*) fit with linear models for which the slope

837   was used to estimate the sample rate (n=16 culture wells; **Methods**). **d.** A comparison

838   of FACS-based single-cell RNA sampling rates of rabies virus infected EGFP+ cells

839   (n=16 culture wells). A linear model of slope = 0.24 described the relationship (R^2 =

840   0.96 and p = -1.2e-11), suggesting scRNA was sampled from 24% of FACS-enriched

841   cells (**Methods**).

842

843

844

845

846

847

Extended Data Figure 5

a

Molecular Identities of Subclusters
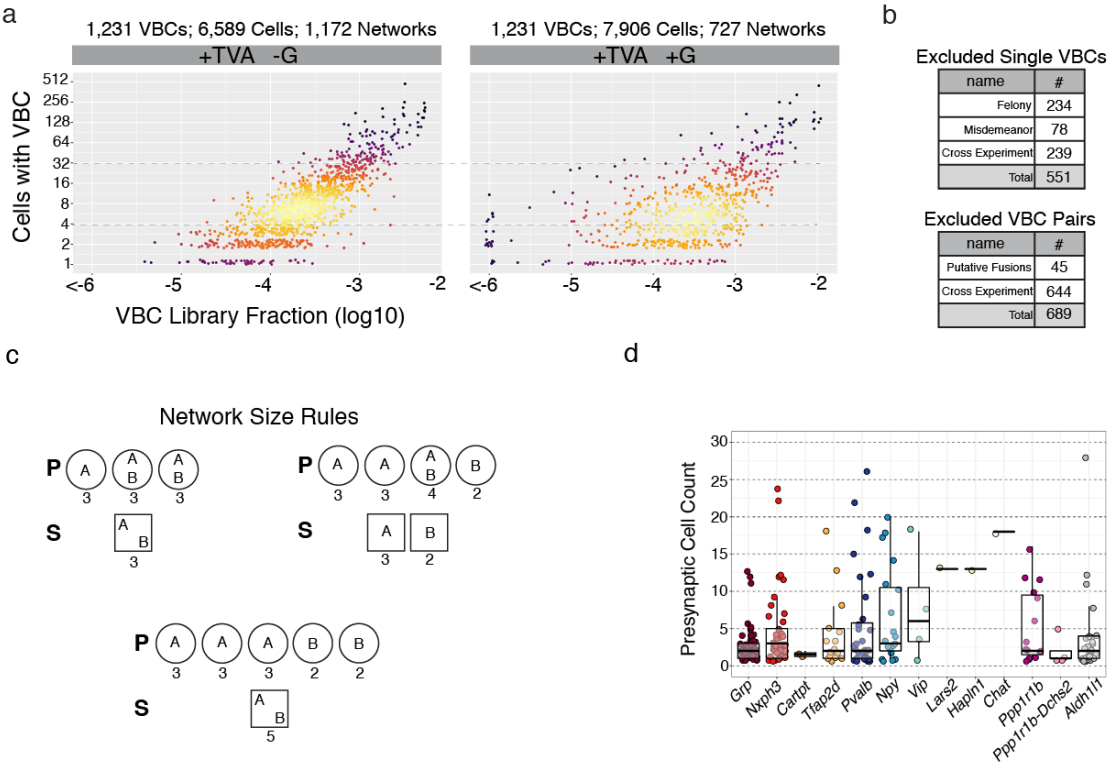

b

Example Subcluster Marker Genes

c

d
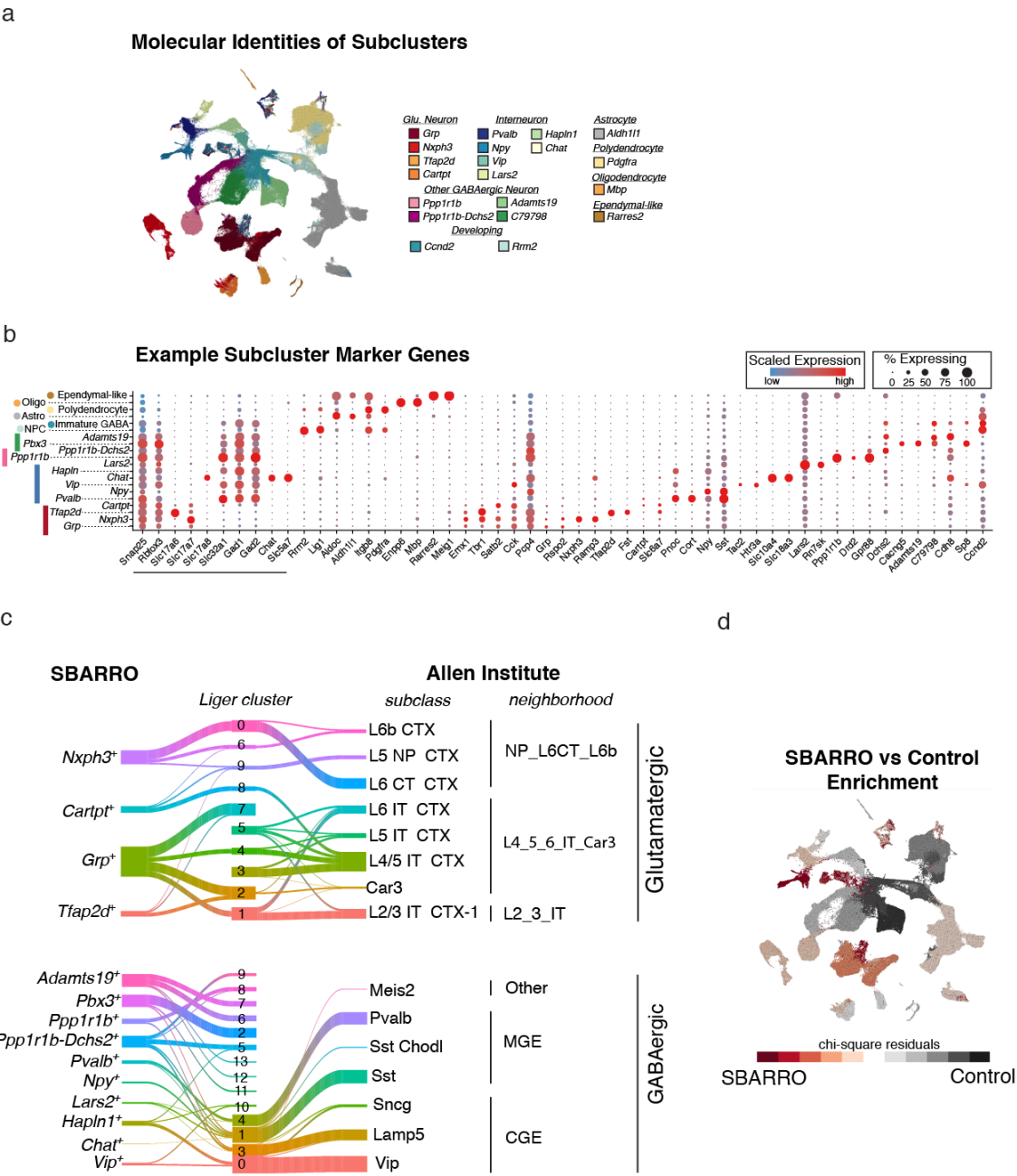
Extended Data Figure 5. Assigning molecular identities to SBARRO scRNA

profiles. a. UMAP embedding color-coded by granular molecular subtypes

(subclusters). **b.** Dotplot of example marker gene expression patterns across

40

854   subcluster populations. Common markers for neurons and neuron types are

855   underlined; additional genes pairs were selected for each population based on

856   differential expression analysis (**Methods**). **c.** Sankey plots showing molecular

857   homologies between scRNA profiles from SBARRO control cells *in vitro* and adult

858   mouse cortex *in vivo*[18] following LIGER analysis of glutamatergic (SBARRO, n = 24,155

859   profiles; Allen Institute, n= 38,899) and GABAergic (SBARRO, n = 54,713 profiles; Allen

860   Institute, n= 18,163) neurons (**Methods**). **d.** Quantifying enrichment or depletion of

861   SBARRO libraries as compared to control scRNA profiles. Color-code shows chi-square

862   residuals after the number of SBARRO/control RNA profiles within each coarse

863   molecular population are compared to dataset totals (**Methods**).

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

Extended Data Figure 6

881

882

Extended Data Figure 6. SBARRO inference of synaptic networks through VBC-based CIPs. a. The effect of rabies virus spread on the relationship between VBC library abundance and number of cells in which each VBC was ascertained. *Right*, a single experiment ("SCC07_1e3_A") in which EnvA-RV*dG-EGFP*_{VBC} founder infections were complemented with glycoprotein endowing monosynaptic retrograde spread (+G). *Left*, a version of the G- starter cell corpus (**Fig. 2f**), consisting exclusively of founder infections, randomly down-sampled to match equivalent VBC numbers (n=1,231). **b.** Tables of VBCs (*top*) and VBC pairs (*bottom*) excluded from network inference using EnvA-RV*dG-EGFP*_{VBC} (**Methods**). **c.** Schematic describing how the "Network Size" parameter was calculated for starter cells (squares) and presynaptic cells (circles) based

893   on examples using two VBCS, "A" and "B". Network Size values are listed below each

894   cell. **d.** Inferred presynaptic network sizes by starter cell subtype for all identified

895   starter cells.

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

Extended Data Figure 7
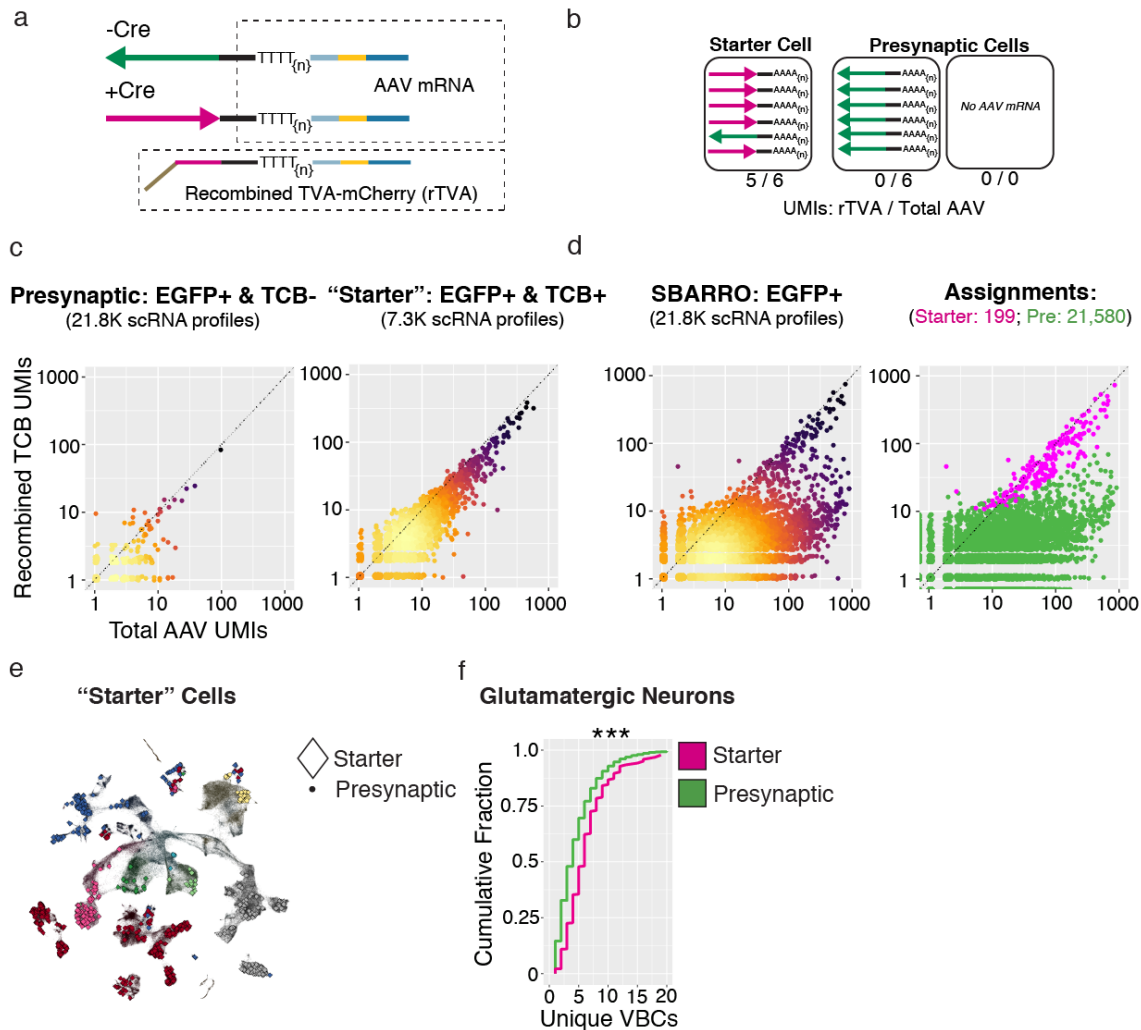
922 **Extended Data Figure 7. Assigning starter cell identities to SBARRO scRNA**

923 **profiles. a-c.** Identifying starter cells from SBARRO scRNA profiles. Starter cells are

924 endowed with functional *TVA-mCherry* and *G* mRNAs after Cre-mediated

925 recombination of CAG-Flex-TVA-mCherry and CAG-Flex-B19(G) rAAV genomes. **a.**

926 Schematic describing the 3' end structure of recombined (+Cre, magenta) or

927 unrecombined (-Cre, green) rAAV mRNAs (such as those encoding *TVA-mCherry* and

928 *G*) after single-cell barcoding and first strand synthesis. The sequenced region critical

929 to determining the identity of the expressed gene – the black bar in the dashed upper

930 box – is unaffected by recombination, thus the vast majority of single cell counts of

931 rAAV mRNAs enabled by standard 3' scRNA-seq (i.e. Drop-seq, inDrop and 10x) are

932 not recombination-informative. To generate sequencing libraries selectively for

933 recombined rAAV mRNAs, we amplified and independently sequenced (using Illumina

934 flowcells) only recombined *TVA-mCherry* transcripts (bottom dashed box; **Methods**).

935 **b**. Cartoon illustrating recombined vs unrecombined rAAV mRNA content of starter

936 and presynaptic cells. Starter cell RNA profiles are enriched for recombined rAAV

937 mRNAs, while presynaptic cells are enriched for unrecombined molecules or have no

938 detectable rAAV mRNA counts. **c**. Validation of rAAV mRNA signatures for starter and

939 presynaptic cells after physical separation via FACS. Scatter plots comparing UMI

940 counts of recombined *TVA-mCherry* mRNA vs Total rAAV mRNA for scRNA profiles

941 resulting from FACS-based separation and independent library generation of

942 presynaptic (*left*, EGFP+/TVA-mCherry-) or starter (*right*, EGFP+/TVA-mCherry+)

943 cells. Points falling along the dotted unity line are cell profiles for which all rAAV

944 mRNAs are from recombined *TVA-mCherry* transcripts. scRNA profiles from sorted

945 starter cells have higher counts of Total rAAV mRNAs and those counts are largely

946 from recombined *TVA-mCherry mRNAs*. Lighter colors indicate higher point densities.

947 **d**. Assigning starter cell identities to SBARRO scRNA profiles (n=21,580) from a

948 single experiment ("SCC07_1e2_C"). *Left*, scatterplot of recombined *TVA-mCherry* vs

949 Total rAAV mRNA counts. *Right*, color-coded by starter (n=199) or presynaptic

950 (n=21,580) assignment based on the results of a binomial testing in which starter RNA

951 profiles exhibit statistical enrichments for recombined *TVA-mCherry* counts (versus

952 Total rAAV counts) and Total rAAV counts (versus all UMIs; **Methods**). **e.** UMAP

953 locations of starter and presynaptic cells. **f.** Cumulative distribution of unique VBCs

954   across glutamatergic neuron starter and presynaptic cells (*** = p < 2.2e-16,

955   Kolmogorov–Smirnov Test).

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977
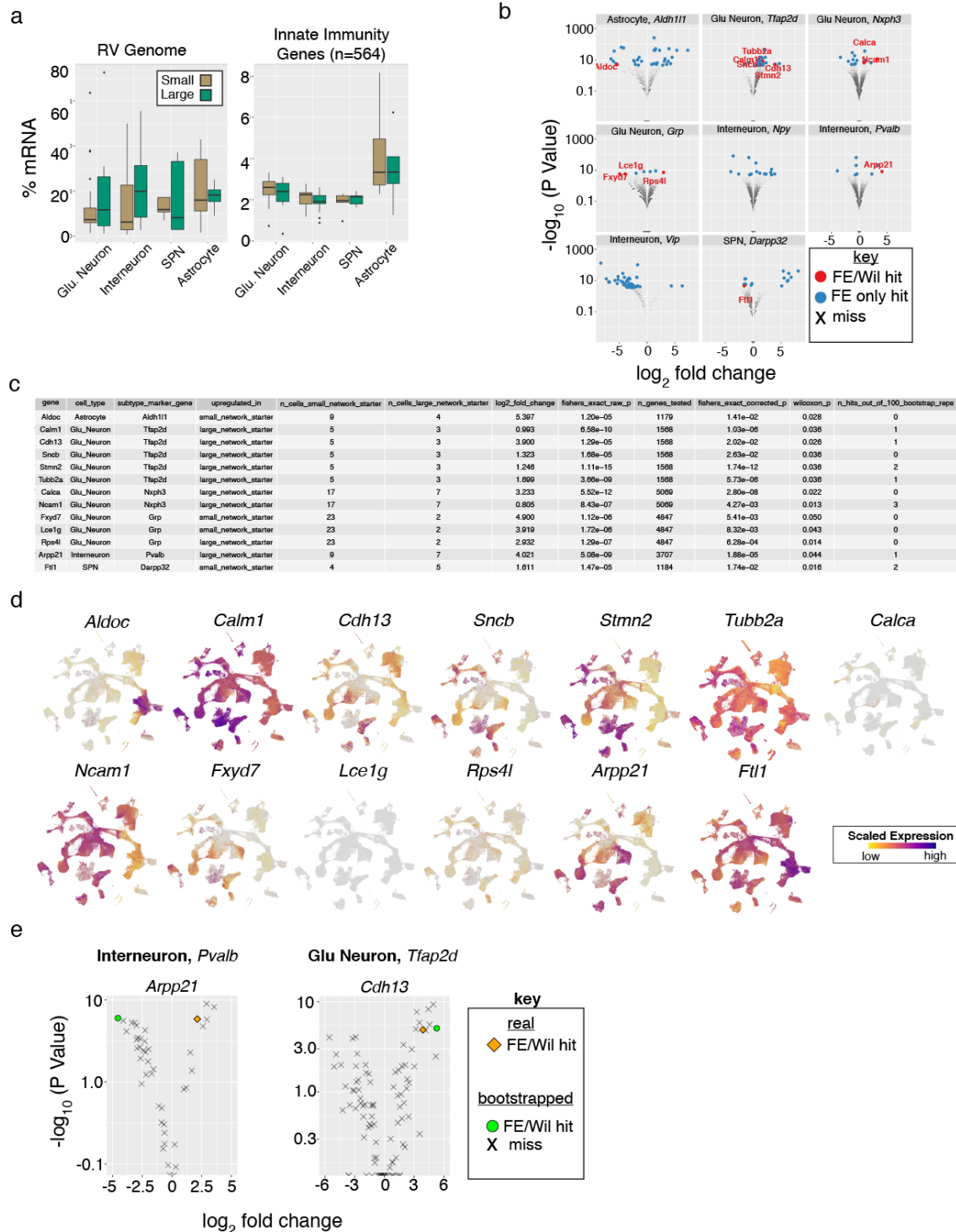
978
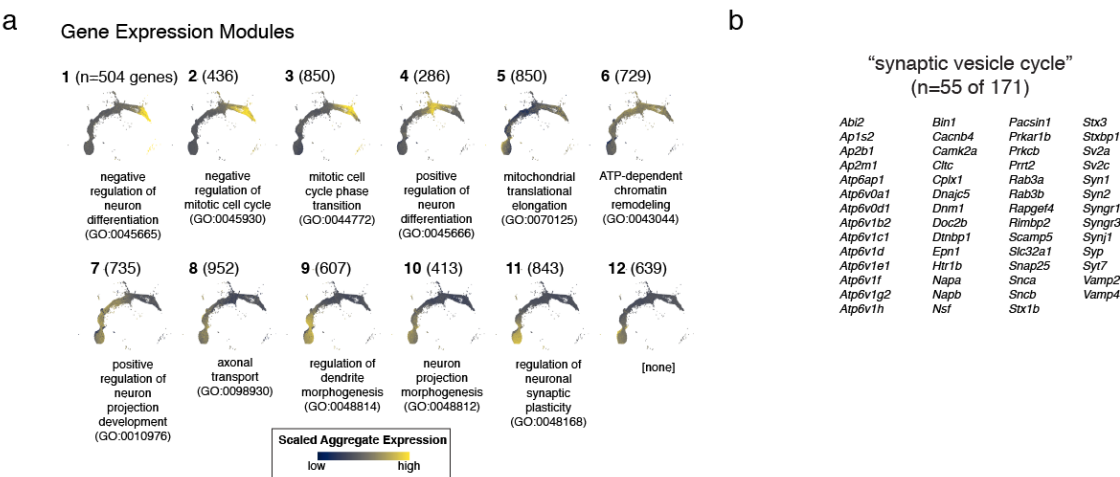
Extended Data Figure 8

a

b

c

d

e

981   **Extended Data Figure 8. Properties of postsynaptic starter cell infection and**

982   **host cell RNAs associated with presynaptic network size inferences. a.** Viral load

983   (% of all mRNAs derived from 5 rabies virus genes; *left*) and innate immunity expression

984   scores (aggregated from 584 curated genes[22]; *right*) across starter cell RNA profiles

985   (n=144) did not show detectable differences across "large" or "small" presynaptic

986   network size groupings for four major brain cell types ($p > 0.05$, Wilcoxon Test). **b.**

987   Volcano plots illustrating results from differential expression testing of starter cell RNA

988   profiles comparing "large" and "small" presynaptic network size categories by starter cell

989   subtypes. UMI counts for each gene of sufficient expression were aggregated by

990   inferred presynaptic network size category then compared (Fisher's Exact Test;

991   **Methods**). Genes passing corrected p value thresholds ($p < 0.05$, blue dots) were

992   further tested for differences in single-cell scaled expression (Wilcoxon Test; **Methods**).

993   Those genes that pass this additional test ($p < 0.05$) were considered hits and labeled

994   (red dots). **c**. Summary table describing differential expression results for those genes

995   identified in b. **d**. Expression plots for the genes identified in c. **e**. Volcano plots

996   comparing differential expression results for *Arpp21* in *Pvalb*+ Interneurons (*left*) and

997   *Cdh13* in *Tfap2d*+ Glutamatergic Neurons (*right*) in real data or 100 permuted replicates

998   in which starter cell RNA profiles were randomly replaced by presynaptic profiles of the

999   same subtype (**Methods**). The real data is shown with a gold diamond; permuted

1000   replicates passing aggregate UMI (Fisher's Exact Test) and scaled expression

1001   comparisons (Wilcoxon Test) are shown as green circles; all other comparisons are

1002   shown with grey crosses.

1003

1004

1005

1006

1007

Extended Data Figure 9

a

Gene Expression Modules

1 (n=504 genes) **2** (436) **3** (850) **4** (286) **5** (850) **6** (729)

negative regulation of neuron differentiation (GO:0045665) | negative regulation of mitotic cell cycle (GO:0045930) | mitotic cell cycle phase transition (GO:0044772) | positive regulation of neuron differentiation (GO:0045666) | mitochondrial translational elongation (GO:0070125) | ATP-dependent chromatin remodeling (GO:0043044)

7 (735) **8** (952) **9** (607) **10** (413) **11** (843) **12** (639)

positive regulation of neuron projection development (GO:0010976) | axonal transport (GO:0098930) | regulation of dendrite morphogenesis (GO:0048814) | neuron projection morphogenesis (GO:0048812) | regulation of neuronal synaptic plasticity (GO:0048168) | [none]

Scaled Aggregate Expression
low   high

b

"synaptic vesicle cycle"
(n=55 of 171)

Abi2	Bin1	Pacsin1	Stx3
Ap1s2	Cacnb4	Prkar1b	Stxbp1
Ap2b1	Camk2a	Prkcb	Sv2a
Ap2m1	Cltc	Prrt2	Sv2c
Atp6ap1	Cplx1	Rab3a	Syn1
Atp6v0a1	Dnajc5	Rab3b	Syn2
Atp6v0d1	Dnm1	Rapgef4	Syngr1
Atp6v1b2	Doc2b	Rimbp2	Syngr3
Atp6v1c1	Dtnbp1	Scamp5	Synj1
Atp6v1d	Epn1	Slc32a1	Syp
Atp6v1e1	Htr1b	Snap25	Syt7
Atp6v1f	Napa	Snca	Vamp2
Atp6v1g2	Napb	Sncb	Vamp4
Atp6v1h	Nsf	Stx1b	

Extended Data Figure 9. Molecular correlates of rabies virus transmission

during SPN development. a. Modules of gene expression (n=12) identified with

Monocle3 (**Methods**). For each numbered module, the count of associated genes is

shown parenthetically, module expression is color-coded by aggregate expression

and a representative enrichment for biological process gene ontologies categories is

shown (adjusted $p < 0.05$). **b.** Names of rabies virus-transmission correlated "synaptic

vesicle cycle" genes (n=55 of 171 in SynGO category GO:0099504).

METHODS

Barcoding rabies virus plasmids and RNA genomes. Rabies virus rescue encapsidates RNA genomes from DNA templates[51]. 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. 1f** and **Extended Data Fig. 2a**); 2) more uniformly amplify plasmid DNA through optimized bacterial transformation and plate-based growth conditions (**Extended Data 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[10] and then exacerbated by biases in viral replication. Our 7-9 day protocol is three-fold faster and achieves titers equivalent or higher than published protocols ($1\times10^{8-9}$ IU/mL; **Extended Data Fig. 2d**)[15,52]. 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 (**Extended Data Fig. 1c** and **Supplementary Table 1**); the resulting paired UMI-barcode sequences were used to count individual

1042   molecules. Inflation of barcode sequences and UMI counts due to mutations arising

1043   during library amplification and Illumina sequencing were accounted for (see

1044   description in Results) using a custom algorithm for post-hoc mutation correction. See

1045   "*Quantifying barcodes from plasmids and rabies virus genome*" section below for

1046   details.

1047

1048   *PCR-based plasmid barcoding*. To generate plasmid libraries in which individual

1049   circular plasmids encode unique barcode sequences, we developed a PCR-based

1050   molecular workflow in which a bipartite barcode cassette can be targeted to arbitrary

1051   regions of a non-barcoded plasmid template (**Extended Data Fig. 1b**). We applied

1052   our system to the SAD-B19 genome plasmid in which the G gene has been replaced

1053   by *EGFP* (cSPBN-4GFP, Addgene #52487[15]), targeting the barcode cassette to the 3'

1054   UTR of *EGFP* adjacent to the viral polyadenylation sequence[53]. To introduce each half

1055   of the barcode cassette, whole-plasmid PCR was performed with forward and reverse

1056   primers targeting the desired region. Each primer contains 3' plasmid-complementary

1057   sequence followed by a 5' tail with 10 bps of random nucleotides further flanked by a

1058   restriction cassette which includes the PlutI ("GGCGCC") restriction site (pSPBN-GFP

1059   Barcoding, Forward Primer: B19_barcode_F, Reverse Primer: B19_barcode_R;

1060   **Supplementary Table 1**). During PCR (See *"Barcoding PCR"* protocol), each round

1061   of primer hybridization and extension introduces a unique barcode, resulting in a

1062   linear, double-stranded amplicon collection in which unique 10 bp barcodes have

1063   been introduced into the 5' terminus of each DNA strand. The desired ~14.5 kb

1064   amplicons were size-selected using standard low-gel agarose (Sigma-Aldrich, A9414)

1065   electrophoresis and cleaned (Zymo Research, Gel DNA Recovery Kit #D4001), then

1066   re-cleaned and concentrated to >200 ng/μl (Zymo Research, DNA Clean &

1067   Concentrator-25 #D4033). To efficiently circularize the amplicons using the barcode

1068   restriction cassette and to remove remaining template plasmid and unwanted linear

1069   products, we developed a series of enzymatic reactions that consecutively performed

1070   in the same tube, saving time and avoiding DNA damage and loss due to repeated

1071   purification (See *"Plasmid Circularization Protocol"*). Briefly, DpnI digest removes

1072   remaining methylated plasmid DNA; PlutI restriction and T4 ligation circularize the

1073   amplicons thus covalently bonding each of two 10 bp barcodes into a 36 bp barcode

1074   cassette; and RecBCD selectively degrades linear DNA over circularized plasmid

1075   containing non-complementary barcode sequences, typically enriching the percentage

1076   of circularized product ~3.5 fold (from ~$20\pm0.7\%$ to $71\pm4\%$, n=6 experiments, $\pm$

1077   denotes s.e.m).

1078

1079   <u>Barcoding PCR</u>

1080   1. PCR (25 μl reaction)

Reagent	Volume (μl)
Q5 High-Fidelity 2X Master Mix (NEB, #M0515)	12.5
Forward primer (B19_barcode_F)	1.25
Reverse primer (B19_barcode_R)	1.25
Template plasmid (cSPBN-4GFP, 0.4 ng/μl)	5
Water	5

1081   <u>Cycle Conditions</u>

1082   <u>1.</u> 98°C for 30s; <u>2.</u> 98°C for 10s; <u>3.</u> 72°C for 30s; <u>4.</u> 72°C for 480s; <u>5.</u> Go to Step 2-4,

1083   35 times; <u>6.</u> 72°C for 60s

1084

1085   Plasmid Circularization Protocol

1086   1. Digest (50 μl reaction)

Reagent	Volume (μl)
PlutI (10,000 U/mL, NEB R0713S)	0.5
DpnI (20,000 U/mL, NEB R0176S)	0.25
CutSmart Buffer (10x, NEB B7204S)	5
Linear Barcoded DNA	500 ng
Water	To 50 μl

1087   Cycle Conditions

1088   1. 37°C for 1 hr; 2. 80°C for 20 min (heat inactivation)

1089

1090   2. Ligation (spike-in, +5.1 μl)

Reagent	Volume (μl)
T4 DNA Ligase (10,000 U/ml, NEB M0202T)	0.1 (200 U)
ATP (10 mM, NEB P0756L)	5

1091   Cycle Conditions

1092   1. 4°C for 2 hrs; 2. 65°C for 20 min (heat inactivation)

1093

1094   3. Circular plasmid enrichment (spike-in, +8.8 μl)

Reagent	Volume (μl)
Exonuclease V (RecBCD, 10,000 U/ml, NEB M0345L)	1
NEBuffer 4 (10x, NEB B7004S)	1.4
ATP (10 mM, NEB P0756L)	6.4

1095   Cycle Conditions

1096   1. 37°C for 1 hr; 2. 70°C for 30 min (heat inactivation)

1097

1098

1099   *Amplifying barcoded DNA plasmids.* Rabies virus rescue requires tens of micrograms

1100   of supercoiled rabies virus genome plasmid for cell transfection. To amplify and

1101   supercoil DNA plasmid libraries carrying hyper-diverse barcodes in a manner that

1102   minimizes loss and skew of barcoded plasmid representation, we transformed eight

1103   vials of chemically competent One Shot OmniMAX 2 T1[R] cells (ThermoFisher

1104   Scientific, C854003) in parallel each with 200 ng of DNA from the *Plasmid*

1105   *Circularization Protocol*. After 1 hour of recovery growth, cultures were combined and

1106   2 ml of the cell mixture was spread over n=8 large plates (24.5 x 24.5cm; Corning,

1107   CLS431111[54]) containing LB Agar (Sigma-Aldrich, L2897) and Ampicillin (100ug/mL;

1108   Sigma-Aldrich, A5354) and grown over night at 37°C. Colonies were scraped from

1109   each plate with 15 ml LB and pelleted through centrifugation (6000g for 15 min at

1110   4°C). Plasmids were isolated from cell pellets using the EndoFree Plasmid Maxi Kit

1111   (0.45 g cells/column; Qiagen, 12362). Sequencing-based barcode quantification (see

1112   below *Quantifying barcodes from plasmids and rabies virus genomes*) comparing

1113   plasmids prepared from pooled transformants grown on plates (as described) versus

1114   liquid culture (250 ml) demonstrated that plate-based growth dramatically reduced

1115   overrepresentation of plasmid barcodes (**Extended Data Fig. 2c**), presumably by

1116   homogenizing clonal growth rates.

1117

1118   *Rescuing barcoded rabies virus libraries. De novo* rescue of negative-stranded RNA

1119   viruses requires transfection-based encapsidation of positive-stranded RNA genomes

1120   with N, P and L proteins; the minimal replication-competent nucleocapsid [51]; Genomes

1121   lacking the G gene additionally require G protein such that replicating particles can

1122   spread cell-to-cell [15]. Two properties of rescue create challenges for generating

1123   particle libraries with millions of unique and uniformly abundant genomes. First, cells

1124   in which rescue events occur are rare (<1:10,000 transfected cells [10]), creating a

1125   limited number of cellular environments in which encapsidation can occur. Second,

1126   state-of-the-art packaging protocols serially infect fresh cultured cells to increase viral

1127   titer; increasing, with each passage, the opportunities for individual clones gain a

1128   replication advantage. To develop a rabies virus rescue protocol for barcoded

1129   genomes, we first systematically characterized how barcode abundances behaved

1130   after transfection and each passage stage of a widely used protocol[15]. We observed

1131   that 1) minimally and on average, hundreds of unique rabies virus genomes were

1132   encapsidated per encapsidation-competent cell and that 2) viral passages tended to

1133   reduce the number of unique barcoded genomes and distort their relative abundances

1134   (**Extended Data Fig. 3**). Therefore, we increased the total number of encapsidation-

1135   competent cells by optimizing large-scale rabies virus transfection and created a one-

1136   step rescue protocol capable of generating rabies virus libraries with millions of

1137   unique genomes at similarly high titers (~2.5 x 10^9 IU/ml) but 4-fold faster (5 days)

1138   than published protocols [15,52]. Pseudotyping with non-native coat proteins requires an

1139   additional 6 days. Specifically, poly-L-lysine (Sigma-Aldrich, P4707) coated T-225

1140   flasks containing 85-95% confluent HEK-293T/17 cells (ATCC, CRL-11268) were each

1141   transfected (Xfect, Takara #631318) with a DNA cocktail containing the 1) the

1142   barcoded rabies virus plasmid library (131.36 μg) and CAG-promoter driven plasmids

1143   for T7 polymerase (23.66 mg, Addgene 59926) and SAD-B19 helper proteins (N,

1144   52.11 μg, Addgene 59924; P, 30.15 μg, Addgene 59925; L, 23.70 μg, Addgene

1145   59922; G, 20.26 μg, Addgene 59921). Cells were maintained with DMEM with

1146   GlutaMAX supplement, pyruvate, high glucose media (Thermo Fisher Scientific,

1147   10569010) supplemented with 5% fetal bovine serum (Thermo Fisher Scientific,

1148   10082147) and 1x antibiotic-antimycotic (Thermo Fisher Scientific, 15240062) and

1149   incubated at 35°C with 5% CO_2. Five days post-transfection, culture media was collected

1150   for either 1) unpseudotyped rabies virus recovery or 2) EnvA pseudotyping. For EnvA

1151   pseudotyping, BHK-EnvA cells (Columbia Univ. Zuckerman Virus Core), initially grown in

1152   15 cm dishes (Corning, 08-772-24) to 85-95% confluence, were infected with filtered

1153   media (0.22 μm PES; Corning, 431097) from the transfected T-225 plates; cells are then

1154   rinsed, pelleted and re-plated first in 15 cm plates then again in T-225 flasks.

1155   Specifically, following 6 hours of incubation with particle-containing media, cells from

1156   each plate are rinsed with two rounds of cold DPBS (+Ca, +Mg), trypsinized with 5 mL of

1157   trypsin-EDTA (Thermo Fisher Scientific, 25300-054) for 30 seconds at 35°C, pelleted

1158   with centrifugation (300 g, 4 min) in DMEM + 10% FBS and then re-plated in 15 cm

1159   plates and allowed to incubate overnight (~16-24 hours) before being re-plated in T-225

1160   flasks with DMEM + 5% FBS. T-225 plate media is supplemented with 3-5 mL of DMEM

1161   + 5% FBS each day for 4 days before being collected and concentrated. Specifically,

1162   collected media is incubated with benzonase nuclease (1:1000 dilution; Millipore Sigma,

1163   70664) for 30 minutes at 37°C and filtered (0.22 μm PES). For ultracentrifugation

1164   (Beckman Coulter, SW32Ti rotor), 2 mL of 20% (w/v) sucrose in DPBS (-Ca, -Mg) is

1165   prepared in ultracentrifuge tubes (Beckman Coulter, 344058) to which the divided EnvA-

1166   pseudotyped viral media is added before pelleting (20,000 RPM for 2 hours at 4°C).

1167   Residual media is removed and viral pellets are each resuspended in 15 μL of DPBS (-

1168   Ca, -Mg) on ice before orbital shaking at 4°C for 8 hours. Volumes are then combined,

1169   aliquoted, and stored at -80°C. Titers were established by quantifying infected HEK-TVA

1170   and HEK-293T/17 cells in 12-well plates (80% confluence) using serial dilutions. To

1171   ensure EnvA-pseudotyping was complete, < 2 HEK-293T/17 cells per well were

1172   tolerated following infection with 1 μL with full-strength sample.

1173

1174   *Quantifying barcodes from plasmids and rabies virus genomes.* UMI-based counting

1175   of barcodes from DNA plasmids and RNA genomes was accomplished with similar

1176   molecular (See "*UMI-based counting of genome and plasmid barcodes*" and

1177   **Extended Data Fig. 2**) and informatic workflows (See "*UMI-based counting of*

1178   *genome and plasmid barcodes*"). RNA genomes were extracted using the ZR Viral

1179   RNA kit (Zymo Research, R1041) from particles ascertained from 1) end stage high-

1180   titer viral aliquots or 2) from cell culture media used for rabies virus rescue after PEG-

1181   based precipitation (Abcam, ab102538) and quantified using the High Sensitivity RNA

1182   ScreenTape assay (Agilent, 5067-5579). To count barcode abundances of individual

1183   RNA genomes or DNA plasmids, an oligonucleotide (B19_UMI_F) containing a SMRT

1184   PCR handle, 12 bp UMI and 33 bps of barcode-adjacent homologous sequence were

1185   hybridized then polymerase-extended through the barcode region. Remaining RNA

1186   genomes were selectively digested using RNase H (New England Biolabs, M0297S)

1187   and reactions were cleaned with Agencourt AMPure XP beads (1:1 volume; Beckman

1188   Coulter, A63881) retaining first-strand cDNA. The UMI-tagged genomic cDNA or

1189   plasmid DNA strands were then selectively amplified (14-18 PCR cycles; 16 median)

1190   using primers which introduce the Illumina P5 (P5-TSO_Hybrid) and indexed P7

1191   (P7i1-L5UTR_seq) sequences. Amplicon libraries were sequenced on an Illumina

1192   MiSeq or NextSeq550 using a custom primer (Read1CustomSeqB) to seed 110 Read

1193   1 cycles. Base pairs (bp) 1-12 were assigned as the UMI. The two 10 bp viral

1194   barcodes were informatically extracted from the barcode cassette using a custom

1195   algorithms based on local sequence alignment algorithm and

1196   ("*TagReadWithRabiesBarcodes*" & "*FilterValidRabiesBarcodes*"). To account for

1197   artifactual barcode sequences created by mutations acquired during the library

1198   amplification and sequencing, we developed an algorithm to identify and collapse

1199   "families" of barcodes with similar sequences likely related through acquired

1200   mutations *("CollapseTagWithContext, MUTATIONAL_COLLAPSE=true"*). Specifically,

1201   after ordering barcodes most to least abundant, we considered each barcode as a

1202   "parent" and identified "siblings" sequences within Hamming distance of 1 of the

1203   "parent" barcode. The process was then iterated for each new "sibling" until no new

1204   "siblings" were discovered. The entire barcode family was assigned the sequence of

1205   the "parent." UMI-parent barcode sequence pairs were then used to count each

1206   "parent" barcode in the library (after collapsing UMI-barcode sequences in which the

1207   UMIs associated with the same "parent" were Hamming distance <= 1). This approach

1208   drastically reduced the inflation of barcode sequences and counts due to library

1209   preparation and sequencing (**Extended Data Fig. 2b**).

1210

1211   UMI-based counting of RNA genome and DNA plasmid barcodes

1212   1a. RNA Genomes - UMI Hybridization (24 μl reaction)

Reagent	Volume (μl)
Rabies ssRNA genomes or dsDNA plasmid (5-25 ng)	--
UMI oligo (10 μM; B19_UMI_F)	4
dNTPs (10 mM;)	4

Water	To 24 µl

1213   <u>Cycle Conditions</u>

1214   <u>1.</u> 72°C for 4 min; <u>2.</u> 4°C

1215

1216   1b. RNA Genomes - Reverse Transcription (spike-in, +16 ml)

Reagent	Volume (µl)
20% Ficoll PM 400 (Sigma Aldrich, 26873-85-8)	4
NxGen RNAse Inhibitor (Lucigen, 30281-2)	0.5
Maxima H- RT (Thermo Scientific, EP0751)	2
5x Maximal H- RT Buffer	8
Water	1.5

1217   <u>Cycle Conditions</u>

1218   <u>1.</u> 42°C for 90 min; <u>2.</u> 85°C 5 min (inactivation)

1219

1220   1c. RNA Genomes – RNase H Treatment (spike-in, +2 ml)

Reagent	Volume (µl)
RNase H (5,000 U/mL, NEB M0297S)	4

1221

1222   1. DNA Plasmids - UMI Hybridization & polymerization (50 µl reaction)

Reagent	Volume (µl)
dsDNA plasmid (5-25 ng)	--
2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602)	25
UMI oligo (10 µM; B19_UMI_F)	2

Water	To 50 μl

1223 <u>Cycle Conditions</u>

1224 <u>1.</u> 98°C for 3 min; <u>2.</u> 68°C 30 s; <u>3.</u> 72°C 20 s

1225

1226 2. Illumina Adaptor PCR (50 μl reaction)

Reagent	Volume (μl)
UMI-tagged DNA	5
P5 Primer (10 μM; P5-TSO_Hybrid)	1
P7 Primer (10 μM; P7i1-L5UTR_seq)	1
2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602)	25
Water	To 50 μl

1227 <u>Cycle Conditions</u>

1228 <u>1.</u> 98°C for 30s; <u>2.</u> 98°C for 10s; <u>3.</u> 72°C for 30s; <u>4.</u> 72°C for 60s; <u>5.</u> Go to Step 2-4,

1229 13-17 times; <u>6.</u> 72°C for 120s

1230

1231 **Synaptic cell culture.** Cells were dissociated from the cortex or striatum of

1232 embryonic day 16 (E16) C57Blk6/N mouse brains and maintained for 14 days *in vitro*

1233 (DIV). rAAVs were transduced on DIV 5 to functionalize "starter" cells. On DIV 12,

1234 EnvA-RV*dG-EGFP*$_{VBC}$ libraries were transduced and infection was allowed to proceed

1235 for 72-96 hours before scRNA-seq libraries were generated. Pregnant C57Blk6/N

1236 dams (Charles River Laboratories) were heavily anesthetized by isoflurane inhalation,

1237 decapitated and the brains of embryonic pups (litter size = 4-9) removed in ice-cold

1238 1X Dissociation Media ("DM"; containing (in mM): 10.52 $MgCl_2$ (Sigma-Aldrich, M2393);

1239   10.53 HEPES (Sigma-Aldrich, H3375); 1.32 Kynurenic Acid (Sigma-Aldrich, K3375) in

1240   HBSS (Thermo Fisher Scientific, 14175079)) in which cortex or cortex and striatum

1241   were dissected from each brain, pooled and incubated in sterile-filtered (0.22 μm;

1242   Corning, 431097) DM+Papain/L-Cysteine (3.4 units Papain and 0.172 mM L-Cysteine;

1243   Worthington Biochemical, LK003178) for 3-5 min at 37° C. Brain tissue is then washed

1244   twice with 2-3 mL sterile-filtered DM+Trypsin Inhibitor (1 mg/mL; Sigma Aldrich,

1245   T9253) and incubated in the 3rd wash for 3-5 min at 37° C. DM+Trypsin Inhibitor is

1246   replaced with 6 mL of sterile-filtered "Plating Media" ("PM"; containing: DMEM (ATCC,

1247   30-2002) and 10% FBS (ATCC, 30-2020)) in which digested brain volumes are

1248   titrated into single cells with a pipetteman equipped with a 5 mL pipet tip. Cell

1249   concentration was measured by diluting cells 1 to 5 in PM, mixed with an equal

1250   volume of 0.4% Trypan Blue (Thermo Fisher Scientific, 15250-061), and quantified

1251   using the Countess II Automated Cell Counter (Life Technologies). Each well of 6-well

1252   cell culture treated plates coated with 0.1% Poly-L-ornithine (3 ug/mL; Sigma Aldrich,

1253   P4957) were seeded with ~750K cells and maintained with sterile-filtered neurobasal

1254   medium (Thermo Fisher Scientific, 21103049), supplemented with serum-free B-27

1255   (Thermo Fisher Scientific,17504044), GlutaMAX (Invitrogen, 35050061) and

1256   Penicillin:Streptomycin (VWR, 45000-652). For imaging experiments, cells were

1257   seeded on glass coverslips (Fisher Scientific, 12-546) coated with 0.1% Poly-L-

1258   ornithine and Laminin (5 ug/mL; Thermo Fisher Scientific, 23017015). On DIV 5, a

1259   cocktail of three rAAVs was used to functionalize starter cells. Our starter cell strategy

1260   was designed to deliver consistent, high MOIs of rAAV per starter cell while flexibly

1261   controlling the number of starter cells in each culture through Cre delivery.

1262   Specifically, 1 μl CAG-Flex-TVA-mCherry ("TCB"; serotype, 2-9; titer, 2.2x10^{13}

1263   genomes/mL; MOI, ~$2.9e^4$; UNC Vector Core) and 1 μl CAG-Flex-B19G (serotype, 2-

1264   9; titer, 1.6×10^{13} genomes/mL; MOI, ~$2.1e^4$; UNC Vector Core) were added to each

1265   well along with 1 μl of Syn1-EBFP-Cre (serotype, 2-1; titer, 6×10^{12} genomes/mL; MOI,

1266   ~8-0.08; Addgene, 51507-AAV1) delivered at full strength or diluted $1:10^3$-10^4. At DIV

1267   12, 1 μl of EnvA-RV*dG-EGFP*VBC library (titer, 0.19-1.1×10^{10} IU/mL; Total cell MOI, 2.5-

1268   14.8) was added to each well. Epifluorescence imaging was used to monitor the

1269   progress of infections, including starter cell locations and morphology (based on TVA-

1270   mCherry fluorescence) as wells as rabies virus transduction and spread from starter

1271   cells (based EGFP fluorescence). , EnvA-RV*dG-EGFP*VBC transduction was completely

1272   dependent on the TVA receptor, since no EGFP fluorescence was observed in

1273   equivalent experiments in which the CAG-Flex-TVA-mCherry rAAV was excluded. To

1274   prepare cultures grown on coverslips for imaging or *in situ* hybridization experiments,

1275   neurobasal medium was removed and each culture well was rinsed three times with

1276   1X PBS and then fixed with fresh 4% paraformaldehyde at room temperature for 30

1277   min, followed by three rinses with 1x PBS. For fluorescence imagining, coverslips

1278   were slide-mounted and nuclei counterstained using ProLong Gold Antifade (Thermo

1279   Fisher Scientific, P36934). For *in situ* hybridizations, cover slips were dehydrated

1280   through a series of brief (~1 min) ethanol washes (50%, 70% and 100% EtOH) before

1281   being stored at -20°C in 100% EtOH.

1282   **Sequencing single-cell mRNAs: host cell, rabies virus barcodes and**

1283   **recombined rAAV.** scRNA-seq libraries were generated using the Chromium Single

1284   Cell 3' v2 or v3 Chemistry platform (10x Genomics), prepared following kit guidelines,

1285   and sequenced to a depth of ~45K reads per cell (Illumina NovaSeq 6000).

1286   Sequences were aligned using STAR v2.4.0a[55] against a composite genome

1287   consisting of GRCm38.81, barcoded cSPBN-4GFP and rAAV accessory sequences

1288   (including the 3' UTR and TVA-mCherry, rabies G and Cre coding sequences) using a

1289   workflow similar to that described for Drop-seq[56]. To create input cell suspensions, a

1290   protocol developed for the adult mouse brain[30] was adapted for *in vitro* synaptic

1291   cultures. Culture wells were first incubated for ~20 min at 37°C with 1.8 mL of

1292   Dissociation Media (DM) containing Papain and Protease 23 [30] until detachment of the

1293   cell monolayer. Cultures were gently swirled and incubated for an additional 5 min.

1294   Each well was then supplemented with 1 mL of DM before transfer into a 5 mL

1295   eppendorf tube in which cells were pelleted through centrifugation (300g for 5 min).

1296   The supernatant was removed and replaced with 1 mL of ice-cold DM in which cells

1297   were titrated by successively smaller bore polished glass Pasteur pipets. The cells

1298   were then re-pelleted and resuspended in 0.5 mL of DM before being filtered through

1299   a pre-wet 40 μm cell strainer (Corning, 352340). For SBARRO experiments, rabies

1300   infected cells were enriched from total cell suspensions through fluorescent activated

1301   cell sorting (FACS) using the MoFlo Astrios EQ cell sorter (Beckman Coulter; 70 μm

1302   nozzle) into 25 μl of DM. RV-derived EGFP fluorescence was used to gate for

1303   SBARRO cells. For experiments in which distinct scRNA-seq libraries were created

1304   for starter and presynaptic cells, mCherry fluorescence (driven by cre-recombined

1305   rAAV genomes encoding TVA-mCherry) was used as an additional gate to sort starter

1306   (GFP+/ TVA-mCherry+) or presynaptic (EGFP+/ TVA-mCherry-) populations. Post

1307   hoc FACS analysis was performed with FloJo software (BD Biosciences). For scRNA-

1308   seq libraries downstream of FACS, 1.7K-33K cells (based on FACS counts) were

1309   loaded per single-cell RNA capture reaction. To generate scRNA-seq libraries for

1310   which total cell suspensions were used as input, cell concentrations were quantified

1311 using the Countess II Automated Cell Counter (Life Technologies) and 10K-16K cells

1312 were loaded per single-cell RNA capture reaction. Rabies virus barcoded EGFP and

1313 cre-recombined TVA-mCherry rAAV 3' mRNAs were independently amplified (See

1314 "*Selective mRNA Adaptor PCR*" protocol; Rabies EGFP: ~255 bp amplicon, 10-14

1315 cycles; rAAV TVA-mCherry: ~1,125 bp amplicon, 34 cycles) from single-cell cDNA

1316 using primers that introduced the Illumina P5 site (P5-TSO_Hybrid) and indexed P7

1317 site onto barcoded 3' EGFP (BC_Seq_P7i`x`_GFP_v4c) or recombined TVA-mCherry

1318 (P7`x`_TCB_CreOn_v4). Barcoded EGFP and recombined TVA-mCherry libraries

1319 were multiplexed and sequenced separately on an Illumina NextSeq500 using a High

1320 Output 150 cycle Kit (Stock Read1 primer; Library concentrations: EGFP, 1.8 pM with

1321 20% PhiX; TVA-mCherry, 0.4 pM with 50% PhiX; Cycle distributions: Read1=28,

1322 Read2=98, Index=8; Reads per library: EGFP, 43M-90M; TVA-mCherry, 121K-8.6M).

1323 To generate integer counts of recombined TVA-mCherry transcripts per cell,

1324 sequences generated from recombined TVA-mCherry library were aligned using

1325 STAR v2.4.0a [55] against a composite genome consisting of GRCm38.81, the RV*dG-*

1326 *EGFP*$_{VBC}$ genome and rAAV accessory sequences (including the 3' UTR and TVA-

1327 mCherry, rabies G and Cre coding sequences). The sequences of UMIs associated

1328 with each gene and cell barcode were collapsed within an edit distance of 2. To

1329 quantify the number of TVA-mCherry mRNAs derived from cre-recombined rAAV

1330 genomes per cell, UMI counts mapping to the TVA-mCherry coding sequence or the

1331 3' UTR were summed. To discover and quantify the RV-derived VBCs in the 3' UTR of

1332 EGFP mRNA, raw VBC sequences were informatically extracted from each read (as

1333 described above for plasmids and viral genome sequences). To accurately

1334 reconstruct and count VBCs in each single-cell, we leveraged the single-cell nature of

1335 the data to informatically account for two types of artifacts: 1) the inflation of barcode

1336 sequences generated by mutations during library amplification and sequencing and 2)

1337 swapping of non-adjacent VBC and cell barcode/UMI sequences due to strand

1338 displacement during PCR amplification ("CollapseTagWithContext,

1339 ADAPTIVE_EDIT_DISTANCE=true" & "BipartiteRabiesVirusCollapse"). To account for

1340 mutations, we assumed that in individual cells, closely related barcode sequences

1341 were likely to originate from mutations introduced during library preparation or

1342 sequencing rather than independent infections of rabies virus particles with similar 20

1343 bp genomic barcodes. Thus we evaluated Hamming edit distance relationships across

1344 all sufficiently abundant VBCs (Inclusion Threshold: >= 3 ("No RG" experiments) or 5

1345 ("SCC" Experiments) UMIs) found within each cell. From these edit distance

1346 distributions, many low-abundance "sibling" VBCs with sequences similar to a single,

1347 more numerous "parent" VBC were assigned the VBC of the "parent"; collapsing these

1348 mutationally-related VBC "families" corrected the strong artifactual correlation present

1349 in the raw data in which cells with more VBC UMIs also tended to have more unique

1350 VBCs (**Extended Data Fig. 3a**) reduced the number of included CBC-VBC counts by

1351 82.4%. In the single-cell cDNA, cell barcode/UMI sequences are separated from the

1352 VBC cassette by >20 bps - including tracts of A/T homopolymers – providing an

1353 opportunity for mispairing of critical barcode sequences during PCR through strand-

1354 displacement or mispriming. To account for mispairing events, in cells with multiple

1355 VBCs, we developed a collapse algorithm based on fraction of shared UMI sequences

1356 (within edit distance 2) shared across each pair of VBCs. For pairs with >50% UMI

1357 sharing, the "sibling" VBC with fewer UMIs was assigned the VBC of the more

1358 abundant "parent", enforcing that CBC-UMI barcodes should not be used by more

1359 than a single VBC. The ratio of within-cell VBC collapse events due to UMI sharing

1360 versus total CBC-VBC counts averaged 0.13±0.02 (s.e.m) across experiments; a

1361   correction which reduced the number CBC-VBCs counts by an additional 1%. Taken

1362   together, these two VBC collapse steps reduced included CBC-VBC counts by 83.4%

1363   as compared to the raw data - thus drastically altering the inferred groupings of single

1364   cells into networks – and also shaped within-cell VBC quantification, altering the UMI

1365   counts for ~15% of VBCs (Change in VBC UMIs: mean, 5.7; median, 2).

1366   Selective mRNA Adaptor PCR (50 μl reaction)

Reagent	Volume (μl)
10x cDNA (~7-12 ng/μl)	1
P5 Primer (10 mM; P5-10x_Hybrid)	1
P7 Primer (10 μM; P7i`x`-GFP or P7i`x`-TCB_CreOn)	1
Q5 High-Fidelity 2x Master Mix (New England Biolabs, M0492L)	25
Water	22

1367   <u>Cycle Conditions</u>

1368   <u>1.</u> 98°C for 30s; <u>2.</u> 98°C for 10s; <u>3.</u> 72°C for 30s; <u>4.</u> 72°C for 60s; <u>5.</u> Go to Step 2-4,

1369   (see above for rabies virus or rAAV cycle number ranges); <u>6.</u> 72°C for 120s

1370

1371   **Identification, clustering and analysis of host cell scRNA profiles.** To discover

1372   the molecular identities of SBARRO cells, we first distinguished single-cell RNA

1373   libraries from background by leveraging properties of both single-cell RNA and VBC

1374   data from individual experiments. Specifically, using total single-cell RNA data, we

1375   identified cell profiles 1) exclusively associated with cell barcodes provided by 10x

1376   genomics (corresponding to v2 or v3 chemistry) and exhibiting 2) large UMI counts

1377   and low fractions of mitochondrial and ribosomal transcripts, as described

1378   previously[30]. In parallel, we used the mutation-collapsed VBC data (see above) to

1379   filter and retain those cell profiles with at least a single VBC ascertained with >= 3 (v2

1380   chemistry) or >= 5 (v3 chemistry) UMI counts. We used the union of cell barcodes

1381   identified by RNA-based and VBC-based methods to generate digital gene expression

1382   matrices (DGEs) for each experiment[56]. DGEs were input into a two-staged analysis

1383   pipeline based on independent components analysis (ICA)[30], a semi-supervised

1384   approach for grouping scRNA profiles into clusters then subclusters. scRNA profiles

1385   corresponding to cell-cell doublets and cell outliers were identified, flagged and excluded

1386   from downstream analyses as described previously[30]. The identities of clusters and

1387   subclusters were systematically annotated based on molecular marker expression[30,57].

1388   Prior to ICA analysis, DGEs were pruned of 1) genes present on the rabies virus or

1389   mitochondrial genomes and 2) small scRNA profiles (profiles with >= 500 UMIs (SCC

1390   Experiment) or >= 50 genes (noRG Experiment) were retained) to promote high-quality

1391   clustering based on host cell nuclear gene expression. Additional DGEs (subject to the

1392   same gene and cell filtering criteria) were also generated while including rabies virus

1393   genes to aid in the downstream analyses of rabies virus expression:

1394

Experiment	10x	FACS	Reads (x1000)			UMIs (x1000)			Genes (x1000)		
			Range	*Median*	*Mean*	*Range*	*Median*	*Mean*	*Range*	Median	Mean
No RG	V2	No	3.9-805.5	41.7	81.5	1.5-92.3	7.3	12.7	0.17-9.5	2.9	3.5
SCC (SBARRO)	V3	Yes	0.6-5957.8	30.2	53	501-291	11.1	17.6	0.049-11.7	3.8	3.8
SCC (Control)	V3	No	1.2-833.2	41.2	54.7	0.97-159.9	13.3	18.1	0.29-11.6	4.5	4.6

1395

1396   To enhance molecular identification of cells in the SCC experiments, an integrated

1397   analysis of SBARRO and control cell libraries was performed using LIGER[58]. Control

1398   cells were sampled from total cell suspension not subject to FACS and derived from

1399 physically adjacent culture wells seeded with the same cell suspensions as SBARRO

1400 experiments. Input DGEs for LIGER analysis lacked rabies virus and mitochondrial

1401 genes and were filtered to remove "cell-cell doublet" or "outlier" RNA profiles as

1402 identified by upstream ICA-based analysis. LIGER alignment and clustering

1403 (factorization, k=40, lambda=3; quantile alignment, resolution= 0.4, knn_k=20) results

1404 were visualized with UMAP embedding and systematically annotated using marker gene

1405 expression. Of the 130.5K SBARRO scRNA profiles, 28.4K were grouped into three

1406 clusters (cluster 1, 8 and 13) which contained cells of multiple classes and were defined

1407 by expression signatures related to GO biological processes such as "cytokine-

1408 mediated signaling pathway (GO:0019221)" (cluster 13; adjusted p value < 2.1×10^{-14}) or

1409 "PERK-mediated unfolded protein response (GO:0036499)" (cluster 8; adjusted p value

1410 < 0.001). To clarify the molecular identities of these 28.4K cells, we re-aligned these

1411 SBARRO libraries to the control cells from SCC7 and SCC8 using LIGER (factorization,

1412 k=60, lambda=3; quantile alignment, resolution= 0.4, knn_k=20). The resulting analysis

1413 split SBARRO libraries across 32 (of 37 total) clusters which exhibited molecular marker

1414 expression consistent with known cell populations. Clusters were then systematically

1415 annotated in a manner consistent with the initial LIGER analysis guided by the original

1416 and re-aligned cluster identities of control cells. In total, we identified n= 20 "granular"

1417 cell populations which could be grouped into n=11 "coarse" populations. To identify and

1418 visualize genes differentially expressed across granular populations, we used the

1419 FindMarkers() and DotPlot() functions from Seurat [59,60]. To evaluate which cell

1420 populations were sensitive or recalcitrant to rabies virus infection, for each population,

1421 we performed a chi-square test comparing the number of SBARRO and control scRNA

1422 profiles to the dataset totals and used the resulting residuals as a metric of enrichment

1423 or depletion. To determine the relationship between the SBARRO molecular identities

1424 and cortical neuron types from the adult mouse cortex, we used LIGER to jointly analyze

1425 scRNA profiles from 78.8K uninfected SBARRO control cells and 57K cells from various

1426 neocortical regions ascertained by the Allen Institute[18]. We conducted separate analyses

1427 for glutamatergic and GABAergic neurons using the same LIGER parameters

1428 (factorization, k=10, lambda=30; quantile alignment, resolution= 0.2, knn_k=400).

1429

1430 **Transcriptional identification of starter and presynaptic cells.** Starter cells are

1431 functionalized after Cre-mediated recombination inverts TVA-mCherry and rabies

1432 virus B19G transgenes within the FLEX rAAV genome into the sense orientation with

1433 respect to the CAG promoter[61,62]. RNA-based identification of starter cells in a direct

1434 and qualitative manner is complicated using these vectors in the context of 3' scRNA-

1435 seq since 1) low expression caused Cre mRNAs not to be captured with high-

1436 probability in Cre+ cells (caused by the limiting MOI of the Syn1-EBFP-Cre rAAV ($\sim$

1437 8×10^{-3}- 8×10^{-5}); mild RNA Polymerase II recruitment with the *Synapsin1* promoter; and

1438 lack of 3' motifs to promote mRNA stability) and 2) the vast majority of scRNA-seq

1439 reads that align to FLEX rAAV genome do so in the 3' UTR, a region unaffected by

1440 recombination. To overcome these limitations and identify starter cells from scRNA

1441 profiles alone, we developed a protocol to selectively amplify and sequence only TVA-

1442 mCherry mRNAs transcribed from Cre-recombined rAAV genomes (see *"Selective*

1443 *mRNA Adaptor PCR"* above) and a downstream informatic approach using these data

1444 to identify rare, candidate starter cells. Specifically, we developed a binomial test with

1445 experiment-specific success rate parameter (updated using an expectation-

1446 maximization-like approach) to identify cells in which both recombined TVA-mCherry

1447 UMIs (relative to total rAAV UMIs) and total rAAV UMIs (relative to host cell RNA

1448 UMIs) were enriched in a manner unlikely to be due to chance (Bonferroni-corrected p

1449 < 0.01). Properties of RNA expression that distinguish identified starter cells from

1450 presynaptic cells in SBARRO experiments – such as the ratio of recombined TVA-

1451 mCherry UMIs / total rAAV UMIs – were observed in scRNA-seq libraries in which

1452 starter cells were physically separated from presynaptic cells using FACS.

1453

1454 **Cell-type-specific synaptic network inference using rabies virus barcodes.** To

1455 facilitate interactive analysis and discovery of cell-type-specific SBARRO networks,

1456 we created an R Shiny program ("Terminal E") which allows dynamic filtering and

1457 plotting of VBCs and VBC-based synaptic networks. Terminal E integrates data from

1458 VBC libraries and individual SBARRO experiments (organized into "collections" that

1459 enable cross-experiment meta-analyses). Library-level data include VBC abundances

1460 and user-defined, library-specific lists of single VBCs or VBC pairs to exclude from

1461 network inference. SBARRO-level data center around the properties of each cell in

1462 each experiment, which, at minimum, include 1) VBC UMI counts; 2) molecular cell

1463 type identities; and 3. starter or presynaptic assignments. Inferred synaptic networks

1464 are collections of cells which share one or more VBC which are statistically likely to

1465 have entered those cells through clonal replication and spread from a single starter

1466 cell infection. Terminal E supports the inference of such networks through two stages

1467 of VBC filtering. During the first stage, filters completely exclude VBCs from network

1468 consideration. We only considered VBCs with >= 7 UMIs (SCC experiments, v3 10x

1469 chemistry) or >= 3 UMIs (No RG Experiments, v2 10x chemistry). We additionally

1470 removed those VBCs with high abundances in genomes of the infecting library (FI

1471 trust score >= 5; see below). We further identified specific VBCs for exclusion by 1)

1472 comparing their presence and abundance in rabies virus libraries to behavior

1473 transducing 17K starter cell scRNA profiles (containing 28.8K founder infections) or 2)

1474 by comparing across n=23 independent SBARRO experiments. We excluded the

1475 named categories of VBCs below based on the following criteria (**Extended Data Fig.**

1476 **6b**):

1477 1. "Felony" VBCs (n=234). Absent from library but observed in > 1 of 5,015

1478 starter cells (Fig. 2f).

1479 2. "Misdemeanor" VBCs (n=78). Present in library, but observed in > 2 starter

1480 scRNA profiles with library frequency $< 10^{-6}$ or observed in > 8 starter scRNA

1481 profiles with library frequency $< 10^{-5.5}$ (Fig. 2f).

1482 3. "Cross Experiment" VBCs (n=239) Absent from library but observed in > 1 of

1483 23 independent SBARRO experiments.

1484

1485 In the second stage, VBCs or VBC sets are filtered such that those retained were

1486 suitably rare enough to enter the experiment through a single starter cell. Critical to

1487 this stage of experiment-specific VBC filtering is an estimate of the total number of

1488 experiment-specific founder infections starter cells; a subset of which lead to cell-to-

1489 cell spread (spreading founder infections). To estimate the number of spreading

1490 founder infections for each experiment, we developed an analytical approached

1491 designed to mimic founder infections by drawing samples of VBCs from the viral

1492 library. We created distributions describing the number of library draws (n=10

1493 replicates, with VBC replacement) required to match the number of unique VBCs

1494 observed in >=2 cells present in each experiment; median values set the experiment-

1495 specific founder infection estimates. To evaluate whether each VBC or VBC set was

1496 suitably rare enough to be included for network inference, we calculated and assigned

1497 an "founder infection trust" (FI trust score). The "FI trust" score equates to the number

1498 of spreading founder infections that could in theory occur for that VBC or VBC set

1499   before a second founder infection was expected (at a given probability) by leveraging

1500   library VBC frequencies:

1501

1502 $\text{FI trust}_p = \log 10(p) / \log 10(1-f)$

1503

1504   Where f is the frequency of an individual VBC in the library (or, for the VBC set, the

1505   inferred frequency calculated by multiplying the frequencies of each VBC member)

1506   and p is the probability of avoiding a second occurrence of the VBC or VBC set. For a

1507   given experiment, VBC or VBC sets that were retained to infer synaptic networks

1508   were those for which FI trust_p > experiment estimates for founder infections. For SCC

1509   experiments, $p = 0.9$. Inferred networks containing a single scRNA profile assigned as

1510   a starter cell were then split into postsynaptic (i.e. the starter) and presynaptic cells;

1511   networks lacking a starter cell were starter-orphaned and all scRNA profiles were

1512   presumed presynaptic. Terminal E facilitates the comparisons of presynaptic network

1513   size (the number of presynaptic cells) and cell type composition across postsynaptic

1514   starter cells of different types. After first stage filtering, VBCs absent from the library

1515   were presumed to be rare and assigned the lowest library frequency value to match

1516   library VBCs counted with a single UMI. After stratifying presynaptic networks by

1517   starter cell type, presynaptic cell type compositions were compared using a Chi-

1518   Square Test. Presynaptic network sizes were assigned to each scRNA profile using

1519   described rules (**Extended Data Fig. 6c**) and presynaptic network sizes were

1520   compared across starter cell types using a Wilcoxon Rank-Sum test.

1521

1522   **Postsynaptic RNAs associated with the number of presynaptic partner cells.**

1523   Postsynaptic RNA profiles from the SCC experiments identified by as one of four

1524   abundant starter cell types (glutamatergic neurons, interneurons, SPNs or astrocytes;

1525   n=144) were binned based on their inferred presynaptic network size: small (2-4

1526   cells); medium (5-6 cells) or large (7+ cells). These bins approximate inflections in the

1527   total distribution of presynaptic network sizes (**Figure 4b**). To determine if viral load

1528   was associated with presynaptic network size category, the fraction of total cellular

1529   mRNA derived from all five rabies virus genes was compared using a Wilcoxon Test.

1530   A similar strategy was used to test for differences in innate immune responses, using

1531   an aggregate expression score derived from the 564 genes (a subset of 646 total

1532   genes for which we detected a transcript) curated as part of the mouse innate immune

1533   response[63]. To identify candidate gene expression differences associated with

1534   presynaptic network size, we compared starter cell RNA profiles associated with

1535   "small" and "large" presynaptic across cell subtypes rather than types, to avoid

1536   confounds due to differences in subtype compositions. Because these comparisons

1537   involved small numbers scRNA profiles (range: 2 to 28 scRNA profiles; mean, 6.3;

1538   median, 4) we used the following strategy to identify genes and contextualize how likely

1539   such differences were likely to arise by chance. First, we used Fisher's Exact Test to

1540   evaluate differences in UMI counts across scRNA profiles aggregated by small or large

1541   presynaptic size category. We considered only those genes with >= 25 UMIs thus

1542   lessening the burden of multiple hypothesis testing and corrected our p value cut < 0.05

1543   by the number of tests completed within each subtype comparison. Second, for the

1544   genes which passed threshold, we used a Wilcoxon Test to determine whether

1545   normalized RNA levels (removing contributions from rabies virus mRNAs before

1546   normalization) differed across the population of individual cells associated with each

1547   presynaptic network size category, using $p < 0.05$ as our cutoff. To help determine which

1548   of the genes we identified were likely due to chance, we repeated this testing procedure

1549 for 100 permuted comparisons in which each starter cell RNA profile was replaced by a

1550 presynaptic cell RNA profile of the same subtype choose at random. Genes identified by

1551 multiple permuted replicates in the same cell subtype were flagged as potentially

1552 spurious and not considered further. Log fold expression changes were calculated after

1553 normalizing the number of gene-specific UMIs by total UMIs for large and small

1554 presynaptic network categories and then scaling the data to 100,000 transcripts after the

1555 addition of a pseudocount.

1556

1557 **Identifying mRNAs correlated with rabies virus transmission across SPN**

1558 **development.** Monocle3[24,64] was used to calculate pseudotime scores for scRNA

1559 profiles of developing SPNs, from neural precursor cells to mature neurons, after

1560 preprocessing (method=PCA; number of dimensions = 10) and alignment SBARRO

1561 and control libraries (alignment_k=3000). To identify modules of genes with similar

1562 expression levels over pseudotime, all genes were first tested for pseudotime-

1563 associated expression using the graph_test() function Genes with q values = 0 were

1564 retained and modules identified using the find_gene_modules() function with

1565 resolution = 0.001. Pseudotime-ordered cells were grouped into 10 bins. For each bin,

1566 1) the fraction of total cells that were SBARRO (rather than control) in origin were

1567 calculated and 2) a meta-control cell was created by summing control cell UMIs and

1568 normalizing such that mRNA expression values summed to 100,000. Pearson

1569 correlations were calculated for each detected gene by comparing SBARRO fractions

1570 and normalized gene expression values across the pseudotime bins. SBARRO-

1571 correlated genes (n=3,309) were defined as those genes in which r >= 0.75.

1572 SBARRO-correlated genes were tested for gene set enrichment using

1573 PantherGO[25,26,65] via EnrichR[66] and SynGO[29] and compared to two sets (n= 1000

1574   replicate gene selections/set) of n=3,309 control genes, selected either 1) from

1575   expression-matched deciles built from the developmentally mature SPN meta-control

1576   cell (bin = 10) or from 2) at random from genes with expressed RNA.