Rabies virus with a destabilization domain added to its nucleoprotein spreads between neurons only if the domain is removed

Lei Jin[1], Makoto Matsuyama[1], Heather A. Sullivan[1], Mulangma Zhu[1], Thomas K. Lavin[1], YuanYuan Hou[1], Nicholas E. Lea[1], Maxwell T. Pruner[1], María Lucía Dam Ferdínez[1], and Ian R. Wickersham[1]

[1]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA

ABSTRACT

Monosynaptic tracing using rabies virus is an important technique in neuroscience, allowing brain-wide labeling of neurons directly presynaptic to a targeted neuronal population. A 2017 article reported development of a noncytotoxic version – a major advance – based on attenuating the rabies virus by addition of a destabilization domain to the C-terminus of a viral protein. However, this modification did not appear to hinder the ability of the virus to spread between neurons. We analyzed two viruses provided by the authors and show here that both were mutants that had lost the intended modification, explaining the paper's paradoxical results. We then made a virus that actually did have the intended modification and found that it did not spread under the conditions described in the original paper – namely, without an exogenous protease being expressed in order to remove the destabilization domain – but that it did spread, albeit with relatively low efficiency, if the protease was supplied. We conclude that the new approach is not robust but that it may become a viable technique given further optimization and validation.

SIGNIFICANCE STATEMENT

Rabies virus, which spreads between synaptically connected neurons, has been one of the primary tools used by neuroscientists to reveal the organization of the brain. A new modification to rabies virus was recently reported to allow the mapping of connected neurons without adverse effects on the cells' health, unlike earlier versions. Here we show that the conclusions of that study were probably incorrect and based on having used viruses that had lost the intended modification because of mutations. We also show that a rabies virus that does retain the intended modification does not spread between neurons under the conditions reported previously; however, it does spread between neurons under different conditions, suggesting that the approach may be successful if refined further.

INTRODUCTION

Viruses have become important tools for neuroscience(1-16), and "monosynaptic tracing" based on rabies virus (5) has become the primary method of labeling neurons directly presynaptic to some targeted group of neurons (17-20). Its core components are, first, selective infection of the targeted neuronal group with a recombinant rabies virus with a deleted gene (which in all work published to date is the "G" gene encoding the glycoprotein that coats the viral envelope) and, second, complementation of the deletion in the targeted starting neurons, by expression of the deleted gene in *trans*. With all of its gene products therefore present in the starting cells, the virus can fully replicate within them and spreads, as wild-type rabies virus does, to cells directly presynaptic to the initially infected neurons. Assuming that G has not been provided in *trans* in these presynaptic cells too, the deletion-mutant ("ΔG", denoting the deletion of G) virus is unable

to spread beyond them, resulting in labeling of just the neurons in the initially targeted population and ones that are directly presynaptic to them (5).

A drawback of these ΔG (or "first-generation" (21)) rabies viruses is that they are cytotoxic (4, 21, 22), which has spurred several labs to develop less toxic versions. Reardon, Murray, and colleagues (22) showed that simply using ΔG rabies virus of a different parent strain — switching from the original SAD B19 strain to the more neuroinvasive CVS N2c strain (23) — decreased toxicity and increased the efficiency of transneuronal spread. Our own group has taken a more drastic approach, recently introducing "second-generation" rabies viruses from which both G and a second gene, "L", encoding the viral polymerase, have been deleted (21). Although in our published work we have only shown that these second-generation, "ΔGL" viruses are efficient means of direct retrograde targeting of projection neurons, it is at least theoretically possible for ΔGL viruses to be used for monosynaptic tracing, if the second deleted gene were also expressed in *trans*.

Taking a quite different approach, Ciabatti et al. (24) introduced "self-inactivating rabies" ("SiR") viruses, which differed from simple first-generation (ΔG, SAD B19 strain) ones by the addition of a destabilization domain to the C-terminus of one of the viral proteins, so that the protein would be rapidly degraded soon after it was produced. Because the protein in question (the nucleoprotein, encoded by the "N" gene) is essential for viral gene expression and replication, its destabilization was intended to "silence" viral gene expression and prevent replication, making the viruses nontoxic.

These SiR viruses were designed to be unable to replicate unless an exogenous protease (tobacco etch virus protease, TEVP) was expressed in infected cells in order to remove the destabilization (or "PEST" (25)) domain. However, the paper reported that they were able to spread between neurons – which requires replication – just as efficiently as unmodified first-generation viruses did, without the protease being provided at all.

We hypothesized that the viruses that were used for the reported transsynaptic tracing experiments (24) were mutants with premature stop codons at or near the end of the native nucleoprotein gene and before the sequence of the destabilization domain. Rhabdoviruses have high mutation rates (26-31), and production of high-titer rabies virus stocks for *in vivo* injection typically involves repeated passaging on complementing cell lines (32-34), which affords ample opportunity for accumulation of mutants with a selective replication advantage.

Here we show that, in both of the two SiR virus samples that we analyzed, the great majority of viral particles did have mutations in their genomes that caused the complete loss of the intended C-terminal addition to the nucleoprotein, so that they were effectively just ordinary first-generation ΔG rabies viral vectors. We also tested the SiR-CRE virus *in vivo* and found that it was rapidly cytotoxic.

We also show that a ΔG virus that does have the intended modification to the nucleoprotein does not spread transsynaptically *in vivo* in the absence of TEVP, but that it does spread when TEVP is provided, although with lower efficiency than a virus without the modification.

RESULTS

We analyzed samples of two viruses sent directly from the Tripodi lab to MIT two months after their publication(24) and given directly to the Wickersham lab soon afterward, still frozen and unopened, with express permission from Marco Tripodi by email on December 5, 2017: "EnvA/SiR-CRE" (made from genome plasmid Addgene 99608, pSAD-F3-NPEST-iCRE-2A-mCherryPEST) and "EnvA/SiR-FLPo" (made from genome plasmid Addgene 99609, pSAD-F3-NPEST-FLPo-2A-mCherryPEST). Both had the SAD B19 strain of rabies virus as their parent

strain and had been packaged with the avian and sarcoma virus subgroup A envelope glycoprotein ("EnvA") for targeted infection of cells expressing EnvA's receptor, TVA (5).

For comparison with the two SiR viruses, we made five control viruses in our own laboratory: three first-generation vectors, RVΔG-4Cre (21), RVΔG-4FLPo (see Methods), and RVΔG-4mCherry (35), and two second-generation vectors, RVΔGL-4Cre and RVΔGL-4FLPo (21). All of these viruses are on the SAD B19 background, like the SiR viruses. For each of the four recombinase-expressing viruses from our laboratory, we made one preparation packaged with the EnvA envelope protein and one preparation packaged with the native rabies virus (SAD B19 strain) glycoprotein (denoted as "B19G"); RVΔG-4mCherry (used only as a control for the Sanger sequencing) was only packaged with the EnvA envelope protein.

Sequencing of viral genomes: Sanger sequencing

In order to directly test our hypothesis that the SiR viruses had developed premature stop codons removing the PEST domain in a majority of viral particles, we sequenced the genomes of a large number of individual viral particles using two different techniques.

First, we used ordinary Sanger sequencing to determine the sequence in the vicinity of the end of the nucleoprotein gene for 50 to 51 individual viral particles of each of the two SiR viruses and of a first-generation virus from our own laboratory, RVΔG-4mCherry (Figure 1). We ensured the isolation of individual viral genomes by using a primer with a random 8-base index for the reverse transcription step, so that the cDNA copy of each RNA viral genome would have a unique index. Following the reverse transcription step, we amplified the genomes by standard PCR, cloned the amplicons into a generic plasmid, transformed this library into E.coli and sequenced plasmids purified from individual colonies.

As shown in Figure 1, the results confirmed our hypothesis that SiR viruses are prone to loss of the 3' addition to the nucleoprotein gene. Specifically, in the SiR-CRE sample, 100% of the 51 sequenced viral particles had lost the PEST domain. Fifty out of the 51 had the same point mutation in the linker between the end of the native nucleoprotein gene and the TEVP cleavage site, converting a glycine codon (GGA) to a stop codon (TGA) so that the only modification to the C-terminus of the nucleoprotein was the addition of two amino acids (a glycine and a serine). The one sequenced viral particle that did not have this point mutation had a single-base insertion in the second-to-last codon of the native nucleoprotein gene, frameshifting the rest of the sequence and resulting in 15 amino acids of nonsense followed by a stop codon before the start of the PEST domain sequence.

In the SiR-FLPo sample, the population was more heterogeneous: out of 50 sequenced viral particles, 18 had the same stop codon that was found in almost all genomes in the Cre sample, while another 28 had a different stop codon three amino acids upstream, immediately at the end of the native nucleoprotein gene (converting a serine codon (TCA) to a stop codon (TGA)). Four viral particles had no mutations in the sequenced region. Thus 46/50 (92%) of the SiR-FLPo viral particles sequenced had lost the PEST domain.

In contrast, in the first-generation virus from our own lab, RVΔG-4mCherry, none of the 50 viral particles sequenced had mutations in the seque"""//////nced region containing the end of the nucleoprotein gene.

Sequencing of viral genomes: Single-molecule, real-time (SMRT) sequencing

As a second approach to analyzing the mutations present in the SiR viruses, we employed a large-scale sequencing technology: single-molecule, real-time ("SMRT") sequencing, which provides independent sequences of tens of thousands of individual single molecules in a sample in parallel (Figure 2). The results from this advanced sequencing method were quite consistent with the results from the Sanger sequencing presented above. As with the sample preparation for

153 Sanger sequencing, we included a random index (10 bases, in this case) in the reverse
154 transcription primer, so that again the cDNA copy of each RNA viral genome molecule would be
155 labeled with a unique index.
156 SMRT sequencing entails circularization of the DNA amplicons and multiple consecutive
157 passes around the resulting circular molecule, with the redundancy provided by this repeated
158 sequencing of each position increasing the signal to noise ratio and statistical significance of the
159 results. The numbers presented in Figure 2 and below use the default of including only clones
160 that had at least three reads of each base ("circular consensus sequence 3", or "CCS3" in
161 Supplementary File S3). Using the increasingly stringent criteria of requiring either five or eight
162 reads per base (CCS5 or CCS8) reduced the numbers of qualifying genomes in all cases and
163 changed the percentages slightly but gave very similar results overall. Because read accuracy for
164 SMRT sequencing is ≥98% for circular consensus sequencing with 3 passes (see
165 https://www.mscience.com.au/upload/pages/pacbio/technical-note---experimental-design-for-
166 targeted-sequencing.pdf), we used a conservative threshold of 2% frequency of any given point
167 mutation position in order to screen out false positives. Also to be very conservative, for Figure 2
168 we ignored all apparent frame shifts caused by insertions and deletions, because insertions in
169 particular are prone to false positives with SMRT sequencing (36). See Supplementary File S3
170 for details, including details of frameshifts due to insertions; Supplementary Files S4-S6 contain
171 the sequences of the PCR amplicons that would be expected based on published sequences of
172 the three viruses, but to summarize here:
173 As a control, we used a virus from our own laboratory, RVΔG-4Cre (21) (see Addgene
174 #98034 for reference sequence). Out of 17,978 sequenced genomes of this virus, we found no
175 mutations above threshold frequency at the end of N. We did find that 1,706 viral particles (9.49%)
176 had a nonsynonymous mutation (TCT (Ser) → ACT (Thr)) farther up in N at amino acid position
177 419 (31 amino acids upstream of the end of the 450-aa native protein). We do not know if this
178 mutation is functionally significant, although it is not present in CVS N2c (37), HEP-Flury (38),
179 ERA (39), or Pasteur strains (Genbank GU992320), so these particles may effectively be N-
180 knockouts that were propagated by coinfection with virions with intact N (see Discussion for more
181 on such parasitic co-propagating mutants).
182 For the SiR-CRE virus, out of 22,205 viral genomes sequenced, 22,032 had the premature
183 stop codon (GGA -> TGA) in the linker between the native nucleoprotein gene and the TEVP
184 cleavage site sequence. In other words, even without including frameshifts, at least 99.22% of
185 the individual viral particles in the SiR-CRE sample were essentially first-generation ΔG vectors,
186 with the only modification of the nucleoprotein being an additional two amino acids at the C-
187 terminus.
188 For the SiR-FLPo virus, out of 17,086 viral genomes sequenced, 5,979 had the stop codon
189 (GGA -> TGA) in the linker, 8,624 had the stop codon (TCA -> TGA) at the end of N, and a further
190 28 had a different stop codon (TCA -> TAA) at the same position at the end of N. Of these, 305
191 viral particles had premature stop codons at both of these two positions, so that the total number
192 of viral particles with one or both stop codons immediately before the PEST domain was (8624 +
193 5979 + 28 - 305 = 14,326. In other words, again even without including frameshifts, at least
194 83.85% of the individual viral particles in the SiR-FLPo sample were essentially first-generation
195 ΔG vectors, with the only modification of the nucleoprotein being either two amino acids added
196 to, or one amino acid lost from, the C-terminus.
197
198 **Anti-nucleoprotein immunostaining**
199
200 We infected reporter cell lines with serial dilutions of the two EnvA-enveloped SiR viruses
201 as well as the eight recombinase-expressing ones from our own lab: ΔG vs. ΔGL, Cre vs. FLPo,
202 EnvA vs. B19G envelopes. Three days later, we immunostained the cells for rabies virus
203 nucleoprotein and imaged the cells with confocal microscopy.

As seen in Figure 3, we found that the cells infected with the SiR viruses looked very similar to those infected with the first-generation, ΔG viruses. Notably, the viral nucleoprotein, which in the SiR viruses is intended to be destabilized and degrade rapidly in the absence of TEVP, accumulated in the SiR-infected cells in clumpy distributions that looked very similar to those in the cells infected with the first-generation, ΔG viruses. By contrast, the cells infected with the second-generation, ΔGL viruses, which we have shown to be noncytotoxic (21), did not show any such nucleoprotein accumulation, clumped or otherwise, only punctate labeling presumably indicating isolated viral particles or post-infection uncoated viral particles (ribonucleoprotein complexes) that are not replicating.

Longitudinal two-photon imaging *in vivo*

To see whether the SiR viruses kill neurons in the brain, we conducted longitudinal two-photon imaging *in vivo* of virus-labeled neurons in visual cortex of tdTomato reporter mice, as we had done previously to demonstrate the nontoxicity of second-generation rabies virus (21) (Figure 4). Because the SiR viruses were EnvA-enveloped, we first injected a lentivirus expressing EnvA's receptor TVA, then one week later we injected either SiR-CRE or one of two EnvA-enveloped viruses made in our laboratory: the first-generation virus RVΔG-4Cre(EnvA) or the second-generation virus RVΔGL-4Cre(EnvA). Beginning one week after rabies virus injection, we imaged labeled neurons at the injection site every seven days for four weeks, so that we could track the fate of individual neurons over time.

As we found in our previous work (21), our second-generation virus RVΔGL-4Cre did not kill neurons to any appreciable degree: all but a tiny handful of the neurons labeled by this virus at seven days after injection were still present three weeks later in all mice. Also as we have found previously (21), our first-generation virus RVΔG-4Cre did kill neurons, but by no means all of them (see the Discussion for possible reasons for this).

However, we found that the putatively nontoxic SiR-CRE caused a steep loss of neurons much more pronounced than even our first-generation virus did. By 14 days after injection, 70% of cells seen at seven days were dead; by 28 days, 81% were.

There is a possible confound from our use of the tdTomato reporter line Ai14 (which we used primarily because we already had large numbers of mice of this line): because SiR-CRE is actually "SiR-iCRE-2A-mCherryPEST", designed to coexpress mCherry (with an added C-terminal PEST domain intended to destabilize it, as for the nucleoprotein) along with Cre, it is conceivable that some of the SiR-CRE-labeled red cells at seven days were only expressing mCherry and not tdTomato. If the destabilized mCherry were expressed only transiently, as intended(24), and a significant fraction of SiR-CRE virions had mutations in the Cre gene so that they did not express functioning Cre, then it is possible that some of the red cells seen at seven days were labeled only with mCherry that stopped being visible by 14 days, so that it would only look like those cells had died.

We viewed this alternative explanation as unlikely, because Ciabatti et al. injected SiR-CRE in an EYFP reporter line and found no cells labeled only with mCherry and not EYFP at six days and nine days postinjection (see Figure S4 in Ciabatti et al.). Nevertheless, we addressed this potential objection in several ways.

First, we sequenced the transgene inserts (iCre-P2A-mCherryPEST) of 21 individual SiR-CRE viral particles (see Supplementary File S1) and found that only two out of 21 had mutations in the Cre gene, suggesting that there would not have been a large population of cells only labeled by mCherry and not by tdTomato.

Second, we repeated some of the SiR-CRE injections and imaging in a different reporter line: Ai35, expressing Arch-EGFP-ER2 after Cre recombination (40) (Jax 012735). Although we found that the membrane-localized green fluorescence from the Arch-EGFP-ER2 fusion protein was too dim and diffuse at seven days postinjection to be imaged clearly, we were able to obtain

clear images of a number of cells at 11 days postinjection. We found that 46% of them had disappeared only three days later (see Figure 5 and Supplementary Video S1), and 86% had disappeared by 28 days postinjection, consistent with a rapid die-off. Furthermore, we found that the red fluorescence in Ai35 mice, which was due only to the mCherry expressed by the virus, was much dimmer than the red fluorescence in Ai14 mice at the same time point of seven days postinjection and with the same imaging parameters (see Supplementary Figure S1): the mean intensity was 45.86 (arbitrary units, or "a.u.") in Ai14 but only 16.29 a.u. in Ai35. This is consistent with the published findings that tdTomato is a much brighter fluorophore than mCherry (41), particularly with two-photon excitation (42), and it is also consistent with Ciabatti et al.'s addition of a destabilization domain to mCherry's C-terminus (although we also observed that there were still red fluorescent cells present at week four in Ai35 (data not shown)). We therefore redid the counts of labeled cells in our Ai14 datasets to include only cells with fluorescence at seven days of more than 32.33 a.u., the midpoint of the mean intensities in Ai35 versus Ai14 mice, in order to exclude neurons that might have been labeled with mCherry alone. As seen in Figure S2, restricting the analysis to the cells that were brightest at seven days (and therefore almost certainly not labeled with just mCherry instead of either just tdTomato or a combination of both mCherry and tdTomato) made no major difference: 70.0% of SiR-labeled neurons had disappeared by 14 days, and 80.8% were gone by 21 days.

Although in theory it is possible that the disappearance of the infected cells could be due to cessation of tdTomato or Arch-EGFP-ER2 expression rather than to the cells' deaths, because of downregulation by rabies virus of host cell gene expression (43), we view this as highly unlikely. Downregulation of host cell gene expression by rabies virus is neither total ("cells with high expression of RbV transcripts retain sufficient transcriptional information for their classification into a specific cell type." (43)) nor uniform (44); in practice, we saw no evidence of a decline in reporter expression in the infected cells but in fact found the exact opposite. As can be seen in a number of cells in Figures 2 and S3, the cells got brighter and brighter over time, unless they abruptly disappeared. In our experience, including in this case, cells infected with rabies virus increase in brightness until they die, often blebbing and coming apart into brightly labeled pieces, regardless of whether the fluorophore is expressed from a reporter allele (as in this case) or directly by the virus (see Chatterjee et al. 2018 for many more examples of this (21)).

Construction and testing of a virus with an intact PEST domain

We decided to directly test whether a rabies virus with an intact PEST domain fused to its nucleoprotein can spread transsynaptically, with or without TEVP (Figure 6). Beginning with our lab's first-generation virus RVΔG-4Cre(21), we constructed a "self-inactivating" version, "RVΔG-NPEST-4Cre", by adding the coding sequence for the C-terminal addition from Ciabatti et al.(24) to the 3' end of the nucleoprotein gene. To reduce the chance of the PEST domain being lost to nonsense mutations during production of the virus, we made synonymous changes to five codons near the junction of the end of the native nucleoprotein gene and the beginning of the addition, so that those codons were no longer a single mutation away from being stop codons; apart from these five synonymous changes, the nucleotide sequence of the addition was identical to that used by Ciabatti et al.(24) (Figure 6A). We also made a matched "revertant" version, "RVΔG-N*PEST-4Cre", with exactly the same sequence as RVΔG-NPEST-4Cre except with a stop codon three codons into the linker, at the same location as the stop codon that we had found in the overwhelming majority of viral particles in the original SiR-CRE. While the genomes of these two new viruses only differed from each other by one codon, therefore, at the protein level one virus (RVΔG-NPEST-4Cre) had a full-length PEST domain on the end of its nucleoprotein, while the other (RVΔG-N*PEST-4Cre) had only a two-amino-acid (Gly-Ser) addition to the end of its nucleoprotein and otherwise was an ordinary first-generation ΔG virus.

305 Following production of high-titer EnvA-enveloped virus (see Methods), we confirmed that
306 the final stocks retained the intended 3' additions to the nucleoprotein gene by extracting the
307 genomic RNA and used Sanger sequencing on the genomes of 32 viral particles for each virus.
308 All of the clones sequenced for each of the two viruses had the respective intended modifications
309 (Figure 6, panel A; details in Supplementary File S7), except for one N*PEST clone that had a
310 synonymous mutation in the PEST domain after the introduced stop codon (and that was
311 therefore irrelevant. We also found three other incidental mutations in one clone each: one
312 NPEST clone had a synonymous mutation in the nucleoprotein gene (at Ser437, from TCA to
313 TCG), one N*PEST clone had a different synonymous mutation in the nucleoprotein gene (at
314 Asn436, from AAC to AAT), and one NPEST clone had a single point mutation in the intergenic
315 region between the N and P genes).
316 Having verified that the NPEST and N*PEST viruses retained their respective
317 modifications, we tested their ability to spread transsynaptically *in vivo* in the absence of TEVP,
318 using corticostriatal neurons as the starting cells (Figure 6, panel B). We injected an AAV2-
319 retro(11) expressing FLPo into the dorsolateral striatum of three mice each of the tdTomato
320 reporter line Ai14(45), with a cocktail of two helper viruses (AAV1)(46-48) injected into the primary
321 somatosensory cortex (barrel field) in the same surgery. Corticostriatal neurons were therefore
322 coinfected with all three viruses for expression of TVA (to allow infection by EnvA-enveloped RV)
323 and G (to allow spread of the ΔG viruses to presynaptic cells), as well as the fluorophores EGFP
324 (marking TVA expression) and mTagBFP2 (marking G expression). One week after AAV injection,
325 we injected either the NPEST or the N*PEST virus at equalized titers, then perfused the mice
326 either 12 days or 3 weeks after the rabies virus injections. Figure 6 shows the results.
327 The N*PEST virus, which had only an additional two amino acids on the C-terminus of its
328 nucleoprotein, spread very efficiently (bottom rows of images in panels B and C, and rightmost
329 bar in each pair in the charts in panel E). We counted tdTomato-labeled neurons in ipsilateral
330 cortex as well as in thalamus and in contralateral cortex (note that we counted neurons only in
331 every sixth 50 μm section (see Methods), so that the numbers of labeled neurons in the entire
332 brain of each mouse would be approximately six times the numbers given below and in the
333 figures). At 12 days, we found an average of 2,772 tdTomato-labeled neurons in ipsilateral cortex;
334 we also found an average of 12 labeled neurons in contralateral cortex and 40 in thalamus. At 3
335 weeks after rabies virus injection, a time point which is unusually long for a first-generation
336 virus(21) but which we included to match the duration used by Ciabatti et al.(24), the N*PEST
337 virus had spread to vastly more neurons: on average 13,070 in ipsilateral cortex, 253 in
338 contralateral cortex, and 507 in thalamus. Importantly, all this label was not due simply to leaky
339 TVA expression or residual RVG-enveloped virus, as control experiments without G resulted in
340 very minimal labeling (Supplementary Figure S6), but instead indicates efficient transsynaptic
341 spread of the N*PEST virus, consistent with our lab's prior experience with the parent virus RVΔG-
342 4Cre(EnvA) as well as with other first-generation RVΔG vectors.
343 In contrast, the NPEST virus, which had the intact PEST domain fused to its nucleoprotein,
344 showed no clear evidence of transsynaptic spread without TEVP at either time point examined
345 (top rows of images in panels C and D, and leftmost bar in each pair in the charts in panel E). At
346 12 days, we found averages of 10 labeled cells at the injection site, 2 in contralateral cortex, and
347 none at all in thalamus. At 3 weeks, the situation was not much different: we found averages of
348 30 labeled cells at the injection site, 0.333 in contralateral cortex, and still zero in thalamus. These
349 numbers were not significantly different from those in the matched controls without G, for which
350 the means at 12 days were 4.667, 0.667, and 0 (injection site, contralateral cortex, and thalamus,
351 respectively) and the means at 3 weeks were 12.333, 0, and 0. Single-factor ANOVAs were used
352 for all comparisons; all cell counts and results of statistical analyses are given in Supplementary
353 File S8..
354 We then tested the ability of the two viruses to spread when both TEVP and G are supplied
355 (Figure 7). These experiments were done in exactly the same way as the ones without TEVP but

with a third virus, AAV1-TREtight-H2b-emiRFP670-TEVP, included in the helper virus mixture. This third helper virus was of the same design as the G-expressing AAV (AAV1-TREtight-mTagBFP2-B19G) but expressed TEVP (S219V mutant) instead of G and an H2b-fused near-infrared fluorescent protein (emiRFP670) instead of the blue fluorophore mTagBFP2.

With both TEVP and G provided, the NPEST virus did show evidence of spread: at 12 days, there were on average 469 labeled cells in ipsilateral cortex, 5 in contralateral cortex, and 3 in thalamus; at 3 weeks, there were on average 1,257 in ipsilateral cortex, 4 in contralateral cortex, and 18 in thalamus.

The N*PEST virus still labeled many more cells than the NPEST one did under these conditions (panels C and D). At 12 days, the N*PEST virus had labeled on average 4.5 times as many cells in ipsilateral cortex (2,090 cells), twice as many in contralateral cortex (10 cells), and 13,3 times as many in thalamus (40 cells); at 3 weeks, it had labeled 2.7 times as many cells in ipsilateral cortex (3,397 cells), 9.3 times as many in contralateral cortex (37 cells), and 2.4 times as many in thalamus (43 cells). While these differences were not found to be statistically significant, the comparisons were presumably underpowered, due to the very low number of subjects (n=3) we used per group, which was dictated by the limited availability of Ai14 mice.

Comparing the best conditions for the NPEST virus (that is, with TEVP provided) to the best conditions for the N*PEST one (that is, without TEVP provided) gave very large and highly significant differences for almost all comparisons (panels E and F). At 3 weeks, for example, the N*PEST virus had labeled 10.3 times as many cells in ipsilateral cortex (averages of 13,070 vs 1,257; p=0.000119), 14.3 times as many in thalamus as the NPEST virus (252.6 vs 17.6; p=0.00433), and 117.8 times as many cells in contralateral cortex (506.6 vs 4.3; p=0.00550).

DISCUSSION

Our transsynaptic tracing results using virus with an intact PEST domain contradict the earlier claim that such a virus can spread between neurons in the absence of TEVP(24). Our findings from sequencing the SiR viruses from the originating laboratory suggest that the reason for that claim was that the viruses used in the original study had lost the intended modification.

While it is possible that the escape of the two SiR virus samples from the modification intended to attenuate them was a fluke due to bad luck with those two batches, we view this as unlikely, for three reasons. First, the reported finding that viral replication and spread occurred in the absence of TEVP is difficult to understand in the absence of mutations but is easily explained if the viral preparations used for those experiments harbored the kind of mutations that we found in the two preparations to which we had access. Second, the two SiR virus samples that we analyzed had independently developed mutations causing loss of the intended C-terminal addition to the nucleoprotein; we know that the mutations were independent because the two samples were of different viruses so did not both derive from a single compromised parental stock. Third, the mutation profiles of the two viruses were very different: whereas the SiR-CRE sample had the same point mutation in nearly 100% of its viral particles, only a minority of the SiR-FLPo particles had that particular mutation, with the majority having a different point mutation three codons away that had the same result. This suggests that any of the many opportunities for removing the C-terminal addition — creation of a premature stop codon at any one of a number of sites, or a frameshift mutation anywhere in the vicinity — can be exploited by a given batch of virus, greatly increasing the probability of such mutants arising.

While it is clearly possible to make virus with the intended modification to the nucleoprotein, because we have done so here (and the authors of the original paper also report doing so, in a recent preprint (49) in response to our own preprint of an earlier version of this paper), our findings suggest that the approach in its current form is vulnerable to being undermined by viral mutation. Although a number of groups have made recombinant rabies

407 viruses — as well as other rhabdoviruses and other nonsegmented negative-strand RNA viruses
408 — encoding fusions of exogenous proteins to viral proteins (50-60), most of these groups have
409 found that the additions significantly impaired function, and some have found that the viruses
410 rapidly lost C-terminal additions to viral proteins. For example, an attempt to make SAD B19
411 rabies virus with EGFP fused to the C-terminus of the nucleoprotein was unsuccessful, suggesting
412 that large C-terminal additions make the nucleoprotein dysfunctional; the authors of that paper
413 resorted instead to making virus encoding the fusion protein in addition to the wild-type
414 nucleoprotein (53). A vesicular stomatitis virus with GFP fused to the C-terminus of the
415 glycoprotein gene lost the modification within a single passage of the virus because of a point
416 mutation creating a premature stop codon (50). Relatedly, a VSV with its glycoprotein gene
417 replaced with that of a different virus was found to quickly develop a premature stop codon
418 causing loss of the last 21 amino acids of the exogenous glycoprotein, conferring a marked
419 replication advantage to the mutants bearing the truncated version (61).
420 Generalizing from these prior examples as well as our findings here, we suggest that any attempt
421 to attenuate a virus by addition to the C-terminus of a viral protein will be vulnerable to loss of the
422 modification, and that any such virus will therefore need to be monitored very carefully.
423 It is unclear whether further improvements to the design of the viruses can be made to
424 make loss of the PEST domain less likely. Although the synonymous mutations that we made to
425 five codons near the junction of N and the 3' addition are a good start, there are numerous other
426 codons in the immediate vicinity that are also one point mutation away from being stop codons
427 but that do not have synonyms that are not. Furthermore, no such changes would protect against
428 frameshifts.
429 If the viruses used for the transsynaptic tracing experiments in the original paper were
430 actually de facto first-generation, ΔG viruses like the SiR samples that we analyzed, how could
431 the authors have found, in postmortem tissue, cells labeled by SiR-CRE that had survived for
432 weeks? The answer may simply be that, as we have shown in Chatterjee et al. (21) and again
433 here (Figure 4), and as the original authors also report in their new preprint (49), a preparation of
434 first-generation rabies viral vector expressing Cre can leave a large fraction of labeled cells alive
435 for at least months, in contrast to similar ones encoding tdTomato (21) or EGFP (4). Similarly,
436 Gomme et al. found long-term survival of some neurons following infection by a replication-
437 competent rabies virus expressing Cre (62)). Our results with the "revertant" control virus RVΔG-
438 N*PEST-4Cre (Figures 6 and 7, and Supplementary Figures S4-S6), showing thousands of
439 labeled neurons three weeks after rabies virus injection, are consistent with this as well.
440 One reason, in turn, why a preparation of a simple ΔG rabies virus encoding Cre can leave
441 many cells alive may be that not all the virions are in fact first-generation viral particles, because
442 of the high mutation rate that we have highlighted in this paper. We have shown in Chatterjee et
443 al. (21) that a second-generation (ΔGL) rabies virus, which has both its glycoprotein gene G and
444 its polymerase gene L deleted, leaves cells alive for the entire four months that we followed them.
445 However, any first-generation (ΔG) virus that contains a frameshift or point mutation knocking out
446 L will in practice be a ΔGL virus. Indeed, a stop codon or frameshift mutation in any of several
447 other viral genes is likely to have a similar effect as one in L (and it might be that the Ser419Thr
448 mutation that we found in 9.49% of our RVΔG-4Cre virions is just such a knockout mutation of N).
449 Together with the high mutation rate of rabies virus, this means that, within every preparation of
450 first-generation rabies virus there is almost guaranteed to be a population of de facto second-
451 generation variants mixed in with the intended first-generation population and propagated in the
452 producer cells by complementation by the first-generation virions. Any rabies virus preparation
453 (whether made in the laboratory or occurring naturally) can be expected to contain a population
454 of such knockout (whether by substitution, frameshift, or deletion) mutants (related to the classic
455 phenomenon of "defective interfering particles", or mutants with a marked replication advantage
456 (63-65), and the higher the multiplicity of infection when passaging the virus, the higher the
457 proportion of such freeloading viral particles typically will be. This would not necessarily be noticed

458 in the case of a virus encoding a more common transgene product such as a fluorophore, because
459 the expression levels of these by the knockout mutants would be too low to label cells clearly (see
460 Figure 1 in Chatterjee et al. (21)). However, with Cre as the payload, any "second-generation"
461 particles would be able to label neurons but not kill them, because second-generation rabies viral
462 vectors do not kill cells for at least months (21). This explanation would predict that the percentage
463 of neurons surviving infection with a rabies virus encoding Cre will depend on the particular viral
464 preparation that is injected, with some having a greater fraction of knockout particles than others.
465 This could explain why the SiR-CRE virus sample killed cells faster than our own RVΔG-
466 4Cre (Figure 4). This analysis would also presumably apply to first-generation (ΔG) viruses
467 expressing FLPo: while we found that the FLPo-expressing version that we made did not leave
468 as many cells alive as the Cre-expressing version (Supplementary Figure S3) did, that preparation
469 may simply have had fewer mutants with knockout of genes essential for replication.
470 On a positive note, we found that virus with an intact PEST domain does spread
471 between neurons when TEVP is supplied. This suggests that such viruses could in fact become
472 the basis for monosynaptic tracing systems with reduced toxicity. Our finding is complementary
473 to those of the original authors in their recent preprint (49) that an intact SiR virus did not kill
474 labeled neurons for five months: that study was of neurons directly infected by B19G-enveloped
475 virus (i.e., without using EnvA-enveloped virus, expression of TVA to mediate its selective
476 infection, expression of G to complement the ΔG virus, or expression of TEVP to remove the C-
477 terminal addition to the nucleoprotein); they did not show that their intact virus could spread
478 between neurons or that it is nontoxic as it does so. Conversely, we showed here that an intact
479 NPEST virus can spread between neurons, but we did not assay toxicity during this process. It
480 remains to be seen to what degree an NPEST or SiR virus that spreads transsynaptically is
481 cytotoxic, not only to transsynaptically-labeled cells but also to the starting cells, which need to
482 express both G and TEVP to allow replication and spread of the virus. This is not a trivial point,
483 as G is toxic when overexpressed (66), and the original authors' finding that cultured cells
484 rapidly lost TEVP activity (49) suggests that TEVP expressed at sufficient levels may be toxic as
485 well.
486 Although we found that the "intact" NPEST virus spread considerably less than that of
487 the "revertant" control virus, there may be room for improvement: the TEVP expression levels
488 that we engineered are unlikely to happen to be optimal, and the numbers of labeled neurons
489 could increase further with longer survival times. On the other hand, if overexpression of TEVP
490 causes toxicity (see above), increasing it beyond what we achieved could be deleterious.
491 Another consideration is the nine additional amino acids that are unavoidably left on the C-
492 terminus of the nucleoprotein after the rest of the addition is removed by TEVP: while these
493 additional amino acids may not much impair the function of the protein, they are unlikely to help
494 it.
495 It is also unclear to what degree a minority population of revertant mutants that did arise
496 in stocks of otherwise-intact virus would pose a problem for monosynaptic tracing studies. If, for
497 example, ~5% of the virions in a given preparation were revertant mutants (e.g., as the original
498 authors report obtaining after six passages in cells highly expressing TEVP (49)), one might
499 expect the same percentage of labeled presynaptic neurons to be labeled by those mutants and
500 therefore to experience the toxicity of infection by a first-generation, ΔG virus. However, because
501 the process of infecting the starter cells, replicating within them (if provided with G), and spreading
502 to other cells is comparable to an additional passage in cell culture, the percentage of presynaptic
503 cells labeled by the revertant mutants could be higher, and would presumably depend on the level
504 of TEVP expression in each starter cell.
505 In summary, our results suggest that rabies virus with a PEST domain added to its
506 nucleoprotein only spreads between neurons if the PEST domain is removed, whether by
507 expression of TEVP or by mutation. Our finding that virus with an intact PEST domain can spread
508 when TEVP is provided, as the designers had presumably originally intended, raises the

possibility that further optimization and validation could make the SiR approach a viable option for monosynaptic tracing with reduced toxicity.

METHODS

Cloning

Lentiviral transfer plasmids were made by cloning, into pCSC-SP-PW-GFP (67) (Addgene #12337), the following components:

the CAG promoter (68) and a Cre-dependent "FLEX" (6) construct consisting of pairs of orthogonal lox sites flanking a back-to-back fusion of the gene for mTagBFP2 (69) immediately followed by the reverse-complemented gene for mCherry (41), to make the Cre reporter construct pLV-CAG-FLEX-BFP-(mCherry)' (Addgene 115234);

the CAG promoter (68) and a Flp-dependent "FLEX" (6) construct consisting of pairs of orthogonal FRT sites (70) flanking a back-to-back fusion of the gene for mTagBFP2 (69) immediately followed by the reverse-complemented gene for mCherry (41), to make the Flp reporter construct pLV-CAG-F14F15S-BFP-(mCherry)' (Addgene 115235);

the ubiquitin C promoter from pUB-GFP (71) (Addgene 11155) and the long isoform of TVA (72) to make the TVA expression vector pLV-U-TVA950 (Addgene 115236).

The first-generation vector genome plasmid pRV∆G-4FLPo (Addgene 122050) was made by cloning the FLPo gene (73) into pRV∆G-4Cre.

pAAV-synP-F14F15S-splitTVA-EGFP-tTA (Addgene 136917) is a FLP-dependent version of the Cre-dependent helper virus genome plasmid pAAV-synP-FLEX-splitTVA-EGFP-tTA (Addgene 52473) with orthogonal FRT sites(70) instead of orthogonal lox sites.

pCAG-hypBase (Addgene 174376) was made by synthesizing the 1785-bp gene for the "hyperactive" piggyBac transposase iPB7 (74) and cloning it into the EcoRI and NotI sites of pCAG-GFP (71) (Addgene 11150).

pB-CAG-TEVP-IRES-mCherry (Addgene 174377) was made by cloning the CAG promoter (68), a mammalian codon-optimized version (75, 76) of the TEVP gene (S219V mutant) (77), the EMCV IRES (78), and the mCherry gene (41) into pB-CMV-MCS-EF1-Puro (System Biosciences #PB510B-1).

pAAV-synP-FLPo (Addgene 174378) was made by cloning the FLPo gene(73) into the EcoRI and AccIII sites of pAAV-synP-FLEX-EGFP-B19G (Addgene 59333).

pAAV-TREtight-H2b-emiRFP670-TEVP (Addgene 174379) was made by cloning an H2b-emiRFP670 fusion gene (79) (sequence from Addgene 136571 but with the internal kozak sequence replaced by a short GSG linker to prevent translation of fluorophore unfused to H2b), followed by a P2A sequence and the above-described TEVP gene, into the EcoRI and NheI sites of pAAV-TREtight-mTagBFP2-B19G (46) (Addgene 100799).

pRV∆G-NPEST-4Cre (Addgene 174380) was made by cloning a synthesized fragment containing the 3' addition from Ciabatti et al. '17, with synonymous changes to 5 codons (24) in the immediate vicinity of the junction between the nucleoprotein gene and the 3' addition so that more than a

560 single point mutation would be required to convert them to stop codons, into the PmlI and BstI
561 sites of pRVΔG-4Cre(21) (Addgene 98034).
562
563 pRVΔG-N*PEST-4Cre (Addgene 174381) was made identically to the above except that the
564 glycine codon in position 453 was replaced with a stop codon (TGA).
565
566 All of the above novel plasmids have been deposited with Addgene, with the accession numbers
567 given above, and can be purchased from there except for pCAG-hypBase, the distribution of
568 which is not permitted due to intellectual property constraints.
569
570 **Production of lentiviral and adeno-associated viral vectors**
571 Lentiviral vectors were made by transfection of HEK-293T/17 cells (ATCC 11268)
572 as described (80) but with the use of the vesicular stomatitis virus envelope expression plasmid
573 pMD2.G (Addgene 12259) for all vectors except for LV-U-TVA950(B19G), which was made using
574 the rabies virus envelope expression plasmid pCAG-B19GVSVGCD (80). Lentiviral vectors
575 expressing fluorophores were titered as described (81); titers of LV-U-TVA950(VSVG) and LV-U-
576 TVA950(B19G) were assumed to be approximately the same as those of the fluorophore-
577 expressing lentiviral vectors produced in parallel.
578
579 AAV1-synP-F14F15S-splitTVA-EGFP-tTA was packaged as serotype 1 by the UNC vector core
580 (and can be purchased from there as well as from Addgene (catalog # 136917)).
581
582 AAV1-TREtight-mTagBFP2-B19G (which we have described previously(46, 47)), was packaged
583 as serotype 1 by Addgene (catalog # 100798-AAV1).
584
585 AAV1-TREtight-H2b-emiRFP670-P2A-TEVP and AAV2-retro-synP-FLPo were made by
586 transfecting HEK 293T/17 cells with the respective genome plasmid along with pHelper
587 (Cellbiolabs VPK-421) and either pAAV-RC1 (Cellbiolabs VPK-421) or "rAAV2-retro helper"
588 (Addgene 81070)(11), using Xfect Transfection Reagent (Takara 631318) according to the
589 manufacturer's protocol, with collection of supernatant (and replacement with fresh media) at 3
590 days after transfection and collection of supernatant as well as the transfected cells at 5 days
591 after transfection. Virus was pelleted from supernatants using PEG 8000; cells were lysed by four
592 freeze-thaw cycles. Pelleted virus and cell lysate were pooled and treated with benzonase,
593 purified on an iodixanol gradient, then concentrated in an Amicon Ultra-15 centrifugal filter unit
594 (Millipore Sigma UFC9100).
595
596 **Production of titering cell lines**
597 To make reporter cell lines, HEK-293T/17 cells were infected with either pLV-CAG-FLEX-BFP-
598 (mCherry)' or pLV-CAG-F14F15S-BFP-(mCherry)' at a multiplicity of infection of 100 in one 24-
599 well plate well each. Cells were expanded to 2x 15cm plates each, then sorted on a FACS Aria
600 to retain the top 10% most brightly blue fluorescent cells. After sorting, cells were expanded again
601 to produce the cell lines 293T-FLEX-BC and 293T-F14F15S-BC, reporter lines for Cre and FLPo
602 activity, respectively. TVA-expressing versions of these two cell lines were made by infecting one
603 24-well plate well each with LV-U-TVA950(VSVG) at an MOI of approximately 100; these cells
604 were expanded to produce the cell lines 293T-FLEX-BC-TVA and 293T-F14F15S-BC-TVA.
605
606 **Production of TEVP-expressing cell line**
607 The 293T-TEVP cell line was made by transfecting HEK 293T/17 (ATCC CRL-11268) cells with
608 pCAG-hypBase and pB-CAG-TEVP-IRES-mCherry in a 15 cm plate (293T/17) or one well each
609 of a 24 well plate (BHK-B19G2 and BHK-EnvA2) using Lipofectamine 2000 (Thermo Fisher
610 11668019) according to the manufacturer's instructions, then expanding and sorting the cells on

611 a FACSAria (BD Biosciences) for the brightest 20% of red fluorescent cells, then expanding and
612 freezing the sorted cells.
613
614 **Production and titering of rabies viruses**
615 RV∆G-4Cre and RV∆GL-4Cre were produced as described (21, 34), with EnvA-enveloped viruses
616 made by using cells expressing EnvA instead of G for the last passage. Titering and infection of
617 cell lines with serial dilutions of viruses was as described (32), with the 293T-TVA-FLEX-BC and
618 293T-TVA-F14F15S-BC lines used for B19G-enveloped viruses and the 293T-TVA-FLEX-BC and
619 293T-TVA-F14F15S-BC used for the EnvA-enveloped viruses. For the *in vivo* injections, the three
620 EnvA-enveloped, Cre-encoding viruses were titered side by side, and the two higher-titer viruses
621 were diluted so that the final titer of the injected stocks of all three viruses were approximately
622 equal at 1.39E9 infectious units per milliliter.
623 The two new rabies viruses RV∆G-NPEST-4Cre and RV∆G-N*PEST-4Cre were produced
624 as described(21, 34) but with only one amplification passage (P1) between the rescue transfection
625 step and the final passage on EnvA-expressing cells. For the NPEST version, the following
626 additional modifications were made: for the rescue transfection, the new cell line 293T-TEVP (see
627 above) was used instead of HEK 293T-17, and pB-CAG-TEVP-IRES-mCherry (10 µg per 15 cm
628 plate) was included in the plasmid mix. Because pilot testing showed no clear advantage of the
629 293T-TEVP line over transfected 293Ts for passaging, the P1 passage was on HEK 293T/17
630 cells transfected with equal amounts of pCAG-B19G and pB-CAG-TEVP-IRES-mCherry (32 µg
631 of each plasmid per 15 cm plate). The final passage was on BHK-EnvA2 cells(32) transfected
632 with pB-CAG-TEVP-IRES-mCherry. Lipofectamine 2000 (Thermo Fisher 11668019) was used for
633 all transfections, following the manufacturer's protocol.
634 The final stocks of RV∆G-NPEST-4Cre(EnvA) and RV∆G-N*PEST-Cre(EnvA) were
635 titered side by side as described(32) series on 293T-FLEX-BC-TVA cells (see above) in two-fold
636 dilution series. Final titers of the viruses were determined to be 1.86e9 iu/ml for RV∆G-NPEST-
637 4Cre(EnvA) and 2.41e10 iu/mL (normalized) for RV∆G-NEPEST-4Cre(EnvA). Prior to injection *in*
638 *vivo* (see below), the N*PEST virus was diluted 12.96-fold in DPBS to match the titer of the
639 NPEST version.
640
641 **Immunostaining and imaging of cultured cells**
642 Reporter cells (see above) plated on coverslips coated in poly-L-lysine (Sigma) were infected with
643 serial dilutions of RVΔG-4Cre and RVΔGL-4Cre as described (32). Three days after infection,
644 cells were fixed with 2% paraformaldehyde, washed repeatedly with blocking/permeabilization
645 buffer (0.1% Triton-X (Sigma) and 1% bovine serum albumin (Sigma) in PBS), then labeled with
646 a blend of three FITC-conjugated anti-nucleoprotein monoclonal antibodies (Light Diagnostics
647 Rabies DFA Reagent, EMD Millipore 5100) diluted 1:100 in blocking buffer for 30 minutes,
648 followed by further washes in blocking buffer, then finally briefly rinsed with distilled water and air-
649 dried before mounting the coverslips onto microscope slides with Prolong Diamond Antifade
650 (Thermo P36970) mounting medium. Images of wells at comparable multiplicities of infection
651 (~0.1) were collected on a Zeiss 710 confocal microscope.
652
653 **Extraction of viral genomic RNA and preparation for Sanger sequencing**
654 RNA viral genomes were extracted from two Tripodi lab (EnvA/ SiR-CRE and EnvA/Sir-FLPo)
655 and one Wickersham lab (RV∆G-4mCherry) rabies virus samples using a Nucleospin RNA kit
656 (Macherey-Nagel, Germany) and treated with DNase I (37°C for 1 hour, followed by 70°C for 5
657 minutes). Extracted RNA genomes were converted to complementary DNA using an AccuScript
658 PfuUltra II RT-PCR kit (Agilent Technologies, USA) at 42°C for 2 hours with the following
659 barcoded (so that individual viral particles' genomes would be marked with distinct barcodes)
660 primer annealing to the rabies virus leader sequence:
661 Adapter_N8_leader_fp: TCAGACGATGCGTCATGCNNNNNNNNACGCTTAACAACCAGATC

662
663 cDNA sequences from the leader through the first half of the rabies virus P gene were amplified
664 using Platinum SuperFi Green Master Mix (Invitrogen (Thermo Fisher), USA) with cycling
665 conditions as follows: denaturation at 98°C for 30 seconds, followed by 25 cycles of amplification
666 (denaturation at 98°C for 5 seconds and extension at 72°C for 75 seconds), with a final extension
667 at 72°C for 5 minutes, using the following primers:
668 pEX_adapter_fp: CAGCTCAGACGATGCGTCATGC
669 Barcode2_P_rp: GCAGAGTCATGTATAGCTTCTTGAGCTCTCGGCCAG
670
671 The ~2kb PCR amplicons were extracted from an agarose gel, purified with Nucleospin Gel and
672 PCR Clean-up (Macherey-Nagel, Germany), and cloned into pEX-A (Eurofins Genomics, USA)
673 using an In-Fusion HD Cloning Kit (Takara Bio, Japan). The cloned plasmids were transformed
674 into Stellar competent cells (Takara Bio, Japan), and 200 clones per rabies virus sample were
675 isolated and purified for sequencing. For each clone, the index and the 3' end of the N gene were
676 sequenced until sequencing data was collected for over fifty clones per sample: 51 clones from
677 SiR-CRE(EnvA), 50 from SiR-FLPo(EnvA), and 51 from RVΔG-4mCherry(EnvA). Although viral
678 samples may contain plasmid DNA, viral mRNA, and positive-sense anti-genomic RNA, this RT-
679 PCR procedure can amplify only the negative-sense RNA genome: the reverse transcription
680 primer is designed to anneal to the leader sequence of the negative-strand genome so that cDNA
681 synthesis can start from the negative-sense RNA genome, with no other possible templates.
682 Additionally, the PCR amplifies the cDNA, not any plasmids which were transfected into producer
683 cell lines during viral vector production, because the forward PCR primer anneals to the primer
684 used in the reverse transcription, rather than any viral sequence. This RT-PCR protocol ensures
685 that only negative-sense RNA rabies viral genomes can be sequenced.
686
687 **Sanger sequencing of transgenes in SiR viruses**
688 The procedure for sequencing the transgene inserts was the same as above, but with the RT
689 primer being Adaptor_N8_M_fp (see below), annealing to the M gene and again with a random
690 8-nucleotide index to tag each clone, and with PCR primers pEX_adaptor_fp (see above) and
691 Barcode2_L_rp (see below), to amplify the sequences from the 3' end of the M gene to the 5' end
692 of the L gene, covering the iCre-P2A-mCherryPEST (or FLPo-P2A-mCherryPEST) sequence.
693
694 Primers for RT and PCR for Sanger sequencing were as follows:
695 Adaptor_N8_M_fp:
696 TCAGACGATGCGTCATGCNNNNNNNNCAACTCCAACCCTTGGGAGCA
697 Barcode2_L_rp:
698 GCAGAGTCATGTATAGTTGGGGACAATGGGGGTTCC
699
700 **Sanger sequencing analysis of PEST region in SiR and control viruses**
701 Alignment and mutation detection were performed using SnapGene 4.1.9 (GSL Biotech LLC,
702 USA). Reference sequences of the viral samples used in this study were based on deposited
703 plasmids in Addgene: pSAD-F3-NPEST-iCRE-2A-mCherryPEST (Addgene #99608), pSAD-F3-
704 NPEST-FLPo-2A-mCherryPEST (Addgene #99609), and pRVΔG-4mCherry (Addgene #52488).
705 Traces corresponding to indices and mutations listed in Figure 1 and Supplementary File S2 were
706 also manually inspected and confirmed.
707
708 **Single-molecule, real-time (SMRT) sequencing**
709 Double-stranded DNA samples for SMRT sequencing were prepared similarly to the above,
710 except that that the clones generated from each of the three virus samples were tagged with one
711 of the standard PacBio barcode sequences to allow identification of each clone's sample of origin
712 following multiplex sequencing (see https://www.pacb.com/wp-content/uploads/multiplex-target-

enrichment-barcoded-multi-kilobase-fragments-probe-based-capture-technologies.pdf and https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/Barcoding-with-SMRT-Analysis-2.3). This was in addition to the random index (10 nucleotides in this case) that was again included in the RT primers in order to uniquely tag each individual genome.

RNA viral genomes were extracted from two Tripodi lab (SiR-CRE and Sir-FLPo) and one Wickersham lab (RV∆G-4Cre (21); see Addgene #98034 for reference sequence) virus samples using a Nucleospin RNA kit (Macherey-Nagel, Germany) and treated with DNase I (37°C for 1 hour, followed by 70° for 5 minutes). Primers for RT and PCR are listed below. PCR cycling conditions were as follows: denaturation at 98°C for 30 seconds, followed by 20 cycles of amplification (denaturation at 98°C for 5 seconds and extension at 72°C for 75 seconds), with a final extension at 72°C for 5 minutes. This left each amplicon with a 16bp barcode at each of its ends that indicated which virus sample it was derived from, in addition to a 10-nt index sequence that was unique to each genome molecule.

Primers for RT and PCR for SMRT sequencing were as follows:

RV∆G-4Cre:

RT:

Barcode1_cagc_N10_leader_fp:
TCAGACGATGCGTCATCAGCNNNNNNNNNNACGCTTAACAACCAGATC

PCR:

Barcode1_cagc_fp: TCAGACGATGCGTCAT-CAGC
Barcode2_P_rp (see above)

SiR-CRE:

RT:

Barcode5_cagc_N10_leader_fp:
ACACGCATGACACACTCAGCNNNNNNNNNNACGCTTAACAACCAGATC

PCR:

Barcode5_cagc_fp: ACACGCATGACACACT-CAGC
Barcode3_P_rp: GAGTGCTACTCTAGTACTTCTTGAGCTCTCGGCCAG

SiR-FLPo:

RT:

Barcode9_cagc_N10_leader_fp:
CTGCGTGCTCTACGACCAGCNNNNNNNNNNACGCTTAACAACCAGATC

PCR:

Barcode9_cagc_fp: CTGCGTGCTCTACGAC-CAGC
Barcode4_P_rp: CATGTACTGATACACACTTCTTGAGCTCTCGGCCAG

After the amplicons were extracted and purified from an agarose gel, the three were mixed together at 1:1:1 molar ratio. The amplicons' sizes were confirmed on the Fragment Analyzer (Agilent Technologies, USA), then hairpin loops were ligated to both ends of the mixed amplicons to make circular SMRTbell templates for Pacbio Sequel sequencing. SMRTbell library preparation used the PacBio Template Preparation Kit v1.0 and Sequel Chemistry v3. Samples were then sequenced on a PacBio Sequel system running Sequel System v.6.0 (Pacific Biosciences, USA), with a 10-hour movie time.

Bioinformatics for PacBio sequence analysis

For the ~2kb template, the DNA polymerase with a strand displacement function can circle around the template and hairpins multiple times; the consensus sequence of multiple passes yields a

764 CCS (circular consensus sequence) read for each molecule. Raw sequences were initially
765 processed using SMRT Link v.6.0 (Pacific Biosciences, USA). Sequences were filtered for a
766 minimum of read length 10 bp, pass 3, and read score 65. 127,178 CCS reads were filtered
767 through passes 3 and Q10; 89,188 CCS reads through passes 5 and Q20; 29,924 CCS reads
768 through passes 8 and Q30. Downstream bioinformatics analysis was performed using BLASR
769 V5.3.2 for the alignment, bcftools v.1.6 for variant calling. Mutations listed in Figure 2 and
770 Supplementary File S3 were also manually inspected and confirmed using Integrative Genomics
771 Viewer 2.3.32 (software.broadinstitute.org/software/igv/). Analysis steps included the following: 1.
772 Exclude CCS reads under 1000 bases, which may have been derived from non-specific reverse
773 transcription or PCR reactions. 2. Classify the CCS reads to the three samples, according to the
774 PacBio barcodes on the 5' ends. 3. For any CCS reads that contain the same 10-nucleotide
775 random index, select only one of them, to avoid double-counting of clones derived from the same
776 cDNA molecule. 4. Align the reads to the corresponding reference sequence (see Supplementary
777 Files S4-S6). 5. Count the number of mutations at each nucleotide position of the reference
778 sequences.
779
780 **Surgeries and virus injections for two-photon imaging**
781 All experimental procedures using mice were conducted according to NIH guidelines and were
782 approved by the MIT Committee for Animal Care (CAC). Mice were housed 1-4 per cage under a
783 normal light/dark cycle for all experiments.
784
785 Adult (>9 weeks, male and female) Cre-dependent tdTomato reporter Ai14 (82) (Jackson
786 Laboratory #007908) or Arch-EGFP reporter Ai35D (40) (Jackson Laboratory # 012735) mice
787 were anesthetized with isoflurane (4% in oxygen) and ketamine/xylazine (100mg/kg and 10mg/kg
788 respectively, i.p.). Mice were given buprenorphine (0.1 mg/kg s.q.) and meloxicam (2 mg/kg s.q.)
789 as preemptive analgesics, as well as eye ointment (Puralube); the scalp was then shaved,
790 depilated with Nair, and thoroughly rinsed before the mice were mounted on a stereotaxic
791 instrument (Stoelting Co.) with a hand warmer (Heat Factory) underneath the animal to maintain
792 body temperature. The scalp was then disinfected with povidone-iodine, and an incision was
793 made at the appropriate location (see below).
794
795 A 3 mm craniotomy was opened over primary visual cortex (V1). 300 nl of LV-U-TVA950(B19G)
796 (see above) was injected into V1 (-2.70 mm AP, 2.50 mm LM, -0.26 mm DV; AP and LM
797 stereotaxic coordinates are with respect to bregma; DV coordinate is with respect to brain surface)
798 using a custom injection apparatus comprised of a hydraulic manipulator (MO-10, Narishige) with
799 headstage coupled via custom adaptors to a wire plunger advanced through pulled glass
800 capillaries (Wiretrol II, Drummond) back-filled with mineral oil and front-filled with virus solution.
801 Glass windows composed of a 3mm-diameter glass coverslip (Warner Instruments CS-3R) glued
802 (Optical Adhesive 61, Norland Products) to a 5mm-diameter glass coverslip (Warner Instruments
803 CS-5R) were then affixed over the craniotomy with Metabond (Parkell). Seven days after injection
804 of the lentiviral vector, the coverslips were removed and 300 nl of one of the three EnvA-
805 enveloped rabies viral vectors (with equalized titers as described above) was injected at the same
806 stereotaxic coordinates. Coverslips were reapplied and custom stainless steel headplates
807 (eMachineShop) were affixed to the skulls around the windows.
808
809 ***In vivo* two-photon imaging and image analysis**
810 Beginning seven days after injection of each rabies virus and continuing every seven days up to
811 a maximum of four weeks following rabies virus injection, the injection sites were imaged on a
812 Prairie/Bruker Ultima IV In Vivo two-photon microscope driven by a Spectra Physics Mai-Tai Deep
813 See laser with a mode locked Ti:sapphire laser emitting at a wavelength of 1020 nm for tdTomato
814 and mCherry or 920 nm for EGFP. Mice were reanesthetized and mounted via their headplates

815 to a custom frame, again with ointment applied to protect their eyes and with a handwarmer
816 maintaining body temperature. One field of view was chosen in each mouse in the area of maximal
817 fluorescent labelling. The imaging parameters were as follows: image size 512 X 512 pixels (282.6
818 µm x 282.6 µm), 0.782 Hz frame rate, dwell time 4.0 µs, 2x optical zoom, Z-stack step size 1 µm.
819 Image acquisition was controlled with Prairie View 5.4 software. Laser power exiting the 20x
820 water-immersion objective (Zeiss, W plan-apochromat, NA 1.0) varied between 20 and 65 mW
821 depending on focal plane depth (Pockel cell value was automatically increased from 450 at the
822 top section of each stack to 750 at the bottom section). For the example images of labeled cells,
823 maximum intensity projections (stacks of 150-400 µm) were made with Fiji software. Cell counting
824 was performed with the ImageJ Cell Counter plugin. When doing cell counting, week 1 tdTomato
825 labelled cells were defined as a reference; remaining week 1 cells were the same cells at later
826 time point that align with week 1 reference cells but the not-visible cells at week 1 (the dead cells).
827 Plots of cell counts were made with Origin 7.0 software (OriginLab, Northampton, MA). For the
828 thresholded version of this analysis (Supplementary Figure S2), in order to exclude cells that
829 could possibly have been labeled only with mCherry in the SiR-CRE group, only cells with
830 fluorescence intensity greater than the average of the mean red fluorescence intensities of cells
831 imaged in Ai35 versus Ai14 mice at the same laser power at 1020 nm at 7 days postinjection
832 (32.33 a.u.) were included in the population of cells tracked from 7 days onward.
833
834 **Monosynaptic tracing experiments: Sanger sequencing of viral genomic RNA**
835 Sanger sequencing of RVΔG-NPEST-4Cre(EnvA) and RVΔG-N*PEST-4Cre(EnvA) was as
836 described above but with the following modifications. The following barcoded primer was used for
837 RT-PCR.
838 Adaptor-UMI-N_fp_57:
839 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNAGAAGTC
840 CGGAGGCTGTTTAT
841 cDNA sequences from the nucleoprotein gene through the first half of the rabies virus P gene
842 were amplified using Platinum SuperFi Green Master Mix (Invitrogen (Thermo Fisher), USA) with
843 cycling conditions as follows: denaturation at 98°C for 30 seconds, followed by 21 cycles of
844 amplification (denaturation at 98°C for 5 seconds, annealing at 60°C for 10 seconds and extension
845 at 72°C for 21 seconds), with a final extension at 72°C for 5 minutes, using the following primers:
846 i5-anchor_CTAGCGCT_fp_56:
847 AATGATACGGCGACCACCGAGATCTACACCTAGCGCTACACTCTTTCCCTACACGAC
848 P_Sanger_rp_56:
849 CAAGCAGAAGACGGCACGATTTTCCATCATCCAGGTG
850
851 The 700bp PCR amplicons for RVdG-NPEST-4Cre (EnvA) or RVΔG-N*PEST-4Cre (EnvA) virus
852 were cloned into pEX-A (Eurofins Genomics, USA) using an In-Fusion HD Cloning Kit (Takara
853 Bio, Japan) as above. The plasmids were transformed into Stellar competent cells (Takara Bio,
854 Japan), and 32 clones from each rabies virus sample were isolated and purified for sequencing.
855
856 **Monosynaptic tracing experiments: surgeries and virus injections**
857 Following preparation of mice as described above ("Surgeries and virus injections for two-photon
858 imaging") the three helper AAV1s were combined at final titers of
859 3.6E10 gc/ml for AAV1-synP-F14F15S-splitTVA-EGFP-tTA and 6.60E11 gc/ml for AAV1-
860 TREtight-mTagBFP2-B19G (when included) and AAV1-TREtight-H2b-emiRFP670-P2A-TEVP
861 (when included) in DPBS (Fisher, 14-190-250). 250 nl of helper virus mixture was injected into
862 layer 5 of barrel cortex (AP -1.55 mm w.r.t. bregma, LM 3.00 mm w.r.t. bregma, -DV 0.75 mm
863 w.r.t. brain surface) of Ai14 mice; in the same surgery, 300 nl of AAV2-retro-synP-FLPo
864 (1.16E13 gc/ml) was injected into dorsolateral striatum (AP 0.74 mm w.r.t. bregma, LM 2.25 mm
865 w.r.t. bregma, DV -2.30 mm w.r.t the brain surface). 7 days after AAV injection, 300nl of RVΔG-

NPEST-4Cre(EnvA) (1.86E9 iu/ml) or RVΔG-N*PEST-4Cre(EnvA) (diluted 12.96-fold in DPBS from 2.41E10 iu/ml to 1.86E9 iu/ml) was injected in barrel cortex at the same site as the helper AAV mixtures.

Monosynaptic tracing experiments: perfusions and histology

12 days or 3 weeks (depending on experiment; see main text) after injection of rabies virus, mice were transcardially perfused with 4% paraformaldehyde in phosphate-buffered saline. Brains were postfixed overnight in 4% paraformaldehyde in PBS on a shaker at 4°C and cut into 50 μm coronal on a vibrating microtome (Leica, VT-1000S). Sections were collected anterior to posteriorly into 6 tubes containing cryoprotectant. Collection goes on for 15 rounds so that each tube contains a sixth of the collected tissue (15 sections in each tube). Sections were immunostained as described(47) with the following primary antibodies (as applicable) at the following respective dilutions: chicken anti-GFP (Aves Labs GFP-1020) 1:500, with secondary antibodies donkey anti-chicken Alexa Fluor 488 (Jackson Immuno 703-545-155) 1:200. Sections were mounted with Prolong Diamond Antifade mounting medium (Thermo Fisher P36970).

Monosynaptic tracing experiments: microscopy and cell counts

Coronal sections between 1.2mm and -3.3mm relative to bregma were examined under an epifluorescence microscope (Zeiss, Imager.Z2). When necessary due to high density of labeled cells, images were taken with the same microscope for cell counting. tdTomato-labeled neurons in contralateral cortex and thalamus were counted manually with the Cell Counter plugin in ImageJ. For cells at the injection site, when tdTomato expressing cells were few and sparse (usually less than 100 per section), cells coexpressing mTagBFP2 or H2b-emiRFP670 alongside were counted manually adding separate labels to each and then looking for overlapping cells. When tdTomato expressing cells were dense, tdTomato labeled cells were first counted using the Analyze Particle function in ImageJ (size in micron^2: 20-400; circularity: 0.20-1.00). The outline of these cells was then merged on top of images of mTagBFP2 and H2b-emiRFP670 labeled cells for the counting of the overlapping cells. Only one of the six series of sections (i.e., every sixth section: see above) was counted for each mouse.

Images for figures were taken on a confocal microscope (Zeiss, LSM 900). So that the confocal images of brain tissue included in the figures in this paper are representative of each group, the images were taken after the counts were conducted (see above), in each case using the mouse with the middle number of labeled neurons in that group (i.e., neither the highest nor the lowest in the group of 3 mice used for each condition).

ACKNOWLEDGEMENTS

We thank Ernesto Ciabatti and Marco Tripodi for sharing samples of EnvA/SiR-CRE and EnvA/SiR-FLPo and for comments on the manuscript. We thank Ed Callaway, Sean Whelan, Ayano Matsushima, and Kim Ritola for helpful discussion and Jun Zhuang, Soumya Chatterjee, and Ali Cetin for helpful discussion and sharing their own results with SiR viruses. We thank Stuart Levine, Noelani Kamelamela, and Huiming Ding of the MIT BioMicro Center for assistance with SMRT sequencing and bioinformatic data analysis and Sara Beach for helpful feedback on the manuscript. Research reported in this publication was supported by the following BRAIN Initiative awards from the National Institute of Mental Health: U01MH106018 (Wickersham) RF1MH120017 (Wickersham), U01MH114829 (Dong), and U19MH114830 (Zeng).

REFERENCES

1. Ugolini G, Kuypers HG, & Strick PL (1989) Transneuronal transfer of herpes virus from peripheral nerves to cortex and brainstem. *Science* 243(4887):89-91.
2. Ugolini G (1995) Specificity of rabies virus as a transneuronal tracer of motor networks: transfer from hypoglossal motoneurons to connected second-order and higher order central nervous system cell groups. *J Comp Neurol* 356(3):457-480.
3. Kelly RM & Strick PL (2000) Rabies as a transneuronal tracer of circuits in the central nervous system. *J Neurosci Methods* 103(1):63-71.
4. Wickersham IR, Finke S, Conzelmann KK, & Callaway EM (2007) Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* 4(1):47-49.
5. Wickersham IR*, et al.* (2007) Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* 53(5):639-647.
6. Atasoy D, Aponte Y, Su HH, & Sternson SM (2008) A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28(28):7025-7030.
7. Beier KT*, et al.* (2011) Anterograde or retrograde transsynaptic labeling of CNS neurons with vesicular stomatitis virus vectors. *P Natl Acad Sci USA*.
8. Lo L & Anderson DJ (2011) A Cre-dependent, anterograde transsynaptic viral tracer for mapping output pathways of genetically marked neurons. *Neuron* 72(6):938-950.
9. Schwarz LA*, et al.* (2015) Viral-genetic tracing of the input-output organization of a central noradrenaline circuit. *Nature*.
10. Deverman BE*, et al.* (2016) Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat Biotechnol* 34(2):204-209.
11. Tervo DG*, et al.* (2016) A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron* 92(2):372-382.
12. Dimidschstein J*, et al.* (2016) A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nat Neurosci* 19(12):1743-1749.
13. Zeng WB*, et al.* (2017) Anterograde monosynaptic transneuronal tracers derived from herpes simplex virus 1 strain H129. *Mol Neurodegener* 12(1):38.
14. Luo L, Callaway EM, & Svoboda K (2018) Genetic Dissection of Neural Circuits: A Decade of Progress. *Neuron* 98(4):865.
15. Ravindra Kumar S*, et al.* (2020) Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. *Nat Methods* 17(5):541-550.
16. Mich JK*, et al.* (2021) Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep* 34(13):108754.
17. Augustine V*, et al.* (2018) Hierarchical neural architecture underlying thirst regulation. *Nature* 555(7695):204-209.
18. Kohl J*, et al.* (2018) Functional circuit architecture underlying parental behaviour. *Nature* 556(7701):326-331.
19. Evans DA*, et al.* (2018) A synaptic threshold mechanism for computing escape decisions. *Nature* 558(7711):590-594.
20. Kaelberer MM*, et al.* (2018) A gut-brain neural circuit for nutrient sensory transduction. *Science* 361(6408).
21. Chatterjee S*, et al.* (2018) Nontoxic, double-deletion-mutant rabies viral vectors for retrograde targeting of projection neurons. *Nat Neurosci* 21(4):638-646.
22. Reardon TR*, et al.* (2016) Rabies Virus CVS-N2c(DeltaG) Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron* 89(4):711-724.

964 23. Morimoto K, Hooper DC, Spitsin S, Koprowski H, & Dietzschold B (1999) Pathogenicity
965 of different rabies virus variants inversely correlates with apoptosis and rabies virus
966 glycoprotein expression in infected primary neuron cultures. *J Virol* 73(1):510-518.
967 24. Ciabatti E, Gonzalez-Rueda A, Mariotti L, Morgese F, & Tripodi M (2017) Life-Long
968 Genetic and Functional Access to Neural Circuits Using Self-Inactivating Rabies Virus.
969 *Cell* 170(2):382-392 e314.
970 25. Rogers S, Wells R, & Rechsteiner M (1986) Amino acid sequences common to rapidly
971 degraded proteins: the PEST hypothesis. *Science* 234(4774):364-368.
972 26. Steinhauer DA & Holland JJ (1987) Rapid evolution of RNA viruses. *Annu Rev Microbiol*
973 41:409-433.
974 27. Steinhauer DA, de la Torre JC, & Holland JJ (1989) High nucleotide substitution error
975 frequencies in clonal pools of vesicular stomatitis virus. *J Virol* 63(5):2063-2071.
976 28. Holland JJ, De La Torre JC, & Steinhauer DA (1992) RNA virus populations as
977 quasispecies. *Curr Top Microbiol Immunol* 176:1-20.
978 29. Holmes EC, Woelk CH, Kassis R, & Bourhy H (2002) Genetic constraints and the
979 adaptive evolution of rabies virus in nature. *Virology* 292(2):247-257.
980 30. Jenkins GM, Rambaut A, Pybus OG, & Holmes EC (2002) Rates of molecular evolution
981 in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54(2):156-165.
982 31. Combe M & Sanjuan R (2014) Variation in RNA virus mutation rates across host cells.
983 *PLoS Pathog* 10(1):e1003855.
984 32. Wickersham IR, Sullivan HA, & Seung HS (2010) Production of glycoprotein-deleted
985 rabies viruses for monosynaptic tracing and high-level gene expression in neurons.
986 *Nature protocols* 5(3):595-606.
987 33. Osakada F & Callaway EM (2013) Design and generation of recombinant rabies virus
988 vectors. *Nat Protoc* 8(8):1583-1601.
989 34. Wickersham IR & Sullivan HA (2015) Rabies viral vectors for monosynaptic tracing and
990 targeted transgene expression in neurons. *Cold Spring Harb Protoc* 2015(4):375-385.
991 35. Weible AP*, et al.* (2010) Transgenic targeting of recombinant rabies virus reveals
992 monosynaptic connectivity of specific neurons. *The Journal of neuroscience : the official*
993 *journal of the Society for Neuroscience* 30(49):16509-16513.
994 36. Carneiro MO*, et al.* (2012) Pacific biosciences sequencing technology for genotyping
995 and variation discovery in human data. *BMC Genomics* 13:375.
996 37. Wirblich C & Schnell MJ (2011) Rabies virus (RV) glycoprotein expression levels are not
997 critical for pathogenicity of RV. *J Virol* 85(2):697-704.
998 38. Tao L*, et al.* (2010) Molecular basis of neurovirulence of flury rabies virus vaccine
999 strains: importance of the polymerase and the glycoprotein R333Q mutation. *J Virol*
1000 84(17):8926-8936.
1001 39. Prehaud C, Lay S, Dietzschold B, & Lafon M (2003) Glycoprotein of nonpathogenic
1002 rabies viruses is a key determinant of human cell apoptosis. *J Virol* 77(19):10537-10547.
1003 40. Madisen L*, et al.* (2012) A toolbox of Cre-dependent optogenetic transgenic mice for
1004 light-induced activation and silencing. *Nature neuroscience* 15(5):793-802.
1005 41. Shaner NC*, et al.* (2004) Improved monomeric red, orange and yellow fluorescent
1006 proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol*
1007 22(12):1567-1572.
1008 42. Drobizhev M, Makarov NS, Tillo SE, Hughes TE, & Rebane A (2011) Two-photon
1009 absorption properties of fluorescent proteins. *Nat Methods* 8(5):393-399.
1010 43. Huang KW & Sabatini BL (2020) Single-Cell Analysis of Neuroinflammatory Responses
1011 Following Intracranial Injection of G-Deleted Rabies Viruses. *Front Cell Neurosci* 14:65.
1012 44. Prosniak M, Hooper DC, Dietzschold B, & Koprowski H (2001) Effect of rabies virus
1013 infection on gene expression in mouse brain. *Proc Natl Acad Sci U S A* 98(5):2758-
1014 2763.

1015 45. Madisen L*, et al.* (2015) Transgenic mice for intersectional targeting of neural sensors
1016 and effectors with high specificity and performance. *Neuron* 85(5):942-958.
1017 46. Liu K*, et al.* (2017) Lhx6-positive GABA-releasing neurons of the zona incerta promote
1018 sleep. *Nature* 548(7669):582-587.
1019 47. Lavin TK, Jin L, & Wickersham IR (2019) Monosynaptic tracing: a step-by-step protocol.
1020 *J Chem Neuroanat*:101661.
1021 48. Lavin TK, Jin L, Lea NE, & Wickersham IR (2020) Monosynaptic Tracing Success
1022 Depends Critically on Helper Virus Concentrations. *Front Synaptic Neurosci* 12:6.
1023 49. Ciabatti E*, et al.* (2020) Genomic stability of Self-inactivating Rabies.
1024 *bioRxiv*:2020.2009.2019.304683.
1025 50. Dalton KP & Rose JK (2001) Vesicular stomatitis virus glycoprotein containing the entire
1026 green fluorescent protein on its cytoplasmic domain is incorporated efficiently into virus
1027 particles. *Virology* 279(2):414-421.
1028 51. Duprex WP, Collins FM, & Rima BK (2002) Modulating the function of the measles virus
1029 RNA-dependent RNA polymerase by insertion of green fluorescent protein into the open
1030 reading frame. *J Virol* 76(14):7322-7328.
1031 52. Finke S, Brzozka K, & Conzelmann KK (2004) Tracking fluorescence-labeled rabies
1032 virus: enhanced green fluorescent protein-tagged phosphoprotein p supports virus gene
1033 expression and formation of infectious particles. *J Virol* 78(22):12333-12343.
1034 53. Koser ML*, et al.* (2004) Rabies virus nucleoprotein as a carrier for foreign antigens. *Proc*
1035 *Natl Acad Sci U S A* 101(25):9405-9410.
1036 54. Brown DD*, et al.* (2005) Rational attenuation of a morbillivirus by modulating the activity
1037 of the RNA-dependent RNA polymerase. *J Virol* 79(22):14330-14338.
1038 55. Das SC, Nayak D, Zhou Y, & Pattnaik AK (2006) Visualization of intracellular transport of
1039 vesicular stomatitis virus nucleocapsids in living cells. *J Virol* 80(13):6368-6377.
1040 56. Klingen Y, Conzelmann KK, & Finke S (2008) Double-labeled rabies virus: live tracking
1041 of enveloped virus transport. *Journal of virology* 82(1):237-245.
1042 57. Das SC, Panda D, Nayak D, & Pattnaik AK (2009) Biarsenical labeling of vesicular
1043 stomatitis virus encoding tetracysteine-tagged m protein allows dynamic imaging of m
1044 protein and virus uncoating in infected cells. *J Virol* 83(6):2611-2622.
1045 58. Marriott AC & Hornsey CA (2011) Reverse genetics system for Chandipura virus:
1046 tagging the viral matrix protein with green fluorescent protein. *Virus Res* 160(1-2):166-
1047 172.
1048 59. Soh TK & Whelan SP (2015) Tracking the Fate of Genetically Distinct Vesicular
1049 Stomatitis Virus Matrix Proteins Highlights the Role for Late Domains in Assembly. *J*
1050 *Virol* 89(23):11750-11760.
1051 60. Nikolic J, Civas A, Lama Z, Lagaudriere-Gesbert C, & Blondel D (2016) Rabies Virus
1052 Infection Induces the Formation of Stress Granules Closely Connected to the Viral
1053 Factories. *PLoS Pathog* 12(10):e1005942.
1054 61. Case JB*, et al.* (2020) Replication-competent vesicular stomatitis virus vaccine vector
1055 protects against SARS-CoV-2-mediated pathogenesis. *bioRxiv*.
1056 62. Gomme EA, Wirblich C, Addya S, Rall GF, & Schnell MJ (2012) Immune clearance of
1057 attenuated rabies virus results in neuronal survival with altered gene expression. *PLoS*
1058 *Pathog* 8(10):e1002971.
1059 63. Wiktor TJ, Dietzschold B, Leamnson RN, & Koprowski H (1977) Induction and biological
1060 properties of defective interfering particles of rabies virus. *J Virol* 21(2):626-635.
1061 64. Kawai A & Matsumoto S (1977) Interfering and noninterfering defective particles
1062 generated by a rabies small plaque variant virus. *Virology* 76(1):60-71.
1063 65. Clark HF, Parks NF, & Wunner WH (1981) Defective interfering particles of fixed rabies
1064 viruses: lack of correlation with attenuation or auto-interference in mice. *J Gen Virol*
1065 52(Pt 2):245-258.

66. Faber M*, et al.* (2002) Overexpression of the rabies virus glycoprotein results in enhancement of apoptosis and antiviral immune response. *J Virol* 76(7):3374-3381.
67. Marr RA*, et al.* (2004) Neprilysin regulates amyloid Beta peptide levels. *J Mol Neurosci* 22(1-2):5-11.
68. Niwa H, Yamamura K, & Miyazaki J (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108(2):193-199.
69. Subach OM, Cranfill PJ, Davidson MW, & Verkhusha VV (2011) An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS One* 6(12):e28674.
70. Turan S, Kuehle J, Schambach A, Baum C, & Bode J (2010) Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J Mol Biol* 402(1):52-69.
71. Matsuda T & Cepko CL (2004) Electroporation and RNA interference in the rodent retina in vivo and in vitro. *P Natl Acad Sci USA* 101(1):16-22.
72. Bates P, Young JA, & Varmus HE (1993) A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor. *Cell* 74(6):1043-1051.
73. Raymond CS & Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* 2(1):e162.
74. Yusa K, Zhou L, Li MA, Bradley A, & Craig NL (2011) A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* 108(4):1531-1536.
75. Gray DC, Mahrus S, & Wells JA (2010) Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* 142(4):637-646.
76. Yang CF*, et al.* (2013) Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell* 153(4):896-909.
77. Kapust RB*, et al.* (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* 14(12):993-1000.
78. Gallardo HF, Tan C, & Sadelain M (1997) The internal ribosomal entry site of the encephalomyocarditis virus enables reliable coexpression of two transgenes in human primary T lymphocytes. *Gene Ther* 4(10):1115-1119.
79. Matlashov ME*, et al.* (2020) A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales. *Nat Commun* 11(1):239.
80. Wickersham IR*, et al.* (2015) Lentiviral vectors for retrograde delivery of recombinases and transactivators. *Cold Spring Harb Protoc* 2015(4):368-374.
81. Sullivan HA & Wickersham IR (2015) Concentration and purification of rabies viral and lentiviral vectors. *Cold Spring Harb Protoc* 2015(4):386-391.
82. Madisen L*, et al.* (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience* 13(1):133-140.

1107 **FIGURE LEGENDS**
1108
1109 **Figure 1: Single-molecule Sanger sequencing of individual viral genomes revealed that**
1110 **most SiR virions had lost the intended C-terminal modification to the nucleoprotein.**
1111 (A) Schematic of the RT-PCR workflow. In the reverse transcription (RT) step, the RT primer,
1112 containing a random 8-nucleotide sequence, anneals to the 3' rabies virus leader, adding a unique
1113 random index to the 5' end of the cDNA molecule corresponding to each individual viral particle's
1114 RNA genome. In the PCR step, the forward PCR primer anneals to the RT primer sequence and
1115 the reverse PCR primer anneals within the viral phosphoprotein gene P. Both PCR primers have
1116 15-base sequences homologous to those flanking an insertion site in a plasmid used for
1117 sequencing, allowing the amplicons to be cloned into the plasmid using a seamless cloning
1118 method before transformation into bacteria. The resulting plasmid library consists of plasmids
1119 containing up to 4^8 different index sequences, allowing confirmation that the sequences of
1120 plasmids purified from individual picked colonies correspond to the sequences of different
1121 individual rabies viral particles' genomes.
1122 (B) Representative Sanger sequencing data of the 8-bp index and the TEV-PEST sequence.
1123 Mutations are highlighted in red.
1124 (C) Mutation variants and their frequencies in each viral vector sample based on Sanger
1125 sequencing data. No unmutated genomes were found in the SiR-CRE sample: 50 out of 51 had
1126 a substitution creating an opal stop codon just before the TEV cleavage site, and the 51st genome
1127 contained a frameshift which also removed the C-terminal addition. In the SiR-FLPo sample, only
1128 4 out of 50 clones had an intact sequence of the C-terminal addition; the other 46 out of 50 had
1129 one of two *de novo* stop codons at the end of N or the beginning of the TEV cleavage site. In the
1130 sample of RVΔG-4mCherry, a virus from our laboratory included as a control to distinguish true
1131 mutations on the rabies genomes from mutations due to the RT-PCR process, none of the 51
1132 clones analyzed had mutations in the sequenced region.
1133
1134 **Figure 2: Single-molecule, real-time (SMRT) sequencing of thousands of barcoded viral**
1135 **genomes confirmed that most SiR virions had lost the intended C-terminal modification to**
1136 **the nucleoprotein.**
1137 (A) Schematic of workflow for SMRT sequencing. An RT primer with a random 10-nucleotide
1138 sequence anneals to the leader sequence on the negative-sense single-stranded RNA genome.
1139 Forward and reverse PCR primers have distinct SMRT barcodes at their 5' ends for the three
1140 different virus samples. After RT-PCR, each amplicon library consists of amplicons each
1141 containing a SMRT barcode to identify the sample of origin as well as the 10-nucleotide index
1142 (i.e., with a potential diversity of 4^{10} different indices) to uniquely label the individual genome of
1143 origin. SMRT "dumbbell" adaptors are then ligated to the amplicons' ends, making circular
1144 templates which are then repeatedly traversed by a DNA polymerase, resulting in long
1145 polymerase reads each containing multiple reads of the positive and negative strands. The
1146 individual subreads for a given molecule are combined to form circular consensus sequence
1147 (CCS) reads.
1148 (B) High-frequency (>2%) point mutations found in the rabies vector samples based on SMRT
1149 sequencing. Horizontal axis represents nucleotide position along the reference sequences (see
1150 text); vertical axis represents variant frequency. Total number of CCS3 reads (i.e., with at least 3
1151 subreads for each position) are 22,205 for SiR-CRE, 17,086 reads for SiR-FLPo, and 17,978
1152 reads for RVΔG-4Cre. The great majority of SiR-CRE and SiR-FLPo genomes have point
1153 mutations creating premature stop codons at or just after the end of N and before the C-terminal
1154 addition. The only frequent (>2%) mutation found in the control virus, RVΔG-4Cre, was a single
1155 amino acid substitution at position 419 in 9.49% of virions. Insertions and deletions are not shown
1156 here (see text).

(C) Summary of results. In the SiR virus samples, 99.22% of SiR-CRE virions and 83.85% of SiR-FLPo virions had point mutations creating premature stop codons that completely removed the intended C-terminal addition to the nucleoprotein, making them simply first-generation (ΔG) rabies viral vectors. This does not include any insertions or deletions causing frameshifts (see text), which would further increase the percentage of first-generation-type virions in these samples. In the RVΔG-4Cre sample, there were no premature stop codons at or near the end of the nucleoprotein gene.

Figure 3: SiR viruses appeared to cause expression of viral nucleoprotein at levels similar to those of first-generation ΔG viruses.
(A-D) Reporter cells infected with first-generation, ΔG viruses show characteristic bright, clumpy anti-nucleoprotein staining (green), indicating high nucleoprotein expression and active viral replication. Red is mCherry expression, reporting expression of Cre or FLPo; blue is mTagBFP2, constitutively expressed by these reporter cell lines.
(E-H) Reporter cells infected with second-generation, ΔGL viruses show only punctate staining for nucleoprotein, indicating isolated individual viral particles or ribonucleoprotein complexes; these viruses do not replicate intracellularly {Chatterjee, 2018 #6924}. Reporter cassette activation takes longer from the lower recombinase expression levels of these viruses, so mCherry expression is dimmer than in cells infected with ΔG viruses at the same time point.
(I-J) Reporter cells infected with SiR viruses show clumps of nucleoprotein and rapid reporter expression indicating high expression of recombinases, similarly to cells infected with ΔG viruses. Scale bar: 100 μm, applies to all panels.

Figure 4: Longitudinal two-photon imaging *in vivo* showed that SiR virus killed approximately 80% of infected neurons in vivo within 2-4 weeks.
A) Representative fields of view (FOVs) of visual cortical neurons labeled with RVΔG-4Cre (top row), RVΔGL-4Cre (middle row), or SiR-CRE (bottom row) in Ai14 mice (Cre-dependent expression of tdTomato). Images within each row are of the same FOV imaged at the four different time points in the same mouse. Circles indicate cells that are present at 7 days postinjection but no longer visible at a subsequent time point. Scale bar: 50 μm, applies to all images.
B-D) Numbers of cells present at week 1 that were still present in subsequent weeks. While very few cells labeled with RVΔGL-4Cre were lost, and RVΔG-4Cre killed a significant minority of cells, SiR-CRE killed the majority of labeled neurons within 14 days following injection.
E) Percentages of cells present at week 1 that were still present in subsequent imaging sessions. By 28 days postinjection, an average of only 20.5% of cells labeled by SiR-CRE remained.

Figure 5: 86% of SiR-CRE-labeled neurons in Arch-EGFP-ER2 reporter mice disappeared between 11 days and 28 days after injection.
A) Maximum intensity projections of the two-photon FOV shown in Supplementary Video S1 of visual cortical neurons labeled with SiR-CRE in an Ai35 mouse, 11-28 days postinjection. Images are from the same FOV at four different time points. All cells clearly visible on day 11 are circled. In this example, 18 out of 19 cells (red circles) disappeared by a subsequent imaging session. Only one cell (white circle) is still visible on day 28. Numbers below four of the circles mark the cells for which intensity profiles are shown in panel B. Scale bar: 50 μm, applies to all images.
B) Green fluorescence intensity versus depth for the four representative neurons numbered in panel A at the four different time points, showing disappearance of three of them over time.
C) Fraction of cells visibly EGFP-labeled at day 11 still visible at later time points, from four different FOVs in two Ai35 mice. Connected sets of markers indicate cells from the same FOV. 86% of SiR-CRE-labeled neurons had disappeared by 4 weeks postinjection.

1207 **Figure 6: Rabies virus with an intact PEST domain does not spread in the absence of**
1208 **TEVP.**
1209 A) Sequences of the region of the junction between the nucleoprotein gene N and its 3' addition
1210 in SiR-CRE (intended sequence) as well as in our two new viruses RVΔG-NPEST-4Cre and
1211 RVΔG-N*PEST-4Cre. Apart from synonymous mutations we made in five codons (highlighted in
1212 green) to make them less likely to mutate to stop codons, RVΔG-NPEST-4Cre has exactly the
1213 same 3' addition to N as was intended for SiR. RVΔG-N*PEST-4Cre has exactly the same
1214 sequence as RVΔG-NPEST-4Cre apart from a stop codon that we deliberately introduced at the
1215 same location as the nonsense mutation that we had found in almost every SiR-CRE virion, so
1216 that the only modification to the native nucleoprotein is an additional two amino acids (Gly-Ser)
1217 on its C-terminus. Sanger sequencing of 32 viral particles of each virus confirmed that the
1218 intended additions were still present in the final stocks.
1219 B) Diagram of virus injections for monosynaptic tracing experiments. An AAV2-retro expressing
1220 FLPo was injected in dorsolateral striatum, and a FLP-dependent helper virus combination was
1221 injected into barrel cortex to express TVA and G in corticostriatal cells. Rabies virus was
1222 injected seven days later at the same location in barrel cortex, and mice were perfused either 12
1223 days or 21 days after rabies virus injection.
1224 C) At 12 days after injection, the NPEST virus (top row), with an intact PEST domain, shows no
1225 evidence of spread, whereas the N*PEST virus (bottom row), without the PEST domain, has
1226 labeled thousands of neurons in ipsilateral cortex (shown here) and also spread to contralateral
1227 cortex and thalamus (see Supplementary Figure S4 for images). For all panels in this figure as
1228 well as Figure 7 and the related supplementary figures, the red channel shows tdTomato,
1229 reporting Cre expression from the rabies viruses; the green channel shows EGFP, coexpressed
1230 with TVA (and tTA) by the first helper virus; the blue channel shows mTagBFP2, coexpressed
1231 with G by the second helper virus. Scale bar: 200 μm, applies to all images.
1232 D) At 21 days after injection, the N*PEST virus again shows no evidence of spread, while the
1233 N*PEST virus has labeled many thousands of neurons. Scale bar: 200 μm, applies to all
1234 images.
1235 E) Counts of labeled neurons in ipsilateral cortex (left), contralateral cortex (center), and
1236 thalamus (right) at 12 days and 21 days for the two viruses. Each dot indicates the total number
1237 of labeled cells found in a given mouse brain when examining every sixth 50 μm section (see
1238 Methods). All differences in numbers of cells labeled by the two viruses are highly significant for
1239 all conditions (see Supplementary File S8 for all counts and statistical analyses), except for the
1240 numbers of contralateral cells at 12 days.
1241
1242 **Figure 7: Rabies virus with an intact PEST domain does spread when TEVP is supplied,**
1243 **although with lower efficiency than virus without a PEST domain.**
1244 With the experimental design exactly the same as shown in Figure 6, Panel B but with a TEVP-
1245 expressing AAV included in the helper virus mixture, the NPEST virus did spread
1246 transsynaptically, but to a more limited degree than the N*PEST one.
1247 A, B) At both 12 (A) and 21 (B) days after injection, both NPEST (top row) and N*PEST (bottom
1248 row) viruses have labeled many cells in ipsilateral cortex (shown here), although many more
1249 cells are labeled by the NPEST virus. Both viruses have also spread to contralateral cortex and
1250 thalamus (see Supplementary Figure S5 for images).
1251 C - D) Counts of tdTomato+ cells (C) and ratios (D) of number of tdTomato+ cells to number of
1252 "starter cells", defined here as cells coexpressing tdTomato, mTagBFP2, and emiRFP670, with
1253 TEVP supplied for both viruses (in cases in which TEVP is not supplied (see below), starter
1254 cells are defined as cells coexpressing tdTomato and mTagBFP2). Both the absolute numbers
1255 and the ratios are higher (in all cases except for contralateral cortex at 12 days) for the N*PEST
1256 virus than for the NPEST one, although the differences were not significant with the small size
1257 of each group (see text). Note that the inclusion of TEVP reduced the spread of the N*PEST

virus considerably (cf. Figure 6; all counts and statistical comparisons are in Supplementary File S8).

E - F) Comparison of the best condition for the NPEST virus (with TEVP; same data as in panel C and D above) to the best condition for the N*PEST one (without TEVP; same data as in Figure 6) gives highly significant differences for most comparisons, with the advantage of the N*PEST virus over the NPEST version increasing between 12 days and 3 weeks.

LEGENDS FOR SUPPLEMENTARY INFORMATION

Supplementary File S1: >90% of SiR-CRE viral particles with the mCherry gene intact also have the Cre gene intact, suggesting that most of the SiR-CRE-infected cells that disappear over time in tdTomato reporter mice are dying rather than simply stopping expression of mCherry.

We sequenced the transgene inserts for 21 individual SiR-CRE clones (see Methods). 19 out of 21 had no mutations in the Cre gene, and two had one point mutation each (Ala88Val and Arg189Ile). All 21 had an intact mCherry gene. The lack of a large proportion of Cre-knockout mutants is one indication that the majority of red fluorescent neurons in SiR-CRE-injected Ai14 (tdTomato reporter) mice are not labeled only with mCherry, providing evidence that their disappearance is equivalent to their death. Parenthetically, we also partially sequenced the transgene insert for four individual clones of the SiR-FLPo sample, just to confirm the identity of the virus. In one of the four sequenced virions, we found a 15 bp deletion in the FLPo gene as well as a synonymous Thr -> Thr (ACC -> ACA) substitution at position 47 of the mCherry gene; in another, we found a Gly -> Ser mutation at position 4 of the mCherry gene. We did not find mutations in the transgenes of other two SiR-FLPo genomes sequenced (data not shown).

Supplementary Figure S1: mCherry fluorescence from SiR-CRE is much dimmer than tdTomato fluorescence in Ai14 mice, suggesting that the disappearance of the brighter cells in SiR-CRE-injected Ai14 mice indicates their death.

A) Representative images of red fluorescence in SiR-CRE-labeled cells in Ai14 (Cre-dependent expression of tdTomato, top row) and Ai35 (Cre-dependent expression of Arch-EGFP-ER2, bottom row). The three images for each mouse line are from 3 different mice of each line, imaged 7 days following SiR-CRE injection (see Methods), all with the same laser intensity and wavelength (1020 nm). Red fluorescence due only to mCherry (i.e., in Ai35 mice) is obviously much dimmer than that due to tdTomato (i.e., in Ai14 mice). Scale bar: 50 μm, applies to all images.

B) Intensity of red fluorescence of SiR-CRE-labeled cells in Ai14 (left) and Ai35 (right) mice. Data point indicate intensity of individual cells in arbitrary units at the same laser and microscope settings (see Methods). Box plots indicate median, 25th–75th percentiles (boxes), and full range (whiskers) of intensities for each mouse. The average of the mean red fluorescent intensity in each mouse was 48.97 in Ai14 and 15.69 in Ai35 ($p=0.00283 < 0.01$, one-way ANOVA); the midpoint of these means, 32.33, was used as the cutoff for the reanalysis of the data in Ai14 mice to exclude neurons that could have been labeled with mCherry alone.

Supplementary Video S1: Video of 95% of SiR-CRE-labeled neurons in an Arch-EGFP-ER2 reporter mouse disappearing between 11 days and 28 days postinjection.

Two-photon image stacks of a single FOV of visual cortical neurons in an Ai35 mouse imaged at four different time points; time in the video represents depth of focus. Large blobs are glia. 18 out of the 19 neurons visibly labeled with Arch-EGFP-ER2 at 11 days following injection of SiR-Cre are no longer visible 17 days later. White circles indicate cells present at both 11 days and all subsequent imaging sessions; red circles indicate cells present at 11 days but gone by 28 days.

Supplementary Figure S2: 81% of SiR-CRE-labeled neurons in tdTomato reporter mice disappear within 2-4 weeks, even excluding dimmer cells that might have only been labeled with mCherry.
A) Same representative fields of view as in Figure 4 but with circles now marking only cells with intensity at 7 days of greater than 32.33 a.u. (see text and Supplementary Figure S1) that are no longer visible at a subsequent time point. Scale bar: 50 µm, applies to all images.
B-D) Numbers of cells above threshold fluorescence intensity at week 1 that were still present in subsequent weeks. The conclusions from Figure 4 are unchanged: few cells labeled with RV∆GL-4Cre were lost, RV∆G-4Cre killed a significant minority of cells, and SiR-CRE killed the majority of labeled neurons within two weeks following injection.
E) Percentages of cells above threshold at week 1 that were still present in subsequent imaging sessions. By 28 days postinjection, an average of only 19.2% of suprathreshold SiR-CRE-labeled cells remained.

Supplementary File S2: Sanger sequencing data of all clones shown in Figure 1.
51 clones derived from SiR-CRE, 50 from SiR-FLPo, and 51 from RV∆G-4mCherry are identified by their unique indices. All of the indices as well as the sequences corresponding to the 3' end of the nucleoprotein gene are shown.

Supplementary File S3: Summary tables of SMRT sequencing data. These tables show all mutations occurring at positions mutated at greater than 2% frequency in the three virus samples analyzed. Position numbers in these tables refer to the sequences in the three Genbank files below (Supplementary Files S4-S5).

Supplementary File S4: SiR-CRE amplicon reference sequence. This Genbank-format file contains the expected (i.e., based on the published sequence: Addgene #99608) sequence of amplicons obtained from SiR-CRE for SMRT sequencing.

Supplementary File S5: SiR-FLPo amplicon reference sequence. This Genbank-format file contains the expected (i.e., based on the published sequence: Addgene # 99609) sequence of amplicons obtained from SiR-FLPo for SMRT sequencing.

Supplementary File S6: RV∆G-4Cre amplicon reference sequence. This Genbank-format file contains the expected (i.e., based on the published sequence: see Addgene #98034) sequence of amplicons obtained from RV∆G-4Cre for SMRT sequencing.

Supplementary Figure S3: First-generation vector RV∆G-4FLPo appears to be toxic to most cells, unlike the comparable first-generation vector RV∆G-Cre. Although we did not rigorously quantify the effect, our FLPo-encoding RV∆G-4Flpo appears to kill neurons more quickly than does RV∆G-4Cre (cf. Figure 4 and Chatterjee et al. {Chatterjee, 2018 #6924}). In this example field of view, most neurons clearly visible at earlier time points have disappeared by 28 days postinjection, leaving degenerating cellular debris. See Discussion for possible reasons why a preparation of a first-generation vector encoding a recombinase may or may not preserve a large percentage of infected neurons. Scale bar: 50 µm, applies to all panels.

Supplementary File S7: Sequencing results for RV∆G-NPEST-4Cre(EnvA), showing that the intended 3' addition to the nucleoprotein was intact in the final stocks used for monosynaptic tracing experiments.
We sequenced the complete region of the 3' addition to the nucleoprotein region from 32 viral particles for each of the two high-titer, EnvA-enveloped preparations of RV∆G-NPEST-

4Cre(EnvA) and RV∆G-N*PEST-4Cre(EnvA) that we made for our *in vivo* experiments. For RV∆G-NPEST-4Cre(EnvA), none of the 32 clones had any mutations in the 3' addition; one clone had a point mutation in the intergenic region between the N and P genes (from TGTATA to TTTATA), and one other clone had a synonymous mutation in the N gene (Ser437, from TCA to TCG). For RV∆G-N*PEST-4Cre(EnvA), 30 of the 32 clones had no mutations in the sequenced region; one clone had a synonymous mutation in the N gene (Asn436, from AAC to AAT), and one other clone had a synonymous mutation in the PEST region (Pro5, from CCG to CCA) after the stop codon.

Supplementary File S8: Counts and statistical analyses of labeled neurons in transsynaptic tracing experiments.

Sheet 1: Detailed cell counts of every (6th) brain section for all mice, including tdTomato+ cell numbers from ipsilateral cortex, contralateral cortex, and thalamus, as well as counts of starter cells for each animal;

Sheet 2: Total counts of labeled cells for each mouse in each condition, as well as the ratios of tdTomato+ cells to starter cells and the average numbers for each condition;

Sheet 3: P-values of comparisons using single factor ANOVAs.

Supplementary Figure S4: Images of starter cells and input regions in the absence of TEVP.

A-B) Closeup images of virus injection sites in barrel cortex for NPEST (A)and N*PEST (B) viruses, without TEVP. Images are from the same sections as Figure 6C but with a 20x (vs 10x) objective. The circled cells are "starter cells", defined as expressing both tdTomato (reporting Cre expressed by the rabies virus) and mTagBFP2 (coexpressed with G). Scale bar: 50 µm, applies to all images in each panel.

C-D) Images of regions presynaptic to barrel cortex for NPEST (A) and N*PEST (B) viruses, without TEVP. Red indicates tdTomato fluorescence reporting activity of the Cre-expressing rabies viruses. Scale bar: 200 µm, applies to all images in each panel.

Supplementary Figure S5: Images of starter cells and input regions with TEVP supplied.

A-B) Closeup images of virus injection sites in barrel cortex using NPEST (A) and N*PEST (B) viruses, with TEVP supplied in starter cells. Images are from the same sections as Figure 6C but with a 20x (vs 10x) objective. The circled cells are "starter cells", defined as expressing both tdTomato (reporting Cre expressed by the rabies virus) and mTagBFP2 (coexpressed with G); many of these cells also express emiRFP670, coexpressed with TEVP from the third helper virus included in these cases. Scale bar: 50 µm, applies to all images in each panel.

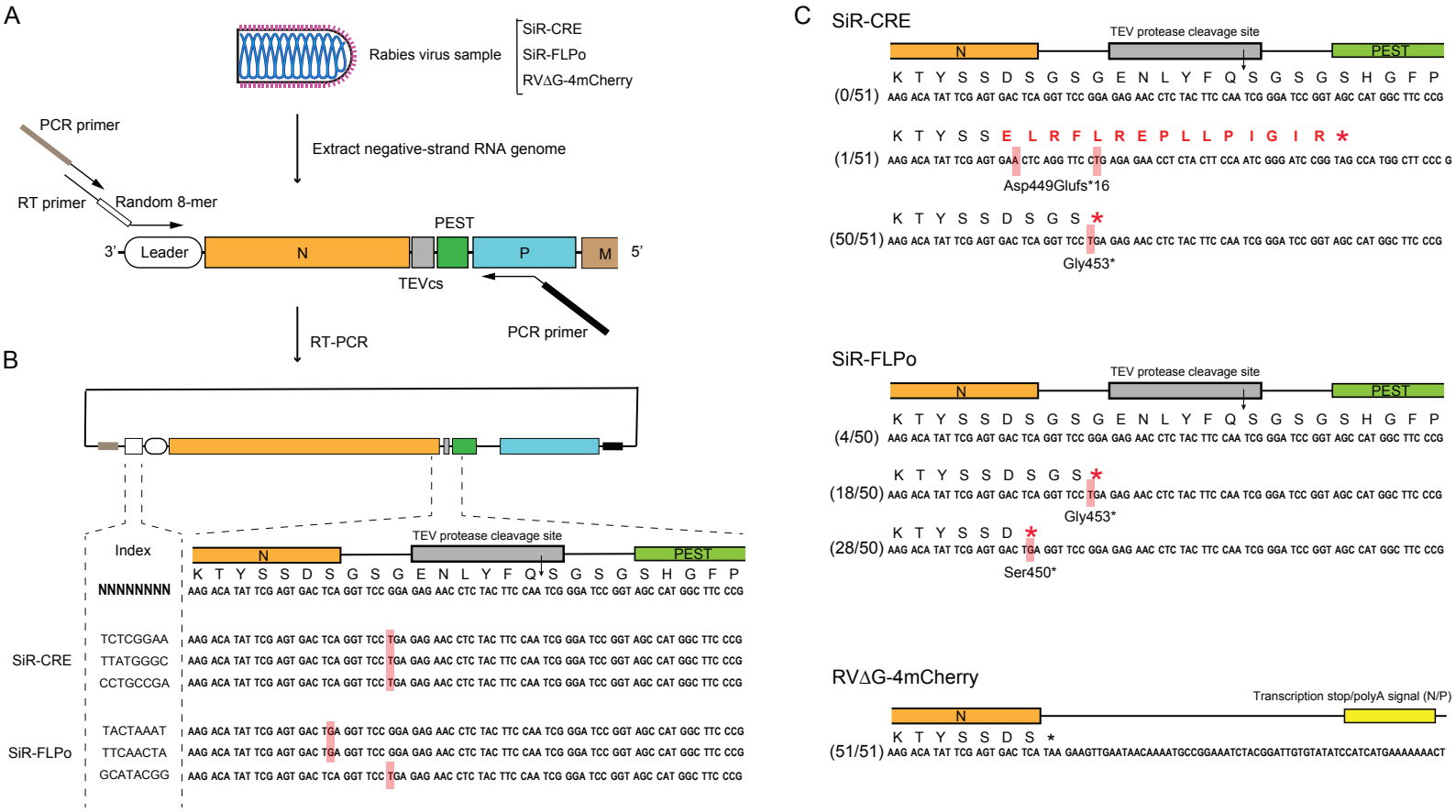
C-D) Images of regions presynaptic to barrel cortex for NPEST (A) and N*PEST (B) viruses, with TEVP supplied in starter cells. The NPEST virus has spread to the same input regions as has N*PEST, although to a more limited degree. Scale bar: 200 µm, applies to all images in each panel.

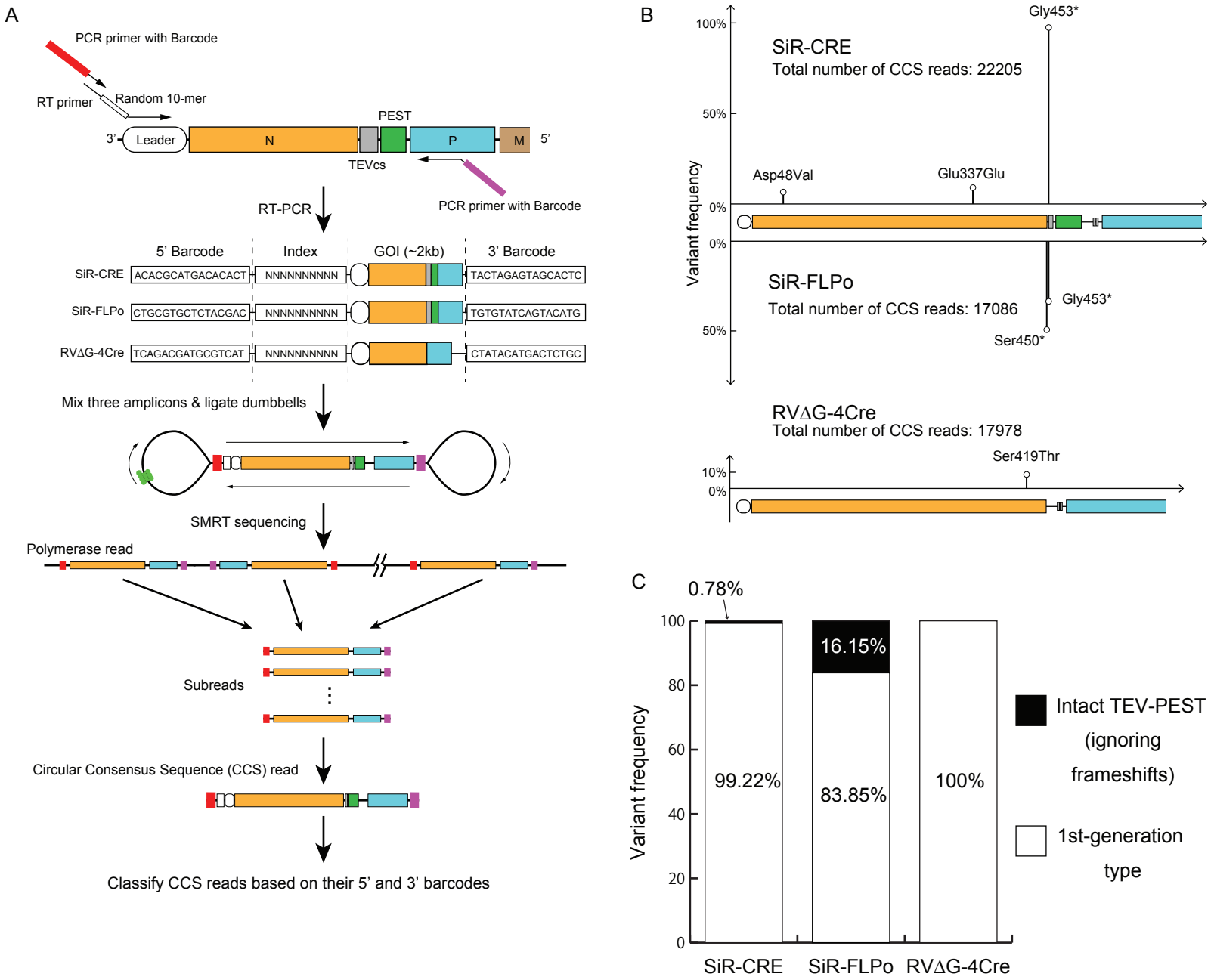
Supplementary Figure S6: Results for control experiments without G, and counts of starter cells for all conditions.

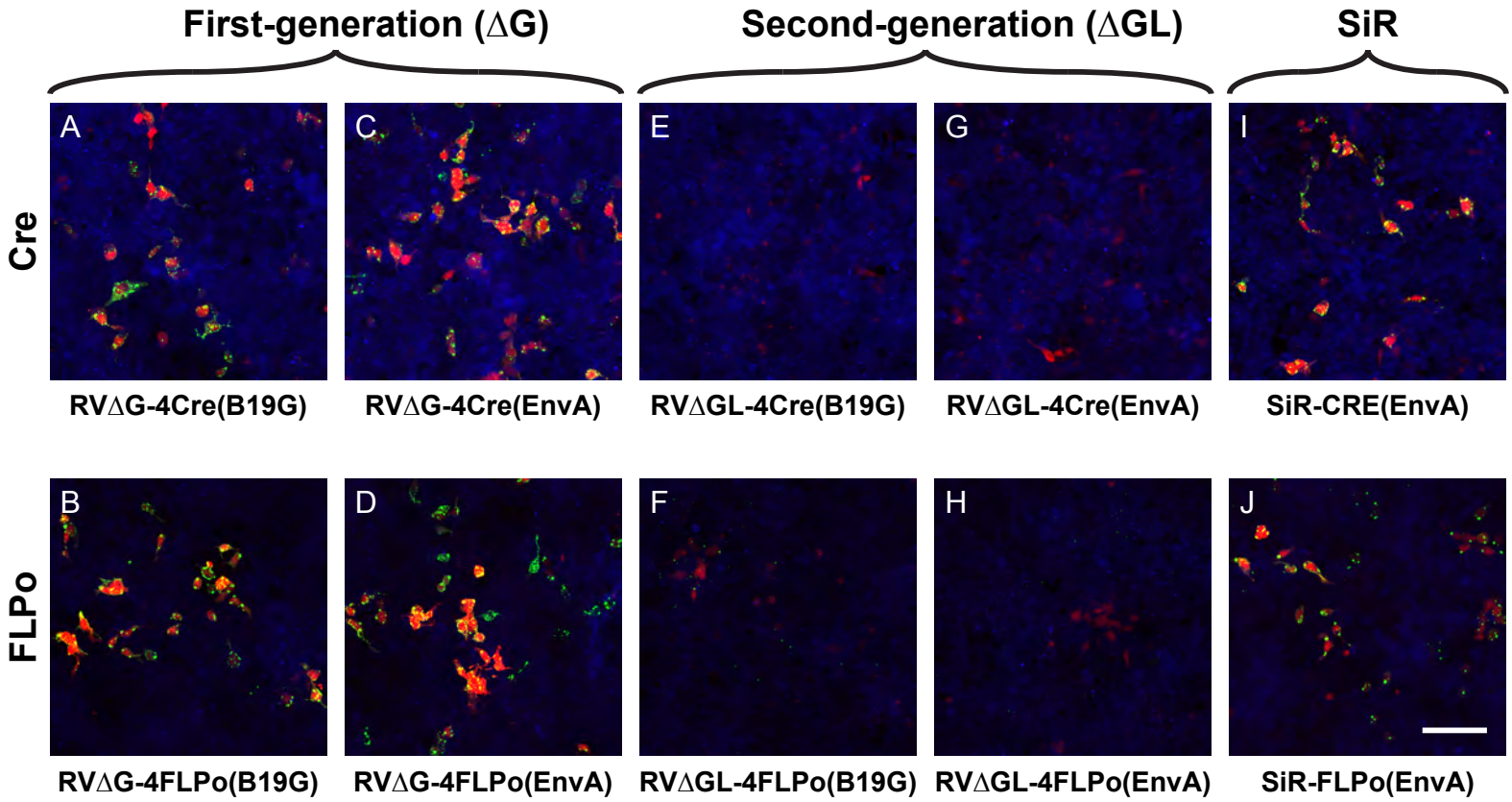
A-B) Images of injection sites for control experiments in which the helper virus expressing G was omitted (but in which the helper virus expressing TEVP was included), using NPEST (A) and N*PEST viruses. Neither virus spreads to any detectable degree without G. Scale bar: 200 µm, applies to all images in each panel.

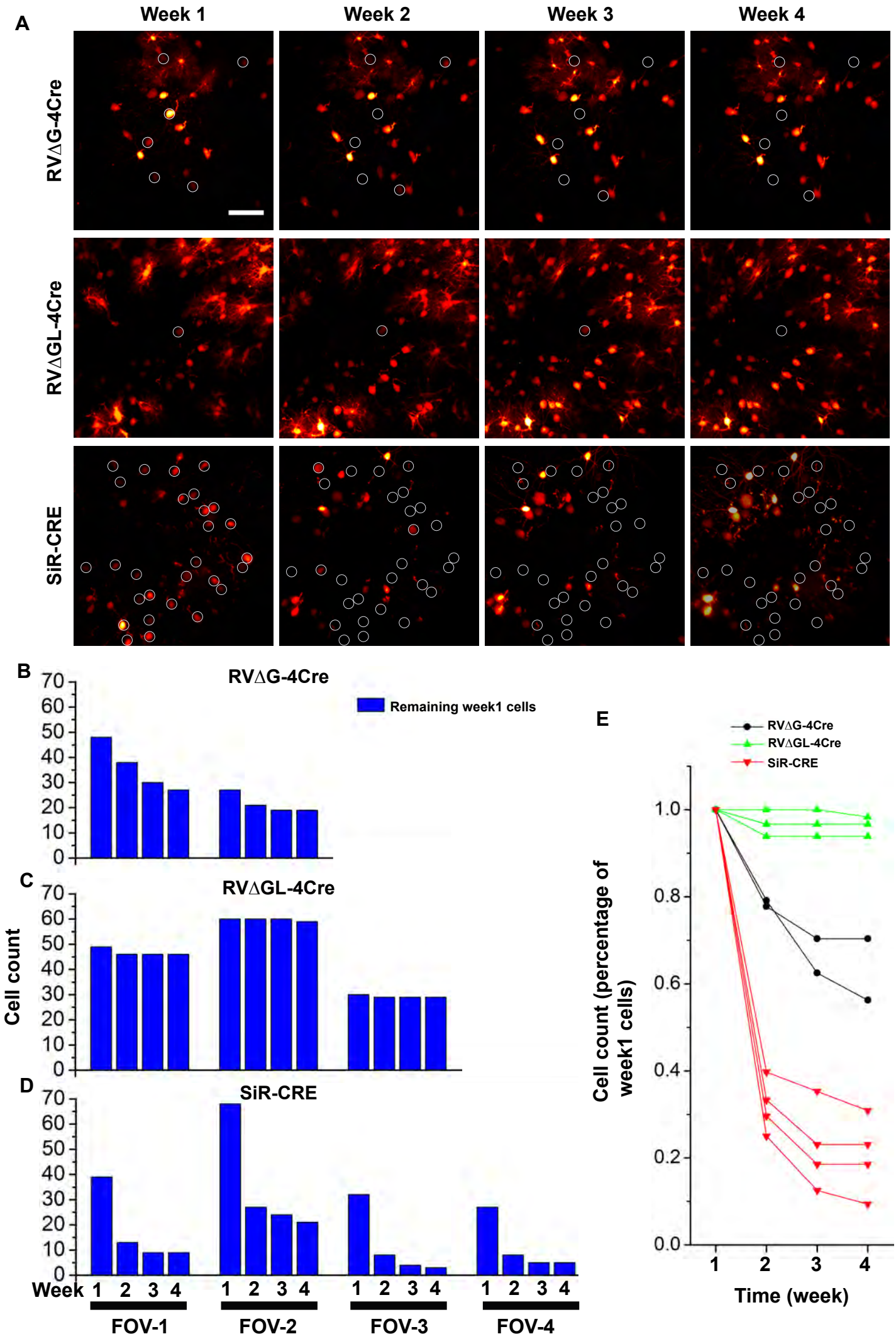
C) Counts of tdTomato-labeled neurons in ipsilateral cortex for the controls without G.

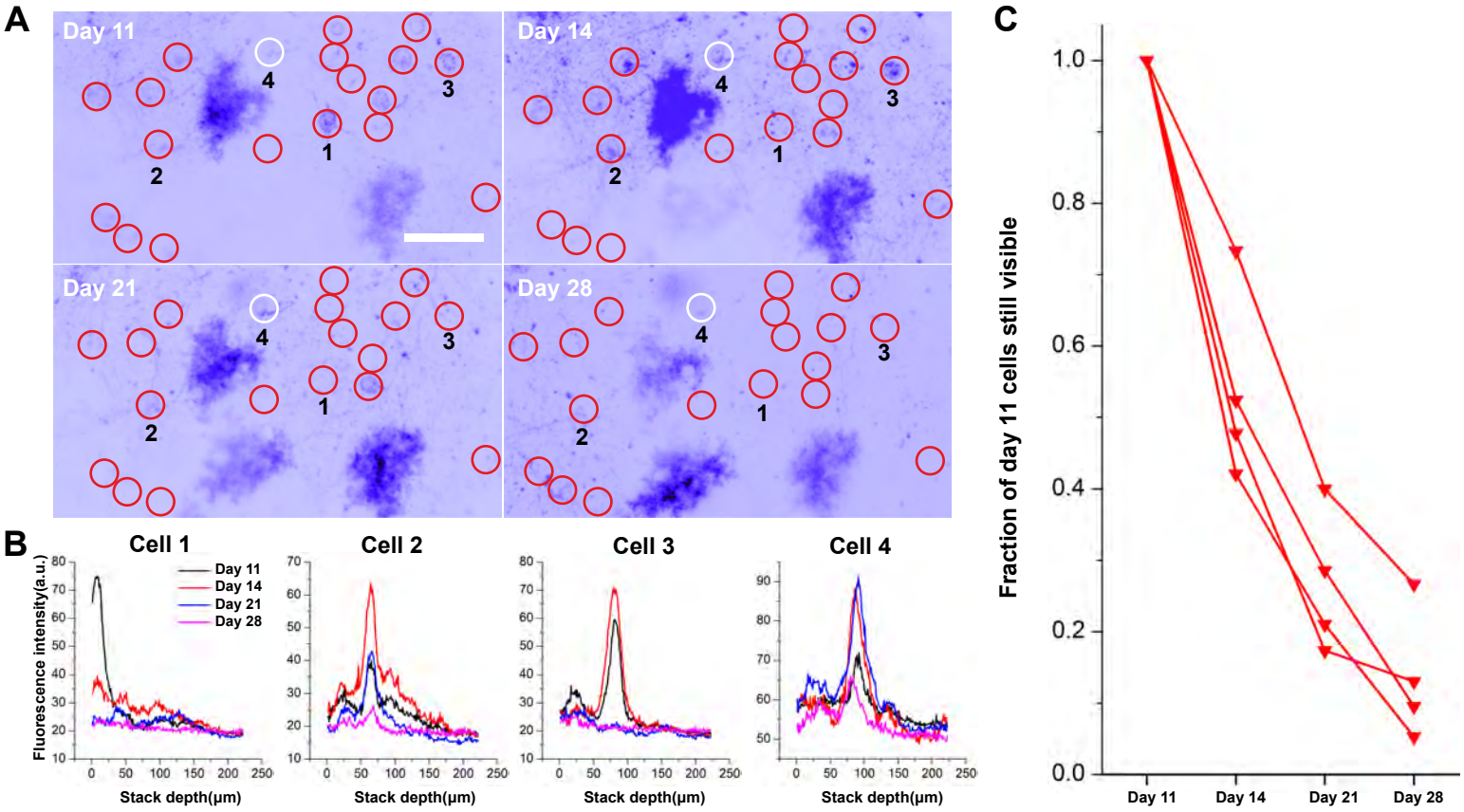
D) Counts of "starter cells" (defined as cells coexpressing tdTomato and mTagBFP2 when TEVP was omitted, and cells coexpressing tdTomato, mTagBFP2, and emiRFP670 when TEVP was supplied) for all conditions.

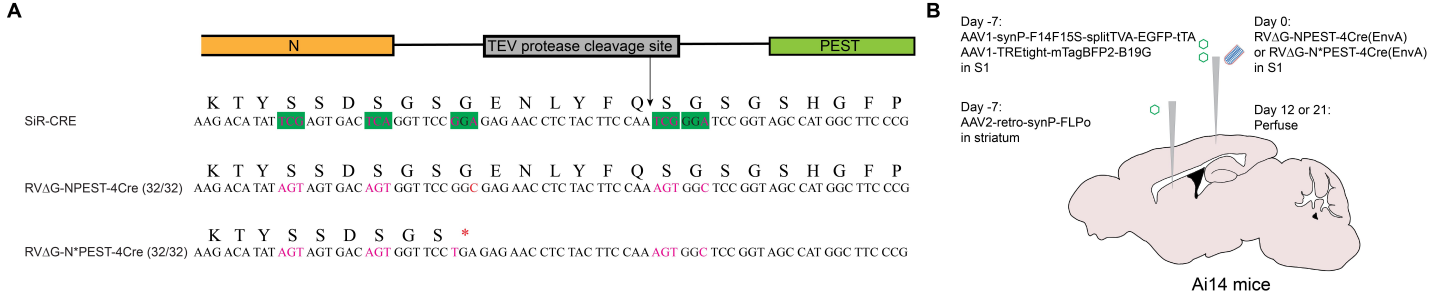

A
Rabies virus sample
SiR-CRE
SiR-FLPo
RVΔG-4mCherry
PCR primer
Extract negative-strand RNA genome
RT primer
Random 8-mer
PEST
3'
Leader
N
P
M
5'
TEVcs
PCR primer
RT-PCR
B
Index
TEV protease cleavage site
N
PEST
K T Y S S D S G S G E N L Y F Q S G S G S H G F P
NNNNNNNN
AAG ACA TAT TCG AGT GAC TCA GGT TCC GGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
SiR-CRE
TCTCGGAA
TTATGGGC
CCTGCCGA
SiR-FLPo
TACTAAAT
TTCAACTA
GCATACGG
C
SiR-CRE
TEV protease cleavage site
N
PEST
K T Y S S D S G S G E N L Y F Q S G S G S H G F P
(0/51) AAG ACA TAT TCG AGT GAC TCA GGT TCC GGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
K T Y S S E L R F L R E P L L P I G I R *
(1/51) AAG ACA TAT TCG AGT GAA CTC AGG TTC CTG AGA GAA CCT CTA CTT CCA ATC GGG ATC CGG TAG CCA TGG CTT CCC G
Asp449Glufs*16
K T Y S S D S G S *
(50/51) AAG ACA TAT TCG AGT GAC TCA GGT TCC TGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
Gly453*
SiR-FLPo
TEV protease cleavage site
N
PEST
K T Y S S D S G S G E N L Y F Q S G S G S H G F P
(4/50) AAG ACA TAT TCG AGT GAC TCA GGT TCC GGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
K T Y S S D S G S *
(18/50) AAG ACA TAT TCG AGT GAC TCA GGT TCC TGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
Gly453*
K T Y S S D *
(28/50) AAG ACA TAT TCG AGT GAC TGA GGT TCC GGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
Ser450*
RVΔG-4mCherry
Transcription stop/polyA signal (N/P)
N
K T Y S S D S *
(51/51) AAG ACA TAT TCG AGT GAC TCA TAA GAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAACT

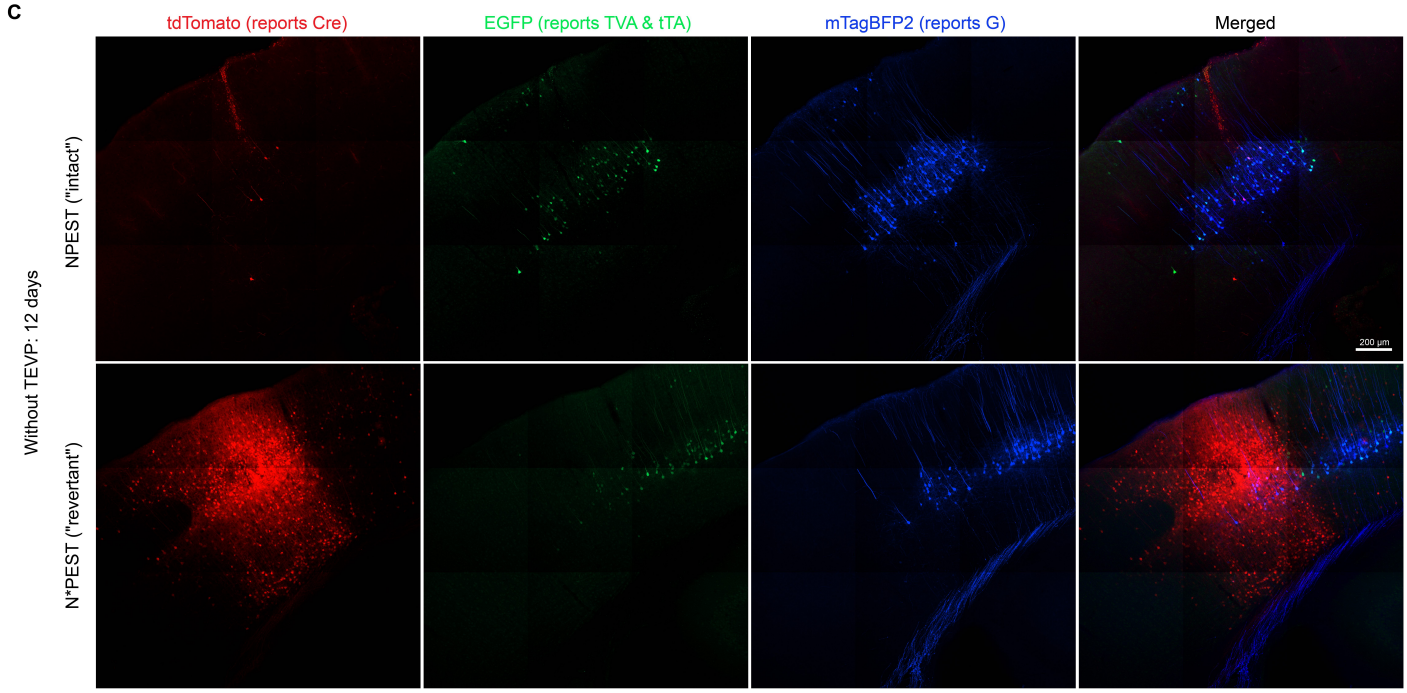

A
PCR primer with Barcode
RT primer
Random 10-mer
3'
Leader
N
PEST
P
M
5'
TEVcs
PCR primer with Barcode
RT-PCR
5' Barcode
Index
GOI (~2kb)
3' Barcode
SiR-CRE
ACACGCATGACACACT
NNNNNNNNNN
TACTAGAGTAGCACTC
SiR-FLPo
CTGCGTGCTCTACGAC
NNNNNNNNNN
TGTGTATCAGTACATG
RVΔG-4Cre
TCAGACGATGCGTCAT
NNNNNNNNNN
CTATACATGACTCTGC
Mix three amplicons & ligate dumbbells
SMRT sequencing
Polymerase read
Subreads
Circular Consensus Sequence (CCS) read
Classify CCS reads based on their 5' and 3' barcodes
B
SiR-CRE
Total number of CCS reads: 22205
Gly453*
Asp48Val
Glu337Glu
100%
50%
0%
Variant frequency
SiR-FLPo
Total number of CCS reads: 17086
Gly453*
Ser450*
50%
0%
RVΔG-4Cre
Total number of CCS reads: 17978
Ser419Thr
10%
0%
C
0.78%
16.15%
99.22%
83.85%
100%
Variant frequency
100
80
60
40
20
0
SiR-CRE
SiR-FLPo
RVΔG-4Cre
Intact TEV-PEST (ignoring frameshifts)
1st-generation type

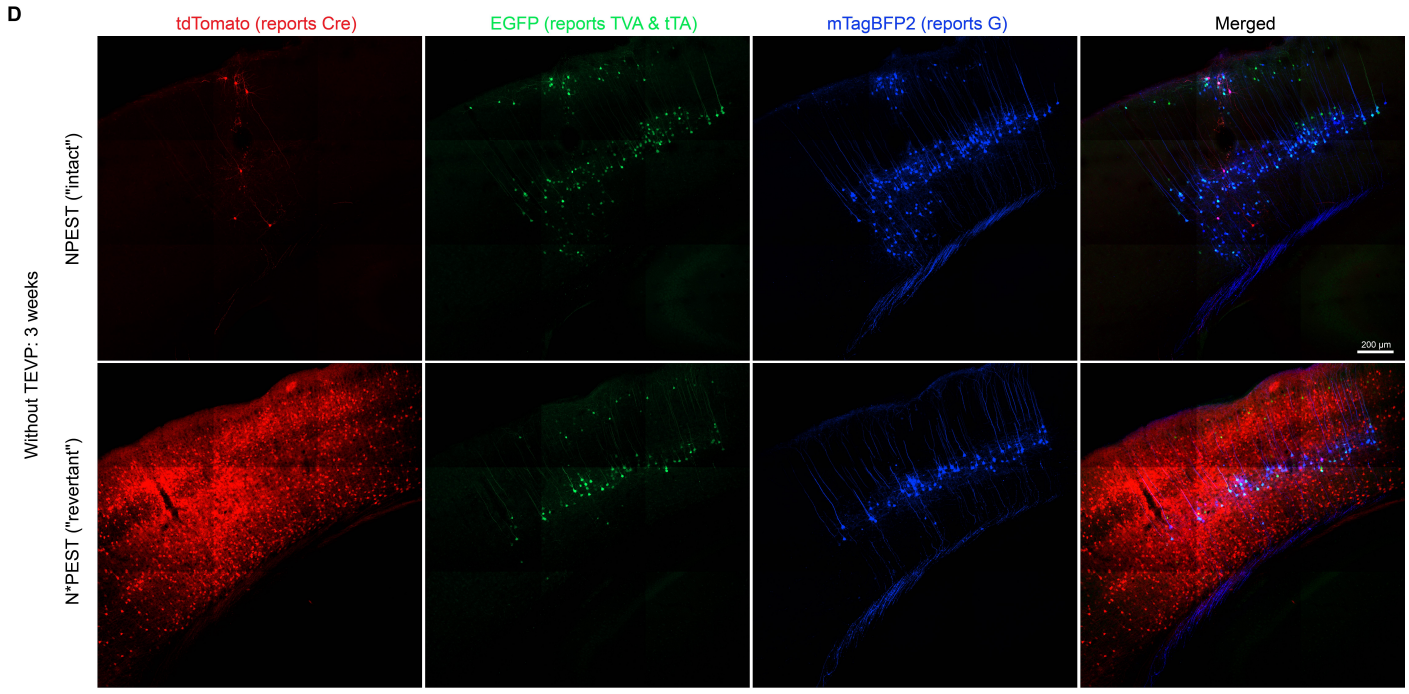

First-generation (ΔG)
Second-generation (ΔGL)
SiR
Cre
FLPo
A
C
E
G
I
B
D
F
H
J
RVΔG-4Cre(B19G)
RVΔG-4Cre(EnvA)
RVΔGL-4Cre(B19G)
RVΔGL-4Cre(EnvA)
SiR-CRE(EnvA)
RVΔG-4FLPo(B19G)
RVΔG-4FLPo(EnvA)
RVΔGL-4FLPo(B19G)
RVΔGL-4FLPo(EnvA)
SiR-FLPo(EnvA)

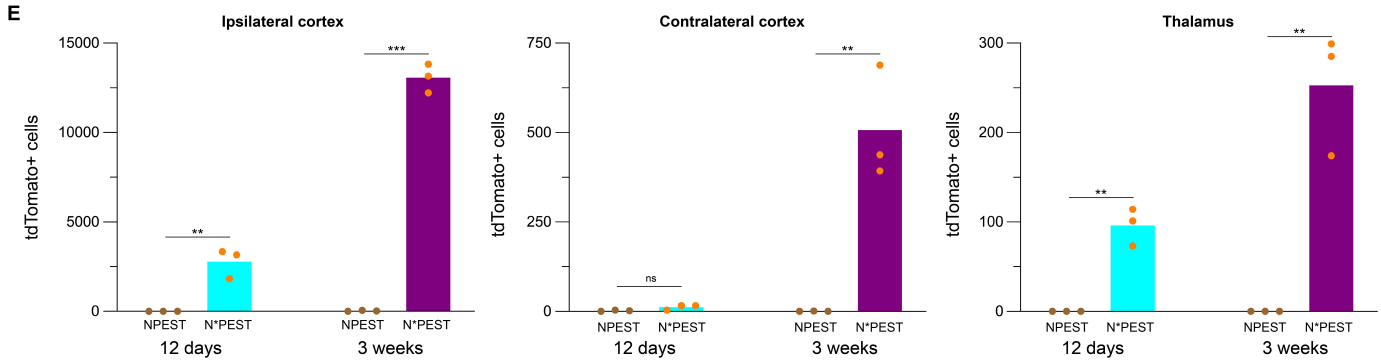

A
Week 1
Week 2
Week 3
Week 4
RVΔG-4Cre
RVΔGL-4Cre
SiR-CRE
B
RVΔG-4Cre
Remaining week1 cells
C
RVΔGL-4Cre
D
SiR-CRE
Cell count
Week
FOV-1
FOV-2
FOV-3
FOV-4
E
RVΔG-4Cre
RVΔGL-4Cre
SiR-CRE
Cell count (percentage of week1 cells)
Time (week)

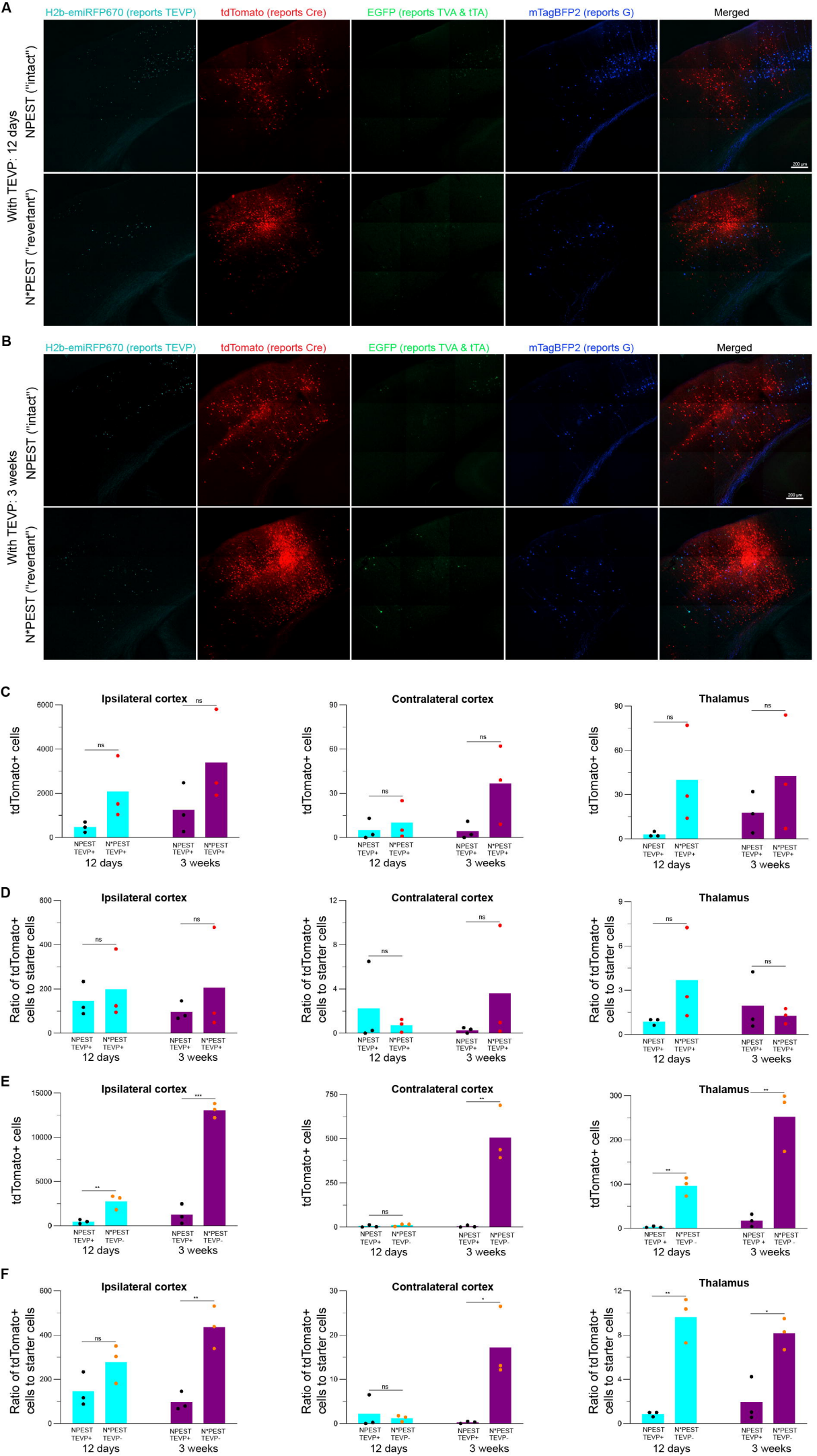

A
Day 11
Day 14
Day 21
Day 28
1
2
3
4
B
Cell 1
Cell 2
Cell 3
Cell 4
Day 11
Day 14
Day 21
Day 28
Fluorescence intensity(a.u.)
Stack depth(μm)
C
Fraction of day 11 cells still visible
Day 11
Day 14
Day 21
Day 28


A
N
TEV protease cleavage site
PEST
SiR-CRE
K T Y S S D S G S G E N L Y F Q S G S G S H G F P
AAG ACA TAT TCG AGT GAC TCA GGT TCC GGA GAG AAC CTC TAC TTC CAA TCG GGA TCC GGT AGC CAT GGC TTC CCG
RVΔG-NPEST-4Cre (32/32)
K T Y S S D S G S G E N L Y F Q S G S G S H G F P
AAG ACA TAT AGT AGT GAC AGT GGT TCC GGC GAG AAC CTC TAC TTC CAA AGT GGC TCC GGT AGC CAT GGC TTC CCG
RVΔG-N*PEST-4Cre (32/32)
K T Y S S D S G S *
AAG ACA TAT AGT AGT GAC AGT GGT TCC TGA GAG AAC CTC TAC TTC CAA AGT GGC TCC GGT AGC CAT GGC TTC CCG
B
Day -7:
AAV1-synP-F14F15S-splitTVA-EGFP-tTA
AAV1-TREtight-mTagBFP2-B19G
in S1
Day 0:
RVΔG-NPEST-4Cre(EnvA)
or RVΔG-N*PEST-4Cre(EnvA)
in S1
Day -7:
AAV2-retro-synP-FLPo
in striatum
Day 12 or 21:
Perfuse
Ai14 mice


C
tdTomato (reports Cre)
EGFP (reports TVA & tTA)
mTagBFP2 (reports G)
Merged
Without TEVP: 12 days
NPEST ("intact")
N*PEST ("revertant")
200 µm


D
tdTomato (reports Cre)
EGFP (reports TVA & tTA)
mTagBFP2 (reports G)
Merged
Without TEVP: 3 weeks
NPEST ("intact")
N*PEST ("revertant")
200 µm


E
Ipsilateral cortex
Contralateral cortex
Thalamus
tdTomato+ cells
15000
10000
5000
0
750
500
250
300
200
100
**
***
ns
NPEST
N*PEST
12 days
3 weeks


A
H2b-emiRFP670 (reports TEVP)
tdTomato (reports Cre)
EGFP (reports TVA & tTA)
mTagBFP2 (reports G)
Merged
With TEVP: 12 days
NPEST ("intact")
N*PEST ("revertant")
200 µm
B
With TEVP: 3 weeks
C
Ipsilateral cortex
Contralateral cortex
Thalamus
tdTomato+ cells
ns
NPEST TEVP+
N*PEST TEVP+
12 days
3 weeks
D
Ratio of tdTomato+ cells to starter cells
E
NPEST TEVP-
N*PEST TEVP-
F