"Self-inactivating" rabies viruses are susceptible to loss of their intended attenuating modification

Makoto Matsuyama[1], Lei Jin[1], Thomas K. Lavin[1], Heather A. Sullivan[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

SUMMARY

An article in *Cell* reported a new form of modified rabies virus that was apparently capable of labeling neurons "without adverse effects on neuronal physiology and circuit function" but that nevertheless was able to spread between neurons as efficiently as the widely-used first-generation deletion-mutant (ΔG) rabies viral vectors. The new "self-inactivating" rabies ("SiR") viruses differed from first-generation vectors only by the addition of a destabilization domain to the viral nucleoprotein. We noticed that the transsynaptic tracing results from that article appeared inconsistent with the strategy described in it: specifically, the viruses were able to spread between neurons even in the absence of the exogenous protease that was meant to be required. We hypothesized that the viruses used were actually mutants that had lost the intended addition to the nucleoprotein, making them *de facto* first-generation viruses. We obtained samples of two SiR viruses from the authors and show here that the great majority of viral particles in both the "SiR-CRE" and "SiR-FLPo" samples were mutants that had lost the intended modification, consistent with our hypothesis. We also found that SiR-CRE killed 70% of infected neurons *in vivo* within two weeks, consistent with the prediction that mutants without the intended modification would share the toxic phenotype typical of first-generation rabies viral vectors. We hypothesize that the same or similar mutations were present in the viruses used in the original article and that this explains the paradoxical reported findings. While it may be possible to successfully make SiR viral preparations that are not dominated by such mutants, and while it may also be possible that such intact SiR viruses are indeed nontoxic to neurons, we predict that it will not be possible to replicate the transsynaptic tracing results from the original paper unless using mutants similar to the ones that we report here.

INTRODUCTION

Monosynaptic tracing based on deletion-mutant rabies virus has become an important tool in neuroscience since its introduction in 2007 (1) and remains the only way of labeling neurons directly presynaptic to some targeted group of neurons in the absence of prior hypotheses (2-5). Its core principles 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 its envelope glycoprotein) and, second, *in vivo* complementation of the deletion, by expression of the deleted gene in *trans*, in the targeted starting neurons. With all of its gene products thus 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 (1).

A drawback of these ΔG (or "first-generation" (6)) rabies viruses is that they are cytotoxic (6-8), which has spurred several labs to develop less-toxic versions. Reardon, Murray, and colleagues (8) 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 (9) — decreased

toxicity and increased transsynaptic labeling efficiency. 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 (6). Deletion of L reduces the transcriptional activity of the virus to extremely low levels, eliminating detectable toxicity but necessitating that the viruses encode Cre or FLPo recombinase (instead of more typical payloads such as fluorescent proteins) and are used in reporter mice or in combination with recombinase-dependent adeno-associated viral vectors (AAVs). While we showed in our paper only that these second-generation, "ΔGL" viruses are efficient means of direct retrograde targeting of projection neurons, ΔGL viruses can also be used (as shown in work so far unpublished) for monosynaptic tracing, at the cost of the additional complexity of expressing the second deleted gene in *trans*.

In another notable development, Ciabatti et al. reported in a paper published in *Cell* that they had created a nontoxic version of ΔG rabies virus that could be used for monosynaptic tracing with no need to express any additional genes other than G (10). The new virus, termed "self-inactivating rabies" virus or "SiR", was simply a ΔG, SAD B19 strain rabies virus except for one modification: the addition of a destabilization domain to the viral nucleoprotein, encoded by "N", the first gene in the rabies viral genome. This destabilization (or "PEST") domain was linked to the nucleoprotein's C terminus by a short linker containing the cleavage site of tobacco etch virus protease (TEVP). The authors' stated intent was that, in the absence of TEVP, newly-produced nucleoprotein (the indispensable protein which encapsidates the viral genome and without which no transcription or replication can occur) would be rapidly degraded because of the attached destabilization domain and that the transcription and replication activity of the virus would therefore be greatly reduced, making the virus nontoxic. To compensate for this reduction in transcription, the SiR viruses encoded Cre or FLPo recombinases and were used in reporter mice or with reporter AAVs, paralleling our own strategy using ΔGL rabies viruses (6). In the presence of TEVP, however, the destabilization domain would be cleaved from the C terminus of the nucleoprotein, which therefore would not be degraded and would allow normal levels of transcription and replication of the virus. In the words of Ciabatti et al., "the virus should be able to transcribe and replicate only when TEVp is present" (10).

While most of the results reported in their paper are consistent with a virus that is conditionally deficient in nucleoprotein production, Ciabatti et al. went on to show that their "SiR" viruses could be used for monosynaptic tracing *in vivo* with no need for TEVP expression at all: the viruses could simply be complemented by G expression in the starting cell population and efficiently spread to putatively presynaptic neurons. In their lack of dependence on TEVP expression, the SiR viruses behaved as if they were simply first-generation, ΔG rabies viruses. This was a paradoxical result: the modified nucleoprotein was supposed to require TEVP-mediated removal of the destabilization domain in order to accumulate in infected cells and allow viral replication, but it appeared not to in the starting cells *in vivo*. Furthermore, despite this apparent ability to replicate perfectly well despite the C-terminal degron that was intended to prevent it from doing so, the virus appeared not to kill neurons: Ciabatti et al. found many surviving putatively-presynaptic neurons at three weeks after rabies virus injection.

We hypothesized an alternative explanation for these results, as follows.

As an incidental finding from the control experiments in our paper on second-generation (ΔGL) rabies virus (6), we discovered that even first-generation (ΔG) virus could apparently be made significantly less toxic simply by switching the transgene that it encodes from some "normal" one like the tdTomato gene to the coding sequence for Cre recombinase. Following an initial die-off of a fraction of neurons infected with a Cre-encoding ΔG virus, a comparably large fraction survived for four months, the longest we followed them, with modest physiological changes in some (6). These findings suggested a "quick and dirty" way of making a less-toxic, but by no means completely nontoxic, monosynaptic tracing system: use a first-generation, ΔG rabies virus encoding a recombinase and do the experiments in reporter mice.

We hypothesized that Ciabatti et al. had inadvertently done exactly that. Rhabdoviruses have high mutation rates (11-16), and production of high-titer rabies virus stocks for *in vivo* injection typically involves repeated passaging on complementing cell lines (17-19), which affords ample opportunity for accumulation of mutants with a selective replication advantage. We therefore hypothesized that the viruses that Ciabatti et al. had actually ended up with and used for their transsynaptic tracing experiments were mutants with premature stop codons at or near the end of the native nucleoprotein gene and before the sequence of the destabilization domain.

If Ciabatti et al.'s viruses had lost the intended C-terminal addition because of mutation, their "SiR-CRE" virus would in practice be a simple first-generation, ΔG rabies viral vector expressing Cre. Because we have shown that such a virus can leave a large percentage of infected cells alive, the presence of surviving cells even at long time points would have led the authors to conclude that they had developed a new kind of rabies virus that did not kill cells but that could be used for monosynaptic tracing by simple complementation by G alone.

Here we show that, in both of the two SiR virus samples to which we had access, the vast 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.

RESULTS

We analyzed samples of two viruses sent directly from the Tripodi lab to MIT in September 2017, two months after the publication in which they were introduced (10): "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). These two viruses, like all of the other rabies viruses described in Ciabatti et al., have the SAD B19 strain of rabies virus as their parent strain. Both of the viruses sent by the Tripodi lab had been packaged with the avian and sarcoma virus subgroup A envelope glycoprotein ("EnvA") for targeted infection of cells expressing EnvA's receptor, TVA (1).

For comparison with the two SiR viruses, we made five control viruses in our own laboratory: three first-generation vectors RVΔG-4Cre (6), RVΔG-4FLPo (see Methods), and RVΔG-4mCherry (20), and two second-generation vectors RVΔGL-4Cre and RVΔGL-4FLPo (6). All of these viruses are also 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 packaged just 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 – 51 individual viral particles of each of the two SiR viruses and for 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 C-terminal addition to the nucleoprotein.

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 sequenced region on 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 Sanger sequencing, we included a random index (10 bases, in this case) in the reverse transcription primer, so that again the cDNA copy of each RNA viral genome molecule would be labeled with a unique index.

SMRT sequencing entails circularization of the DNA amplicons and multiple consecutive passes around the resulting circular molecule, with the redundancy provided by this repeated sequencing of each position increasing the signal to noise ratio and statistical significance of the results. The numbers presented in Figure 2 and below use the default of including only clones that had at least three reads of each base ("circular consensus sequence 3", or "CCS3" in Supplementary File S3). Using the increasingly stringent criteria of requiring either five or eight reads per base (CCS5 or CCS8) reduced the numbers of qualifying genomes in all cases and changed the percentages slightly but gave very similar results overall. Because read accuracy for SMRT sequencing is ≥98% for circular consensus sequencing with 3 passes (see https://www.mscience.com.au/upload/pages/pacbio/technical-note---experimental-design-for-targeted-sequencing.pdf), we used a conservative threshold of 2% frequency of any given point mutation position in order to screen out false positives. Also to be very conservative, for Figure 2 we ignored all apparent frame shifts caused by insertions and deletions, because insertions in particular are prone to false positives with SMRT sequencing (21). See Supplementary File S3 for details, including details of frameshifts due to insertions; Supplementary Files S4-S6 contain the sequences of the PCR amplicons that would be expected based on published sequences of the three viruses, but to summarize here:

As a control, we used a virus from our own laboratory, RVΔG-4Cre (6) (see Addgene #98034 for reference sequence). Out of 17,978 sequenced genomes of this virus, we found no mutations above threshold frequency at the end of N. We did find that 1,706 viral particles (9.49%) had a nonsynonymous mutation (TCT (Ser) → ACT (Thr)) farther up in N at amino acid position

419 (31 amino acids upstream of the end of the 450-aa native protein). We do not know if this mutation is functionally significant, although it is not present in CVS N2c (22), HEP-Flury (23), ERA (24), or Pasteur strains (Genbank GU992320), so these particles may effectively be N-knockouts that were propagated by coinfection with virions with intact N (see Discussion for more on such parasitic co-propagating mutants).

For the SiR-CRE virus, out of 22,205 viral genomes sequenced, 22,032 had the premature stop codon (GGA -> TGA) in the linker between the native nucleoprotein gene and the TEVP cleavage site sequence. In other words, even without including frameshifts, at least 99.22% of the individual viral particles in the SiR-CRE sample were simply first-generation ΔG vectors.

For the SiR-FLPo virus, out of 17,086 viral genomes sequenced, 5,979 had the stop codon (GGA -> TGA) in the linker, 8,624 had the stop codon (TCA -> TGA) at the end of N, and a further 28 had a different stop codon (TCA -> TAA) at the same position at the end of N. Of these, 305 viral particles had premature stop codons at both of these two positions, so that the total number of viral particles with one or both stop codons immediately before the PEST domain was (8624 + 5979 + 28 - 305 = 14,326. In other words, at least 83.85% of the individual viral particles in the SiR-FLPo sample were simply first-generation ΔG vectors, with the only modification of the nucleoprotein being either two amino acids added to, or one amino acid lost from, the C-terminus.

Anti-nucleoprotein immunostaining

We infected reporter cell lines with serial dilutions of the two EnvA-enveloped SiR viruses as well as the eight recombinase-expressing ones from our own lab: ΔG vs. ΔGL, Cre vs. FLPo, EnvA vs. B19G envelopes. Three days later, we immunostained the cells for rabies virus 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 (6), 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 (6) (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 (6), 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. Again as we have found previously (6), our first-generation virus RVΔG-4Cre did kill neurons, but by no means all of them (see the Discussion for a possible reason 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 Ciabatti et al. intended it to be (10), and a significant fraction of SiR-CRE virions had mutations in the Cre gene so 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 (25) (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 (26), particularly with two-photon excitation (27), 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 (28), 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." (28)) nor uniform (29); 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 (6)).

DISCUSSION

Although many of the results presented in Ciabatti et al. (10) seemed internally consistent, the results showing transsynaptic tracing were puzzling. The degron fused to the C terminus of the nucleoprotein was supposed to require removal by TEV protease in order to allow replication of the SiR virus, but the authors reported that the viruses spread efficiently *in vivo* with no TEV protease present at all. The results that we have shown here suggest that the likely explanation is that the SiR viruses had simply lost the intended C-terminal modification of the nucleoprotein through mutation, as we found had happened in both of the two viruses given to us by the authors shortly after publication of their paper.

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 the following three reasons. First, both of the two SiR virus samples to which we had access 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. Second, 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. Third, Ciabatti et al.'s finding that viral replication and spread occurred *in vivo* 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.

If the viruses used for the otherwise-unexplained transsynaptic tracing experiments in Ciabatti et al. were actually *de facto* first-generation ΔG viruses like the SiR samples that we analyzed, how could the authors have found, in postmortem tissue, cells labeled by SiR-CRE that had survived for weeks? The answer may simply be that, as we have shown in Chatterjee et al. (6) and again here (Figure 4), a preparation of first-generation rabies viral vector expressing Cre can leave a large fraction of labeled cells alive for at least months, in stark contrast to similar ones encoding tdTomato (6) or EGFP (7). Similarly, Gomme et al. found long-term survival of some neurons following infection by a replication-competent rabies virus expressing Cre (30)).

The reason, in turn, why a preparation of a simple ΔG rabies virus encoding Cre can leave many cells alive may be that not all the virions are in fact first-generation viral particles, due, fittingly, to the same high mutation rate that we have highlighted in this paper. We have shown in Chatterjee et al. (6) that a second-generation (ΔGL) rabies virus, which has both its glycoprotein gene G and its polymerase gene L deleted, leaves cells alive for the entire four months that we followed them. However, any first-generation (ΔG) virus that contains a frameshift or point mutation knocking out L will in practice be a ΔGL virus. Indeed, a stop codon or frameshift mutation in the several other viral genes is likely to have a similar effect as one in L (and it might be that the Ser419Thr mutation that we found in 9.49% of our RVΔG-4Cre virions is just such a knockout mutation of N). Together with the high mutation rate of rabies virus, this means that, within every preparation of first-generation rabies virus there is almost guaranteed to be a

population of *de facto* second-generation variants mixed in with the intended first-generation population and propagated in the producer cells by complementation by the first-generation virions. Any rabies virus preparation (whether made in the laboratory or occurring naturally) can be expected to contain a population of such knockout (whether by substitution, frameshift, or deletion) mutants (related to the classic phenomenon of "defective interfering particles", or mutants with a marked replication advantage (31-33), and the higher the multiplicity of infection when passaging the virus, the higher the proportion of such freeloading viral particles typically will be. This would not necessarily be noticed in the case of a virus encoding a more common transgene product such as a fluorophore, because the expression levels of these by the knockout mutants would be too low to label cells clearly (see Figure 1 in Chatterjee et al. (6)). However, with Cre as the transgene, any "second-generation" particles would be able to label neurons but not kill them, because second-generation rabies viral vectors do not kill cells for at least months (6). This explanation would predict that the percentage of neurons surviving infection with a rabies virus encoding Cre will depend on the particular viral preparation that is injected, with some having a greater fraction of knockout particles than others.

This could explain why the SiR-CRE virus sample killed cells faster than our own RVΔG-4Cre (Figure 4). This analysis would also presumably apply to first-generation (ΔG) viruses expressing FLPo: while we found that the FLPo-expressing version that we made did not leave as many cells alive as the Cre-expressing version (Supplementary Figure S3) did, that preparation may simply have had fewer mutants with knockout of genes essential for replication. It is unclear whether Ciabatti et al.'s SiR-FLP would leave many cells alive for long durations, as the authors did not show any results indicating that it does, and we did not test it *in vivo* ourselves.

To summarize, we have found that both of the SiR virus samples that we analyzed consisted overwhelmingly of mutants that had lost the modification that was intended to make them nontoxic. We believe that this provides the most likely explanation for Ciabatti et al.'s paradoxical results. However, because we don't know for certain that the viruses that Ciabatti et al. used for their paper had the same problem as did the stocks that they gave us, we cannot rule out the following three-part alternative explanation: first, that their transsynaptic tracing experiments were in fact performed with unmutated viruses that only expressed nucleoprotein molecules with intact PEST domains fused to their C termini; second, that these PEST domains, contrary to their intended behavior, did not cause rapid degradation of the nucleoprotein but instead allowed it to accumulate sufficiently to result in viral replication and spread; third, that the large (159 aa) C-terminal addition did not noticeably interfere with the function of the nucleoprotein, with the result that the modified viruses spread just as efficiently as the unmodified versions, even in the absence of TEVP.

We regard this alternative explanation as implausible, for the reasons discussed above. However, because of this uncertainty, the conservative conclusions from our findings are simply that SiR viruses are susceptible to loss of their intended modification and that the approach should be used with caution and that the published results should be reinterpreted in light of our findings. We further submit that, because there is currently no evidence that the viruses from the original paper were not mutants like both of the batches that we analyzed, the ability of an intact SiR virus to spread transsynaptically, with or without TEVP, remains unproven.

To be clear, we have no reason to believe that it is impossible to make a rabies virus with the C-terminal addition to the nucleoprotein that Ciabatti et al. had intended. A number of groups have made recombinant rabies viruses — as well as other rhabdoviruses and other nonsegmented negative-strand RNA viruses — encoding fusions of exogenous proteins to viral proteins (34-44). However, most of these groups have found that the additions significantly impaired the function of the viral proteins. Most relevantly, an attempt to make SAD B19 rabies virus with EGFP fused to the C terminus of the nucleoprotein was unsuccessful, suggesting that the nucleoprotein is intolerant of large C-terminal additions; the authors of that paper resorted

instead to making virus encoding the fusion protein *in addition to the wild-type nucleoprotein* because the fusion protein was evidently dysfunctional (37).

These prior attempts have also found that rhabdoviruses can rapidly lose C-terminal portions of viral or exogenous proteins. A vesicular stomatitis virus with GFP fused to the C terminus of the glycoprotein gene lost the modification within a single passage of the virus because of a point mutation creating a premature stop codon (34). Relatedly, a VSV with its glycoprotein gene replaced with that of a different virus was found to quickly develop a premature stop codon causing loss of the last 21 amino acids of the exogenous glycoprotein, conferring a marked replication advantage to the mutants bearing the truncated version (45).

Generalizing from these prior examples as well as our findings here, we suggest that any attempt to attenuate a virus by addition to the C terminus of a viral protein will be vulnerable to loss of the addition by creation of either a stop codon or a frameshift, and that any such virus will therefore need to be monitored very carefully.

Despite this, we do not claim that it is impossible, given sufficient care, to make SiR virus with the intended C-terminal addition intact in the great majority of viral particles. A good way to start might be to make synonymous mutations to the codons that had most frequently mutated in the stocks that we analyzed, so that those codons are not a single point mutation away from a stop codon (e.g., GGA -> GGG, TCA -> TCT); this improvement would not prevent loss of the C-terminal addition due to frameshift mutations, however.

Indeed, in response to our posting a preprint of an earlier version of this manuscript in February 2019 (46), the authors of the original paper have recently posted a new preprint in which they report that they are now able to make mutation-free SiR virus and that this intact SiR virus is noncytotoxic when used to directly infect cortical neurons (47). This is a good start but does not change anything about our findings, conclusions, and predictions. Crucially, the new preprint does not include any results testing whether the intact SiR virus can spread transsynaptically, either with TEVP or without it. This suggests that, in over 1.5 years since we brought the problem to light, Ciabatti et al. have not replicated, with unmutated virus, the experiments from their original paper.

We predict that they will not be able to. Specifically: it may be the case that an intact SiR virus is capable of at least some transsynaptic spread if provided with both G and TEVP, although again there are no experiments presented either in Ciabatti et al.'s original paper or in their new preprint that tested this, so whether this phenomenon occurs at all, and with what efficiency if it does occur, are currently unknown. We predict, though, that any such intact SiR virus will be incapable of appreciable transsynaptic spread in the *absence* of TEVP, i.e., under the conditions Ciabatti et al. described in *Cell*. To put it another way, we predict that any SiR virus that is found to be capable of efficient transsynaptic spread under the conditions described in Ciabatti et al. '17 — i.e., without provision of TEVP — will, if closely examined, be found to consist largely of "escape mutants" similar to those that we have described here.

METHODS

Cloning

Lentiviral transfer plasmids were made by cloning, into pCSC-SP-PW-GFP (48) (Addgene #12337), the following components:

the CAG promoter (49) and a Cre-dependent "FLEX" (50) construct consisting of pairs of orthogonal lox sites flanking a back-to-back fusion of the gene for mTagBFP2 (51) immediately followed by the reverse-complemented gene for mCherry (26), to make the Cre reporter construct pLV-CAG-FLEX-BFP-(mCherry)';

the CAG promoter (49) and a Flp-dependent "FLEX" (50) construct consisting of pairs of orthogonal FRT sites (52) flanking a back-to-back fusion of the gene for mTagBFP2 (51) immediately followed by the reverse-complemented gene for mCherry (26), to make the Flp reporter construct pLV-CAG-F14F15S-BFP-(mCherry)';

the ubiquitin C promoter from pUB-GFP (53) (Addgene 11155) and the long isoform of TVA (54) to make the TVA expression vector pLV-U-TVA950.

The first-generation vector genome plasmid pRVΔG-4FLPo was made by cloning the FLPo gene (55) into pRVΔG-4Cre.

The above novel plasmids have been deposited with Addgene with accession numbers 115234, 115235, 115236, and 122050.

Production of lentiviral vectors

Lentiviral vectors were made by transfection of HEK-293T/17 cells (ATCC 11268) as described (56) but with the use of the vesicular stomatitis virus envelope expression plasmid pMD2.G (Addgene 12259) for all vectors except for LV-U-TVA950(B19G), which was made using the rabies virus envelope expression plasmid pCAG-B19GVSVGCD (56). Lentiviral vectors expressing fluorophores were titered as described (57); titers of LV-U-TVA950(VSVG) and LV-U-TVA950(B19G) were assumed to be approximately the same as those of the fluorophore-expressing lentiviral vectors produced in parallel.

Production of titering cell lines

To make reporter cell lines, HEK-293T/17 cells were infected with either pLV-CAG-FLEX-BFP-(mCherry)' or pLV-CAG-F14F15S-BFP-(mCherry)' at a multiplicity of infection of 100 in one 24-well plate well each. Cells were expanded to 2x 15cm plates each, then sorted on a FACS Aria to retain the top 10% most brightly blue fluorescent cells. After sorting, cells were expanded again to produce the cell lines 293T-FLEX-BC and 293T-F14F15S-BC, reporter lines for Cre and FLPo activity, respectively. TVA-expressing versions of these two cell lines were made by infecting one 24-well plate well each with LV-U-TVA950(VSVG) at an MOI of approximately 100; these cells were expanded to produce the cell lines 293T-TVA-FLEX-BC and 293T-TVA-F14F15S-BC.

Production and titering of rabies viruses

RVΔG-4Cre and RVΔGL-4Cre were produced as described (6, 19), with EnvA-enveloped viruses made by using cells expressing EnvA instead of G for the last passage. Titering and infection of cell lines with serial dilutions of viruses was as described (17), with the 293T-TVA-FLEX-BC and 293T-TVA-F14F15S-BC lines used for B19G-enveloped viruses and the 293T-TVA-FLEX-BC and 293T-TVA-F14F15S-BC used for the EnvA-enveloped viruses. For the *in vivo* injections, the three EnvA-enveloped, Cre-encoding viruses were titered side by side, and the two higher-titer viruses were diluted so that the final titer of the injected stocks of all three viruses were approximately equal at 1.39E9 infectious units per milliliter.

Immunostaining

Reporter cells (see above) plated on coverslips coated in poly-L-lysine (Sigma) were infected with serial dilutions of RVΔG-4Cre and RVΔGL-4Cre as described (17). Three days after infection, cells were fixed with 2% paraformaldehyde, washed repeatedly with blocking/permeabilization buffer (0.1% Triton-X (Sigma) and 1% bovine serum albumin (Sigma) in PBS), then labeled with a blend of three FITC-conjugated anti-nucleoprotein monoclonal antibodies (Light Diagnostics Rabies DFA Reagent, EMD Millipore 5100) diluted 1:100 in blocking buffer for 30 minutes, followed by further washes in blocking buffer, then finally briefly rinsed with distilled water and air-

dried before mounting the coverslips onto microscope slides with Prolong Diamond Antifade (Thermo P36970) mounting medium. Images of wells at comparable multiplicities of infection (~0.1) were collected on a Zeiss 710 confocal microscope.

Virus injections and surgery

All experimental procedures using mice were conducted according to NIH guidelines and were approved by the MIT Committee for Animal Care (CAC). Mice were housed 1-4 per cage under a normal light/dark cycle for all experiments.

Adult (>9 weeks, male and female) Cre-dependent tdTomato reporter Ai14 (58) (Jackson Laboratory #007908) or Arch-EGFP reporter Ai35D (25) (Jackson Laboratory # 012735) mice were anesthetized with isoflurane (4% in oxygen) and ketamine/xylazine (100mg/kg and 10mg/kg respectively, i.p.). Mice were given preemptive analgesics buprenorphine (0.1 mg/kg s.q.) and meloxicam (2 mg/kg s.q.) as well as eye ointment (Puralube); the scalp was then shaved and mice mounted on a stereotaxic instrument (Stoelting Co.) with a hand warmer (Heat Factory) underneath the animal to maintain body temperature. Following disinfection with povidone-iodine, an incision was made in the scalp, and a 3 mm craniotomy was opened over primary visual cortex (V1). 300 nl of LV-U-TVA950(B19G) (see above) was injected into V1 (-2.70 mm AP, 2.50 mm LM, -0.26 mm DV; AP and LM stereotaxic coordinates are with respect to bregma; DV coordinate is with respect to brain surface) using a custom injection apparatus comprised of a hydraulic manipulator (MO-10, Narishige) with headstage coupled via custom adaptors to a wire plunger advanced through pulled glass capillaries (Wiretrol II, Drummond) back-filled with mineral oil and front-filled with virus solution. Glass windows composed of a 3mm-diameter glass coverslip (Warner Instruments CS-3R) glued (Optical Adhesive 61, Norland Products) to a 5mm-diameter glass coverslip (Warner Instruments CS-5R) were then affixed over the craniotomy with Metabond (Parkell). Seven days after injection of the lentiviral vector, the coverslips were removed and 300 nl of one of the three EnvA-enveloped rabies viral vectors (with equalized titers as described above) was injected at the same stereotaxic coordinates. Coverslips were reapplied and custom stainless steel headplates (eMachineShop) were affixed to the skulls around the windows.

***In vivo* two-photon imaging and image analysis**

Beginning seven days after injection of each rabies virus and continuing every seven days up to a maximum of four weeks following rabies virus injection, the injection sites were imaged on a Prairie/Bruker Ultima IV In Vivo two-photon microscope driven by a Spectra Physics Mai-Tai Deep See laser with a mode locked Ti:sapphire laser emitting at a wavelength of 1020 nm for tdTomato and mCherry or 920 nm for EGFP. Mice were reanesthetized and mounted via their headplates to a custom frame, again with ointment applied to protect their eyes and with a handwarmer maintaining body temperature. One field of view was chosen in each mouse in the area of maximal fluorescent labelling. The imaging parameters were as follows: image size 512 X 512 pixels (282.6 µm x 282.6 µm), 0.782 Hz frame rate, dwell time 4.0 µs, 2x optical zoom, Z-stack step size 1 µm. Image acquisition was controlled with Prairie View 5.4 software. Laser power exiting the 20x water-immersion objective (Zeiss, W plan-apochromat, NA 1.0) varied between 20 and 65 mW depending on focal plane depth (Pockel cell value was automatically increased from 450 at the top section of each stack to 750 at the bottom section). For the example images of labeled cells, maximum intensity projections (stacks of 150-400 µm) were made with Fiji software. Cell counting was performed with the ImageJ Cell Counter plugin. When doing cell counting, week 1 tdTomato labelled cells were defined as a reference; remaining week 1 cells were the same cells at later time point that align with week 1 reference cells but the not-visible cells at week 1 (the dead cells). Plots of cell counts were made with Origin 7.0 software (OriginLab, Northampton, MA). For the thresholded version of this analysis (Supplementary Figure S2), in order to exclude cells that could possibly have been labeled only with mCherry in the SiR-CRE group, only cells with

fluorescence intensity greater than the average of the mean red fluorescence intensities of cells imaged in Ai35 versus Ai14 mice at the same laser power at 1020 nm at 7 days postinjection (32.33 a.u.) were included in the population of cells tracked from 7 days onward.

Sanger sequencing

RNA viral genomes were extracted from two Tripodi lab (EnvA/ SiR-CRE and EnvA/Sir-FLPo) and one Wickersham lab (RVΔG-4mCherry) rabies virus samples using a Nucleospin RNA kit (Macherey-Nagel, Germany) and treated with DNase I (37°C for 1 hour, followed by 70°C for 5 minutes). Extracted RNA genomes were converted to complementary DNA using an AccuScript PfuUltra II RT-PCR kit (Agilent Technologies, USA) at 42°C for 2 hours with the following barcoded primer annealing to the rabies virus leader sequence:
Adapter_N8_leader_fp: TCAGACGATGCGTCATGCNNNNNNNNACGCTTAACAACCAGATC

cDNA sequences from the leader through the first half of the rabies virus P gene were amplified using Platinum SuperFi Green Master Mix (Invitrogen (Thermo Fisher), USA) with cycling conditions as follows: denaturation at 98°C for 30 seconds, followed by 25 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, using the following primers:
pEX_adapter_fp: CAGCTCAGACGATGCGTCATGC
Barcode2_P_rp: GCAGAGTCATGTATAGCTTCTTGAGCTCTCGGCCAG

The ~2kb PCR amplicons were extracted from an agarose gel, purified with Nucleospin Gel and PCR Clean-up (Macherey-Nagel, Germany), and cloned into pEX-A (Eurofins Genomics, USA) using an In-Fusion HD Cloning Kit (Takara Bio, Japan). The cloned plasmids were transformed into Stellar competent cells (Takara Bio, Japan), and 200 clones per rabies virus sample were isolated and purified for sequencing. For each clone, the index and the 3' end of the N gene were sequenced until sequencing data was collected for over fifty clones per sample: 51 clones from SiR-CRE(EnvA), 50 from SiR-FLPo(EnvA), and 51 from RVΔG-4mCherry(EnvA). Although viral samples may contain plasmid DNA, viral mRNA, and positive-sense anti-genomic RNA, this RT-PCR procedure can amplify only the negative-sense RNA genome: the reverse transcription primer is designed to anneal to the leader sequence of the negative-strand genome so that cDNA synthesis can start from the negative-sense RNA genome, with no other possible templates. Additionally, the PCR amplifies the cDNA, not any plasmids which were transfected into producer cell lines during viral vector production, because the forward PCR primer anneals to the primer used in the reverse transcription, rather than any viral sequence. This RT-PCR protocol ensures that only negative-sense RNA rabies viral genomes can be sequenced.

Sanger sequencing of transgenes in SiR viruses

The procedure for sequencing the transgene inserts was the same as above, but with the RT primer being Adaptor_N8_M_fp (see below), annealing to the M gene and again with a random 8-nucleotide index to tag each clone, and with PCR primers pEX_adaptor_fp (see above) and Barcode2_L_rp (see below), to amplify the sequences from the 3' end of the M gene to the 5' end of the L gene, covering the iCre-P2A-mCherryPEST (or FLPo-P2A-mCherryPEST) sequence.

Primers for RT and PCR for Sanger sequencing were as follows:
Adaptor_N8_M_fp:
TCAGACGATGCGTCATGCNNNNNNNNCAACTCCAACCCTTGGGAGCA
Barcode2_L_rp:
GCAGAGTCATGTATAGTTGGGGACAATGGGGGTTCC

Sanger sequencing analysis

Alignment and mutation detection were performed using SnapGene 4.1.9 (GSL Biotech LLC, USA). Reference sequences of the viral samples used in this study were based on deposited plasmids in Addgene: pSAD-F3-NPEST-iCRE-2A-mCherryPEST (Addgene #99608), pSAD-F3-NPEST-FLPo-2A-mCherryPEST (Addgene #99609), and pRVΔG-4mCherry (Addgene #52488). Traces corresponding to indices and mutations listed in Figure 1 and Supplementary File S2 were also manually inspected and confirmed.

Single-molecule, real-time (SMRT) sequencing

Double-stranded DNA samples for SMRT sequencing were prepared similarly to the above, except that that the clones generated from each of the three virus samples were tagged with one of the standard PacBio barcode sequences to allow identification of each clone's sample of origin 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 (RVΔG-4Cre (6); 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:

RVΔG-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 CCS (circular consensus sequence) read for each molecule. Raw sequences were initially processed using SMRT Link v.6.0 (Pacific Biosciences, USA). Sequences were filtered for a minimum of read length 10 bp, pass 3, and read score 65. 127,178 CCS reads were filtered through passes 3 and Q10; 89,188 CCS reads through passes 5 and Q20; 29,924 CCS reads through passes 8 and Q30. Downstream bioinformatics analysis was performed using BLASR V5.3.2 for the alignment, bcftools v.1.6 for variant calling. Mutations listed in Figure 2 and Supplementary File S3 were also manually inspected and confirmed using Integrative Genomics Viewer 2.3.32 (software.broadinstitute.org/software/igv/). Analysis steps included the following: 1. Exclude CCS reads under 1000 bases, which may have been derived from non-specific reverse transcription or PCR reactions. 2. Classify the CCS reads to the three samples, according to the PacBio barcodes on the 5' ends. 3. For any CCS reads that contain the same 10-nucleotide random index, select only one of them, to avoid double-counting of clones derived from the same cDNA molecule. 4. Align the reads to the corresponding reference sequence (see Supplementary Files S4-S6). 5. Count the number of mutations at each nucleotide position of the reference sequences.

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 Sean Whelan and Ayano Matsushima for helpful discussion and Jun Zhuang, Soumya Chatterjee, and Ali Cetin for helpful discussion and sharing their own results. We thank Stuart Levine, Noelani Kamelamela, and Huiming Ding of the MIT BioMicro Center for assistance with SMRT sequencing and bioinformatic data analysis. Research reported in this publication was supported by BRAIN Initiative awards U01MH106018 (Wickersham), U01MH114829 (Dong), and U19MH114830 (Zeng) from the National Institute of Mental Health.

FIGURE LEGENDS

Legends here are for the main figures only. Legends for Supplementary Information are after the main figure images below.

Figure 1: Single-molecule Sanger sequencing of barcoded viral genomes reveals that most SiR virions have lost the intended C-terminal modification to the nucleoprotein.

(A) Schematic of the RT-PCR workflow. In the reverse transcription (RT) step, the RT primer, containing a random 8-nucleotide sequence, anneals to the 3' rabies virus leader, adding a unique random index to the 5' end of the cDNA molecule corresponding to each individual viral particle's RNA genome. In the PCR step, the forward PCR primer anneals to the RT primer sequence and

the reverse PCR primer anneals within the viral phosphoprotein gene P. Both PCR primers have 15-base sequences homologous to those flanking an insertion site in a plasmid used for sequencing, allowing the amplicons to be cloned into the plasmid using a seamless cloning method before transformation into bacteria. The resulting plasmid library consists of plasmids containing up to 4^8 different index sequences, allowing confirmation that the sequences of plasmids purified from individual picked colonies correspond to the sequences of different individual rabies viral particles' genomes.
(B) Representative Sanger sequencing data of the 8-bp index and the TEV-PEST sequence. Mutations are highlighted in red.
(C) Mutation variants and their frequencies in each viral vector sample based on Sanger sequencing data. No unmutated genomes were found in the SiR-CRE sample: 50 out of 51 had a substitution creating an opal stop codon just before the TEV cleavage site, and the 51st genome contained a frameshift which also removed the C-terminal addition. In the SiR-FLPo sample, only 4 out of 50 clones had an intact sequence of the C-terminal addition; the other 46 out of 50 had one of two *de novo* stop codons at the end of N or the beginning of the TEV cleavage site. In the sample of RVΔG-4mCherry, a virus from our laboratory included as a control to distinguish true mutations on the rabies genomes from mutations due to the RT-PCR process, none of the 51 clones analyzed had mutations in the sequenced region.

Figure 2: Single-molecule, real-time (SMRT) sequencing of thousands of barcoded viral genomes confirms that most SiR virions have lost the intended C-terminal modification to the nucleoprotein.
(A) Schematic of workflow for SMRT sequencing. An RT primer with a random 10-nucleotide sequence anneals to the leader sequence on the negative-sense single-stranded RNA genome. Forward and reverse PCR primers have distinct SMRT barcodes at their 5' ends for the three different virus samples. After RT-PCR, each amplicon library consists of amplicons each containing a SMRT barcode to identify the sample of origin as well as the 10-nucleotide index (i.e., with a potential diversity of 4^{10} different indices) to uniquely label the individual genome of origin. SMRT "dumbbell" adaptors are then ligated to the amplicons' ends, making circular templates which are then repeatedly traversed by a DNA polymerase, resulting in long polymerase reads each containing multiple reads of the positive and negative strands. The individual subreads for a given molecule are combined to form circular consensus sequence (CCS) reads.
(B) High-frequency (>2%) point mutations found in the rabies vector samples based on SMRT sequencing. Horizontal axis represents nucleotide position along the reference sequences (see text); vertical axis represents variant frequency. Total number of CCS3 reads (i.e., with at least 3 subreads for each position) are 22,205 for SiR-CRE, 17,086 reads for SiR-FLPo, and 17,978 reads for RVΔG-4Cre. The great majority of SiR-CRE and SiR-FLPo genomes have point mutations creating premature stop codons at or just after the end of N and before the C-terminal addition. The only frequent (>2%) mutation found in the control virus, RVΔG-4Cre, was a single amino acid substitution at position 419 in 9.49% of virions. Insertions and deletions are not shown 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 appear 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 (6). 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* shows that SiR virus kills 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 disappear 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.

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 RVΔGL-4Cre were lost, RVΔG-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 RVΔG-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: RVΔG-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 RVΔG-4Cre for SMRT sequencing.

Supplementary Figure S3: First-generation vector RVΔG-4FLPo appears to be toxic to most cells, unlike the comparable first-generation vector RVΔG-Cre. Although we did not rigorously quantify the effect, our FLPo-encoding RVΔG-4Flpo appears to kill neurons more quickly than does RVΔG-4Cre (cf. Figure 4 and Chatterjee et al. (6)). 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.

REFERENCES

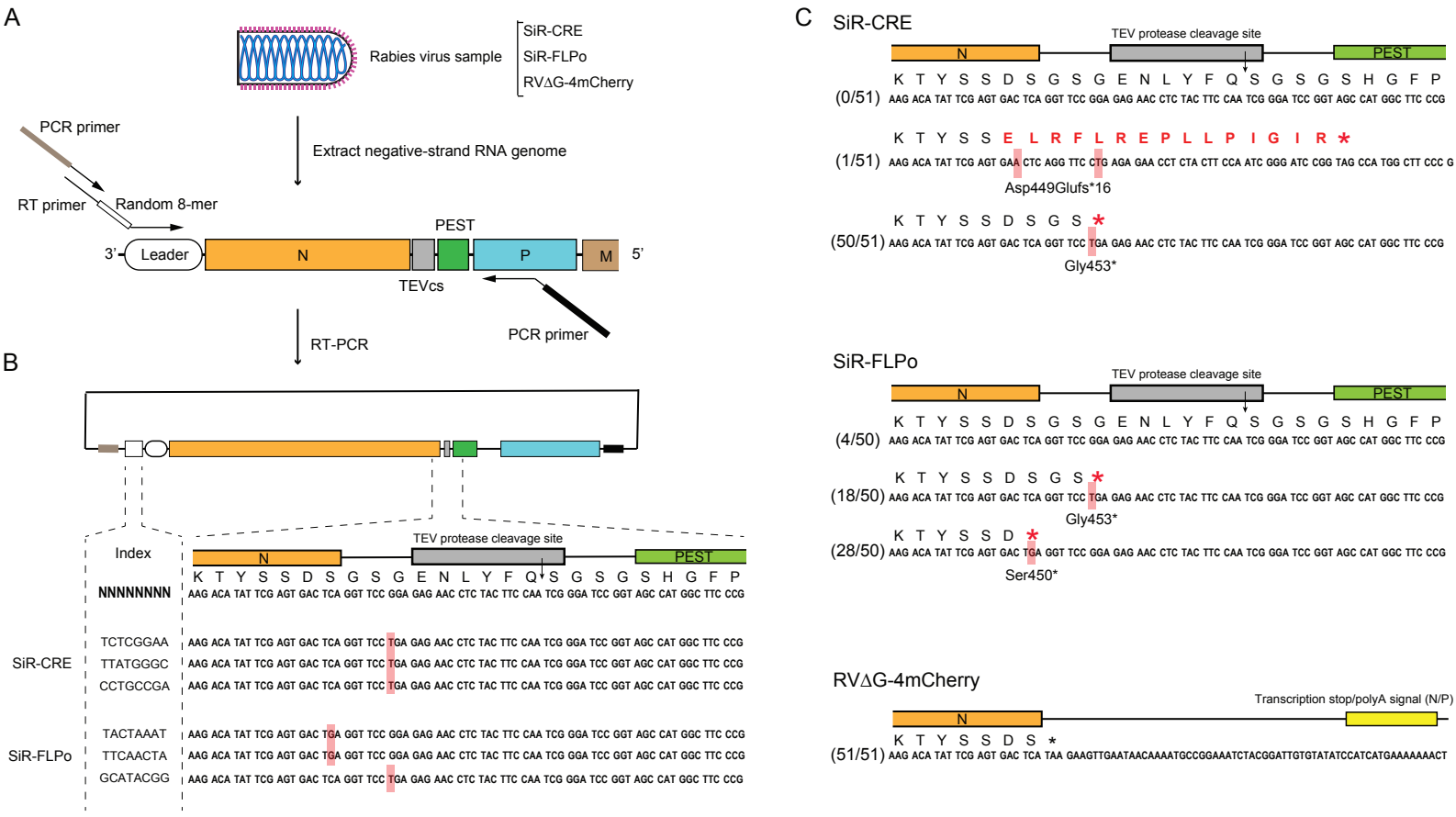
1. Wickersham IR*, et al.* (2007) Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* 53(5):639-647.
2. Augustine V*, et al.* (2018) Hierarchical neural architecture underlying thirst regulation. *Nature* 555(7695):204-209.
3. Kohl J*, et al.* (2018) Functional circuit architecture underlying parental behaviour. *Nature* 556(7701):326-331.
4. Evans DA*, et al.* (2018) A synaptic threshold mechanism for computing escape decisions. *Nature* 558(7711):590-594.
5. Kaelberer MM*, et al.* (2018) A gut-brain neural circuit for nutrient sensory transduction. *Science* 361(6408).
6. Chatterjee S*, et al.* (2018) Nontoxic, double-deletion-mutant rabies viral vectors for retrograde targeting of projection neurons. *Nat Neurosci* 21(4):638-646.
7. Wickersham IR, Finke S, Conzelmann KK, & Callaway EM (2007) Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* 4(1):47-49.
8. Reardon TR*, et al.* (2016) Rabies Virus CVS-N2c(DeltaG) Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron* 89(4):711-724.

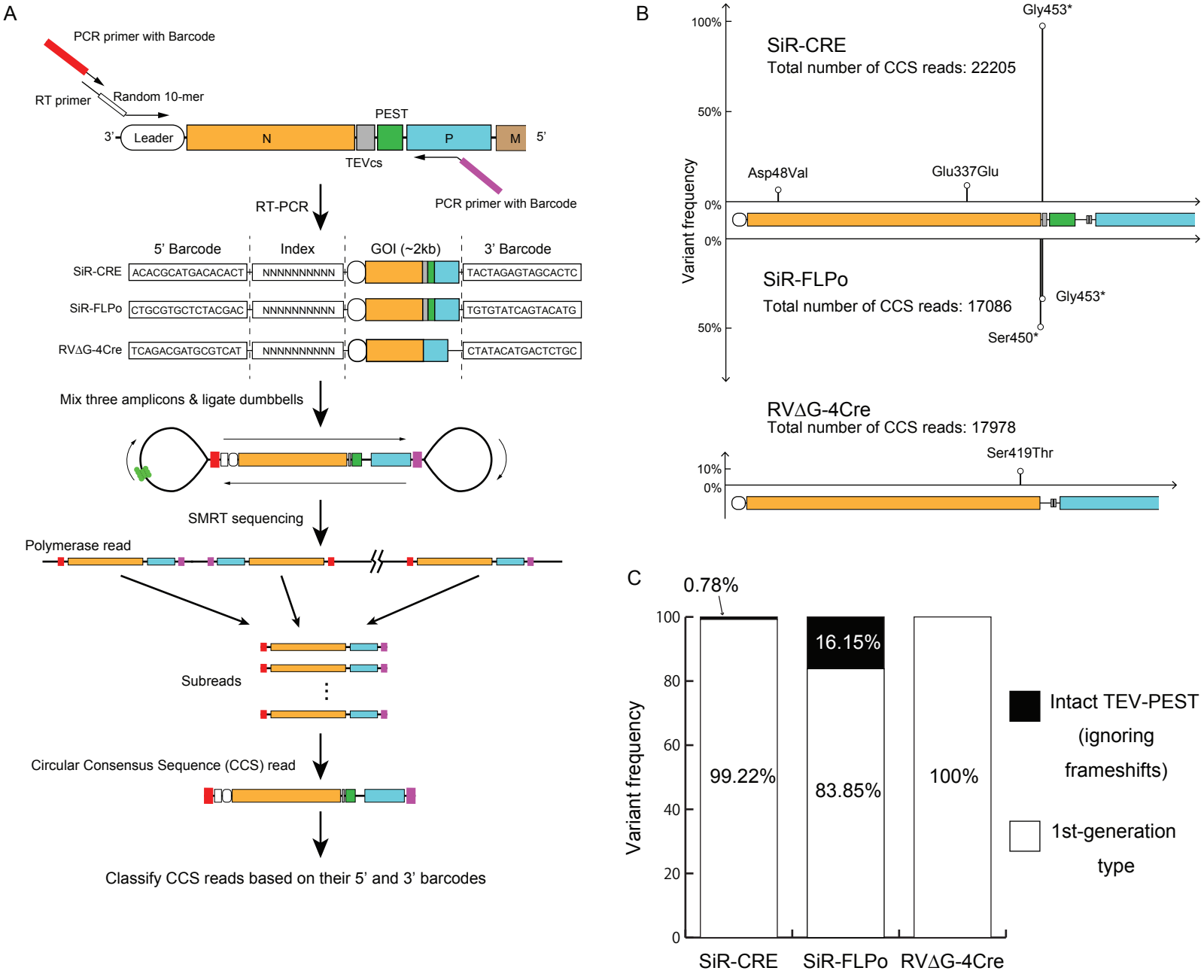
9. Morimoto K, Hooper DC, Spitsin S, Koprowski H, & Dietzschold B (1999) Pathogenicity of different rabies virus variants inversely correlates with apoptosis and rabies virus glycoprotein expression in infected primary neuron cultures. *J Virol* 73(1):510-518.
10. Ciabatti E, Gonzalez-Rueda A, Mariotti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

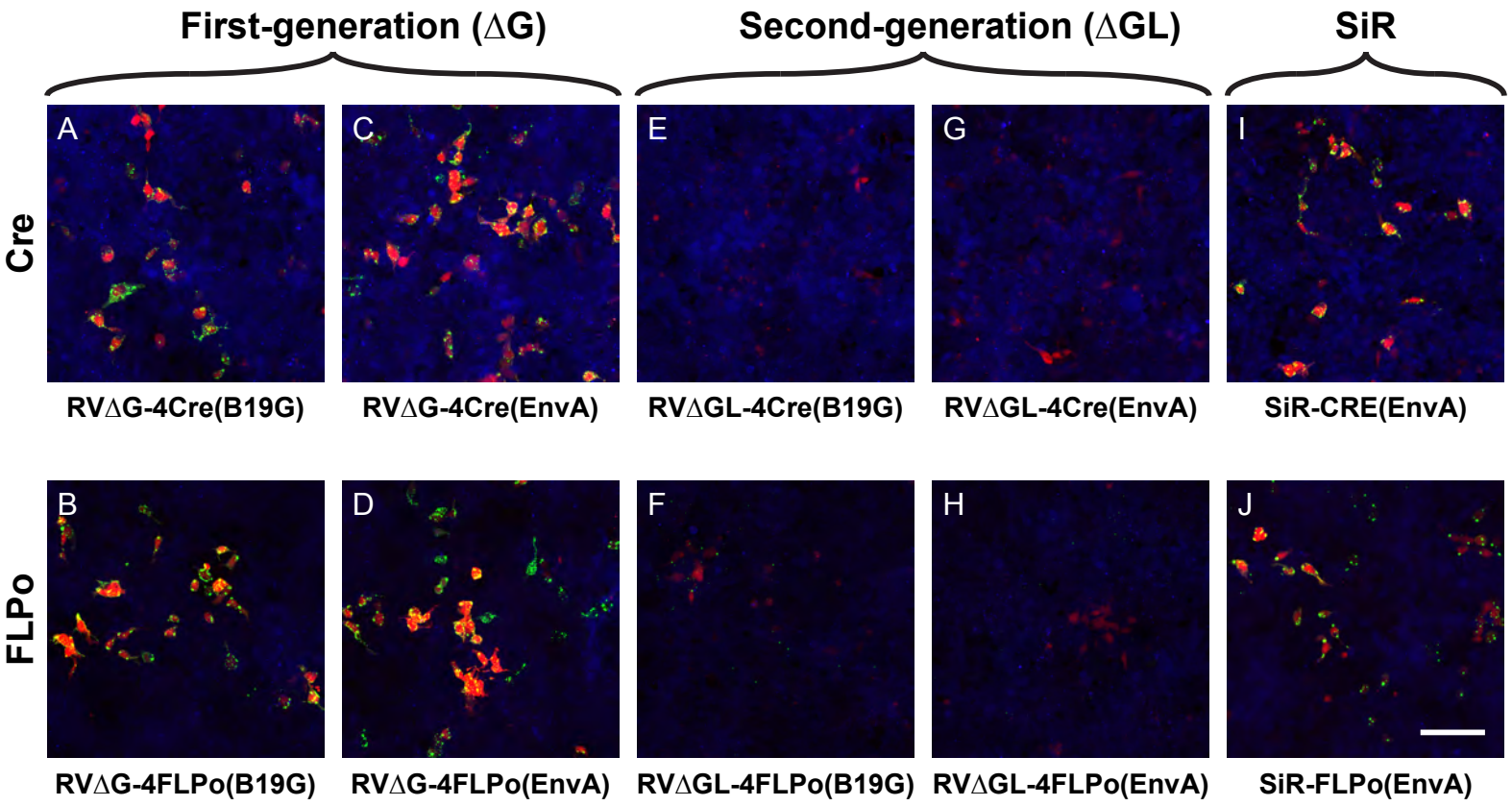
30. Gomme EA, Wirblich C, Addya S, Rall GF, & Schnell MJ (2012) Immune clearance of attenuated rabies virus results in neuronal survival with altered gene expression. *PLoS Pathog* 8(10):e1002971.
31. Wiktor TJ, Dietzschold B, Leamnson RN, & Koprowski H (1977) Induction and biological properties of defective interfering particles of rabies virus. *J Virol* 21(2):626-635.
32. Kawai A & Matsumoto S (1977) Interfering and noninterfering defective particles generated by a rabies small plaque variant virus. *Virology* 76(1):60-71.
33. Clark HF, Parks NF, & Wunner WH (1981) Defective interfering particles of fixed rabies viruses: lack of correlation with attenuation or auto-interference in mice. *J Gen Virol* 52(Pt 2):245-258.
34. Dalton KP & Rose JK (2001) Vesicular stomatitis virus glycoprotein containing the entire green fluorescent protein on its cytoplasmic domain is incorporated efficiently into virus particles. *Virology* 279(2):414-421.
35. Duprex WP, Collins FM, & Rima BK (2002) Modulating the function of the measles virus RNA-dependent RNA polymerase by insertion of green fluorescent protein into the open reading frame. *J Virol* 76(14):7322-7328.
36. Finke S, Brzozka K, & Conzelmann KK (2004) Tracking fluorescence-labeled rabies virus: enhanced green fluorescent protein-tagged phosphoprotein p supports virus gene expression and formation of infectious particles. *J Virol* 78(22):12333-12343.
37. Koser ML*, et al.* (2004) Rabies virus nucleoprotein as a carrier for foreign antigens. *Proc Natl Acad Sci U S A* 101(25):9405-9410.
38. Brown DD*, et al.* (2005) Rational attenuation of a morbillivirus by modulating the activity of the RNA-dependent RNA polymerase. *J Virol* 79(22):14330-14338.
39. Das SC, Nayak D, Zhou Y, & Pattnaik AK (2006) Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells. *J Virol* 80(13):6368-6377.
40. Klingen Y, Conzelmann KK, & Finke S (2008) Double-labeled rabies virus: live tracking of enveloped virus transport. *Journal of virology* 82(1):237-245.
41. Das SC, Panda D, Nayak D, & Pattnaik AK (2009) Biarsenical labeling of vesicular stomatitis virus encoding tetracysteine-tagged m protein allows dynamic imaging of m protein and virus uncoating in infected cells. *J Virol* 83(6):2611-2622.
42. Marriott AC & Hornsey CA (2011) Reverse genetics system for Chandipura virus: tagging the viral matrix protein with green fluorescent protein. *Virus Res* 160(1-2):166-172.
43. Soh TK & Whelan SP (2015) Tracking the Fate of Genetically Distinct Vesicular Stomatitis Virus Matrix Proteins Highlights the Role for Late Domains in Assembly. *J Virol* 89(23):11750-11760.
44. Nikolic J, Civas A, Lama Z, Lagaudriere-Gesbert C, & Blondel D (2016) Rabies Virus Infection Induces the Formation of Stress Granules Closely Connected to the Viral Factories. *PLoS Pathog* 12(10):e1005942.
45. Case JB*, et al.* (2020) Replication-competent vesicular stomatitis virus vaccine vector protects against SARS-CoV-2-mediated pathogenesis. *bioRxiv*.
46. Matsuyama M*, et al.* (2019) “Self-inactivating” rabies viruses are just first-generation, ΔG rabies viruses. *bioRxiv*:550640.
47. Ciabatti E*, et al.* (2020) Genomic stability of Self-inactivating Rabies. *bioRxiv*:2020.2009.2019.304683.
48. Marr RA*, et al.* (2004) Neprilysin regulates amyloid Beta peptide levels. *J Mol Neurosci* 22(1-2):5-11.
49. Niwa H, Yamamura K, & Miyazaki J (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108(2):193-199.
50. 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.

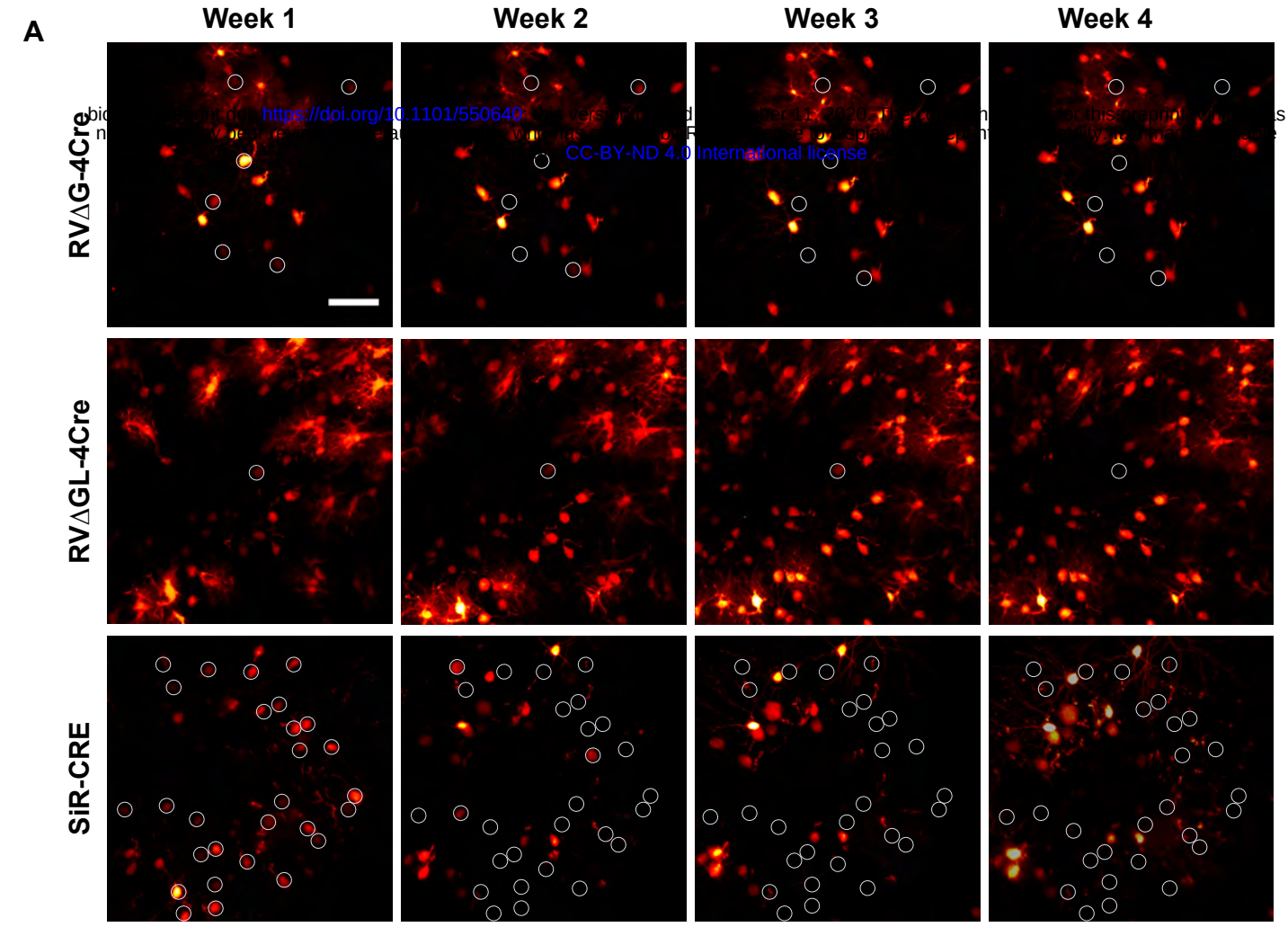
The Journal of neuroscience : the official journal of the Society for Neuroscience 28(28):7025-7030.

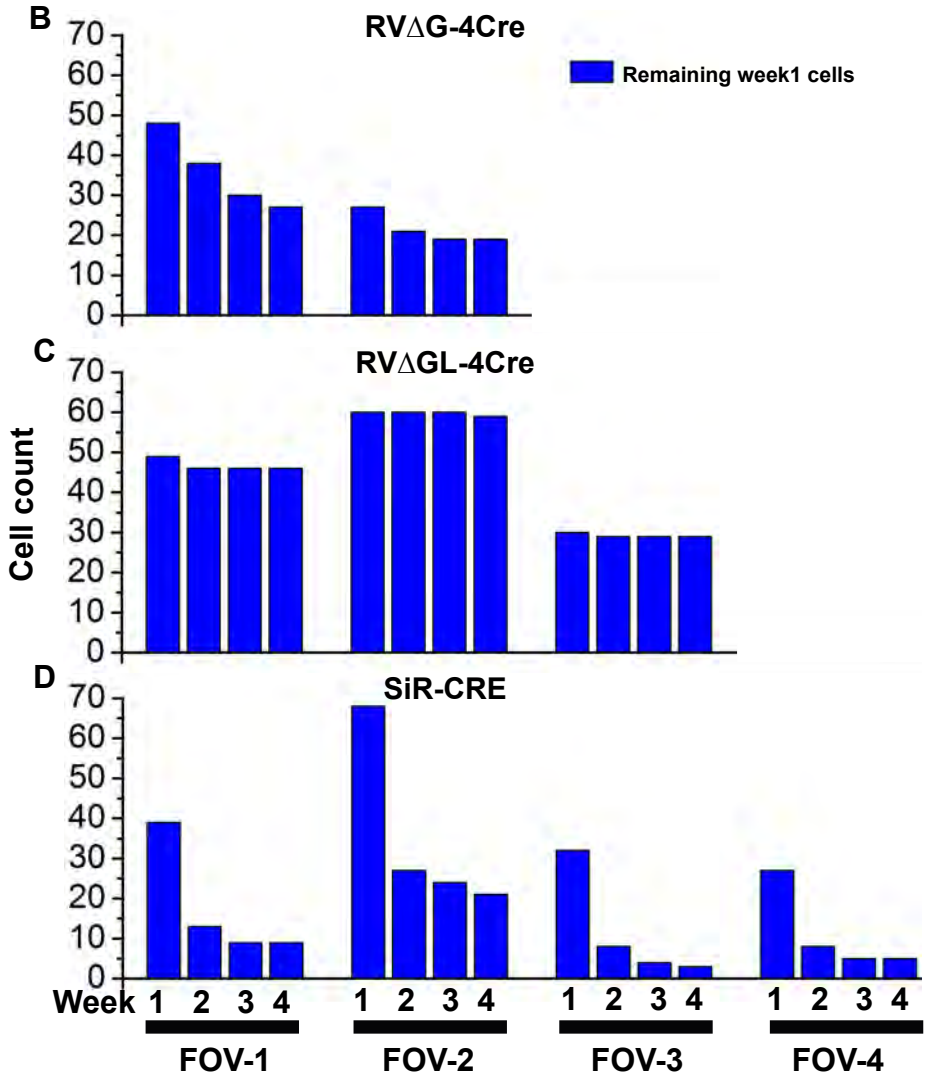
51. Subach OM, Cranfill PJ, Davidson MW, & Verkhusha VV (2011) An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS One* 6(12):e28674.
52. Turan S, Kuehle J, Schambach A, Baum C, & Bode J (2010) Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J Mol Biol* 402(1):52-69.
53. 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.
54. 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.
55. Raymond CS & Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* 2(1):e162.
56. Wickersham IR*, et al.* (2015) Lentiviral vectors for retrograde delivery of recombinases and transactivators. *Cold Spring Harb Protoc* 2015(4):368-374.
57. Sullivan HA & Wickersham IR (2015) Concentration and purification of rabies viral and lentiviral vectors. *Cold Spring Harb Protoc* 2015(4):386-391.
58. 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.

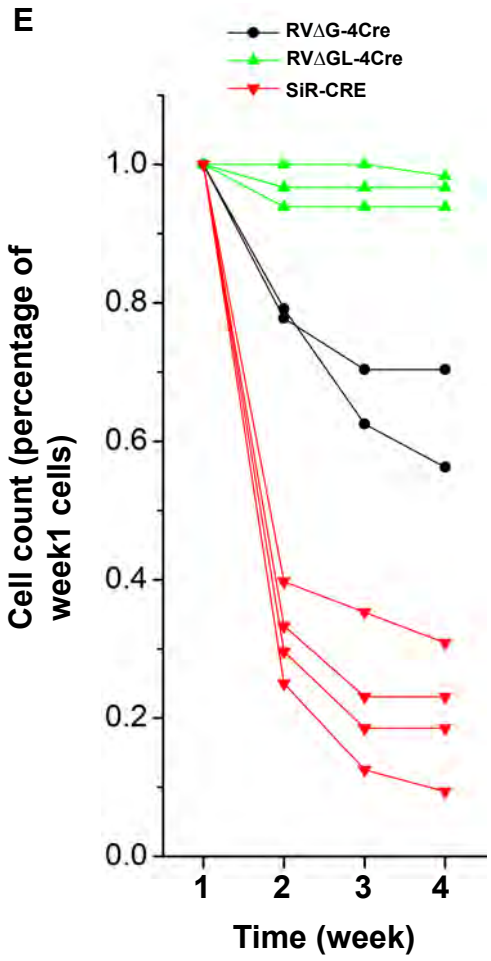


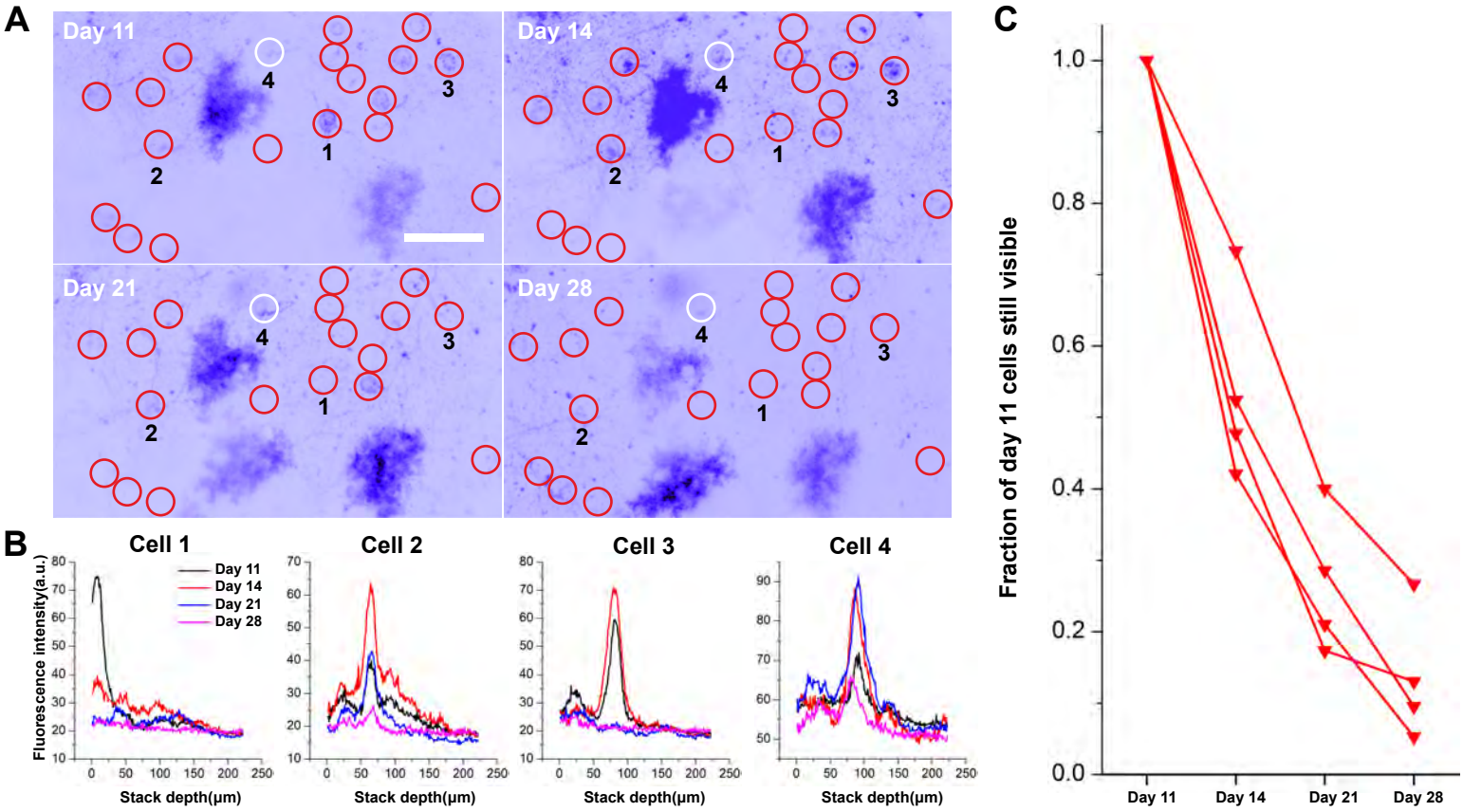


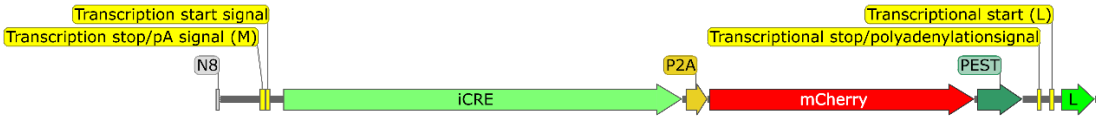












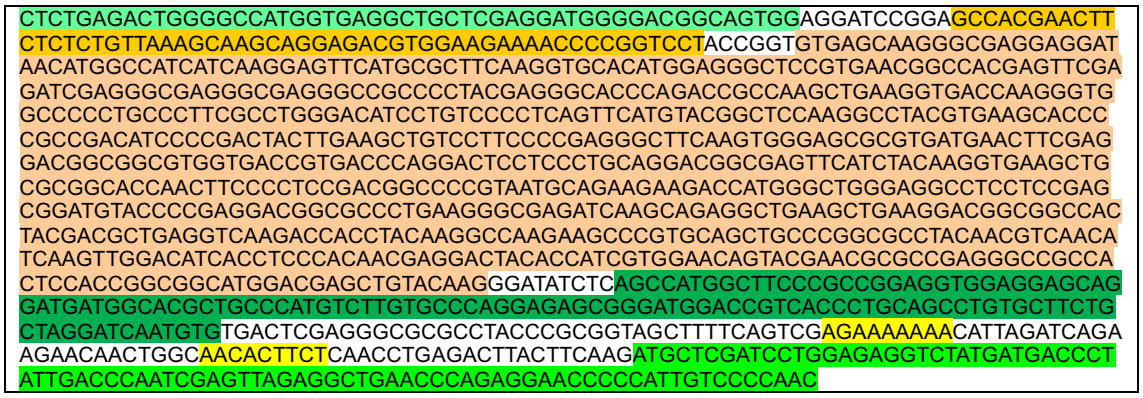
Random 8-nucleotide index
Cre gene, P2A, mCherry gene, PEST, and transcription stop/polyA signal, transcription start signal, L gene, mutation

No mutations in transgene insert of SiR-CRE (19/21 clones)

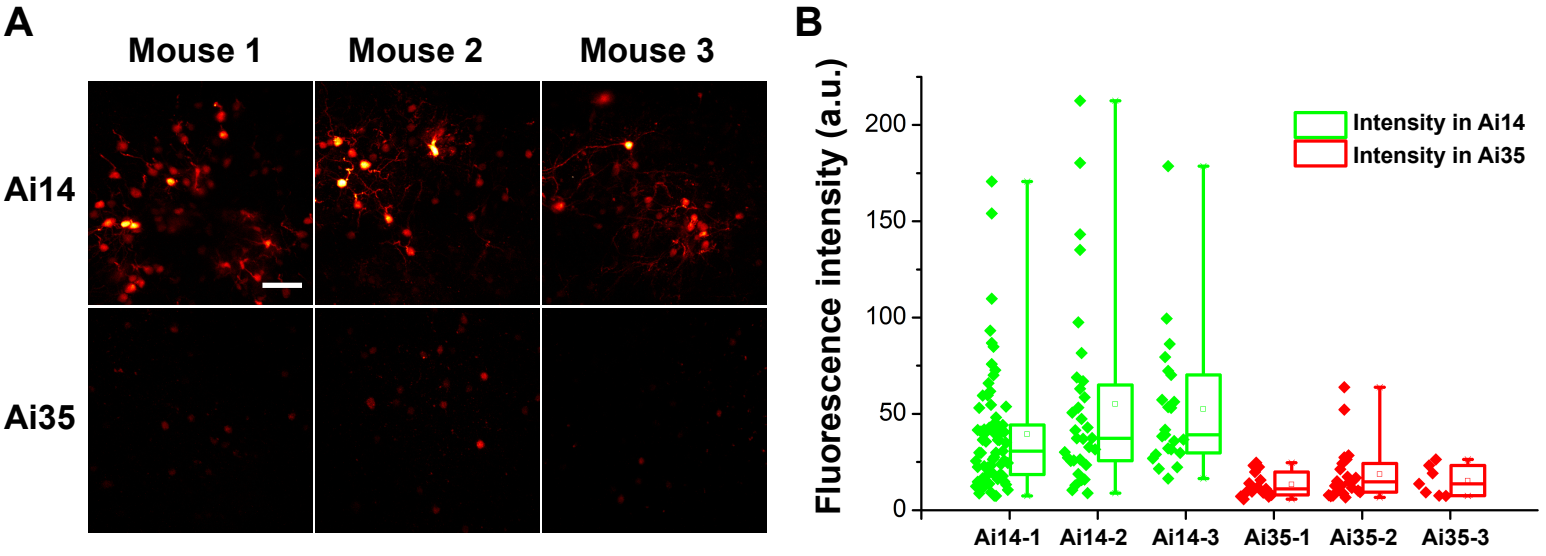
ACTTTCTA, AGCCCAGT, TTATCAGG, AGGGCGTT, AGTGGGCA, GGGCCTTT, TCCCGCAA, GGTCGGGG, GGCCTATA, TGTCTCTC, CCTGTCGG, TAGCACAC, GTATGTGG, TAAGGCCT, CTCACTAT, TTTAGTCC, AACCCGAC, AGCGGCGC, AGCGGGGG
CAACTCCAACCCTTGGGAGCAATATAACAAAAAACATGTTATGGTGCCATTAAACCGCTGCATTTCATCAAAGTCA AGTTGATTACCTTTACATTTTGATCCTCTTGGATGTGAAAAAAACTATTAACATCCCTCAAAGGACCTGCAGGTACG CGGCCGCGGTACCGCCACCATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAAC CTGCCTGCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGGGACAGGCA GGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGGGCTGCCTGGTGCAAGCTGAACA ACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGGGACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGC TGTGAAGACCATCCAACAGCACCTGGGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCT GACTCCAATGCTGTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAAGC AGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTCTGACAGATGCCAGGAC ATCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCAGAATCAGAGT GAAGGACATCTCCCGCACCGATGGTGGGAGAATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACA GCTGGTGTGGAGAAGGCCCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGG CTGATGACCCCAACAACTACCTGTTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGCCACCTCCCAA CTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGATCTATGGTGCCAAGGATGACTCTG GGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTG TGTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGG ACTCTGAGACTGGGGCCATGGTGAGGCTGCTCGAGGATGGGGACGGCAGTGGAGGATCCGGAGCCACGAACT TCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGGTCCTACCGGTGTGAGCAAGGGCGAGGAGGA TAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCG AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGT GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACC CCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA GGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCT GCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGA GCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCA CTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAAC ATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCC ACTCCACCGGCGGCATGGACGAGCTGTACAAGGGATATCTCAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCA GGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCT GCTAGGATCAATGTGTGACTCGAGGGCGCGCCTACCCGCGGTAGCTTTTCAGTCGAGAAAAAAACATTAGATCA GAAGAACAACTGGCAACACTTCTCAACCTGAGACTTACTTCAAGATGCTCGATCCTGGAGAGGTCTATGATGACC CTATTGACCCAATCGAGTTAGAGGCTGAACCCAGAGGAACCCCCATTGTCCCCAAC

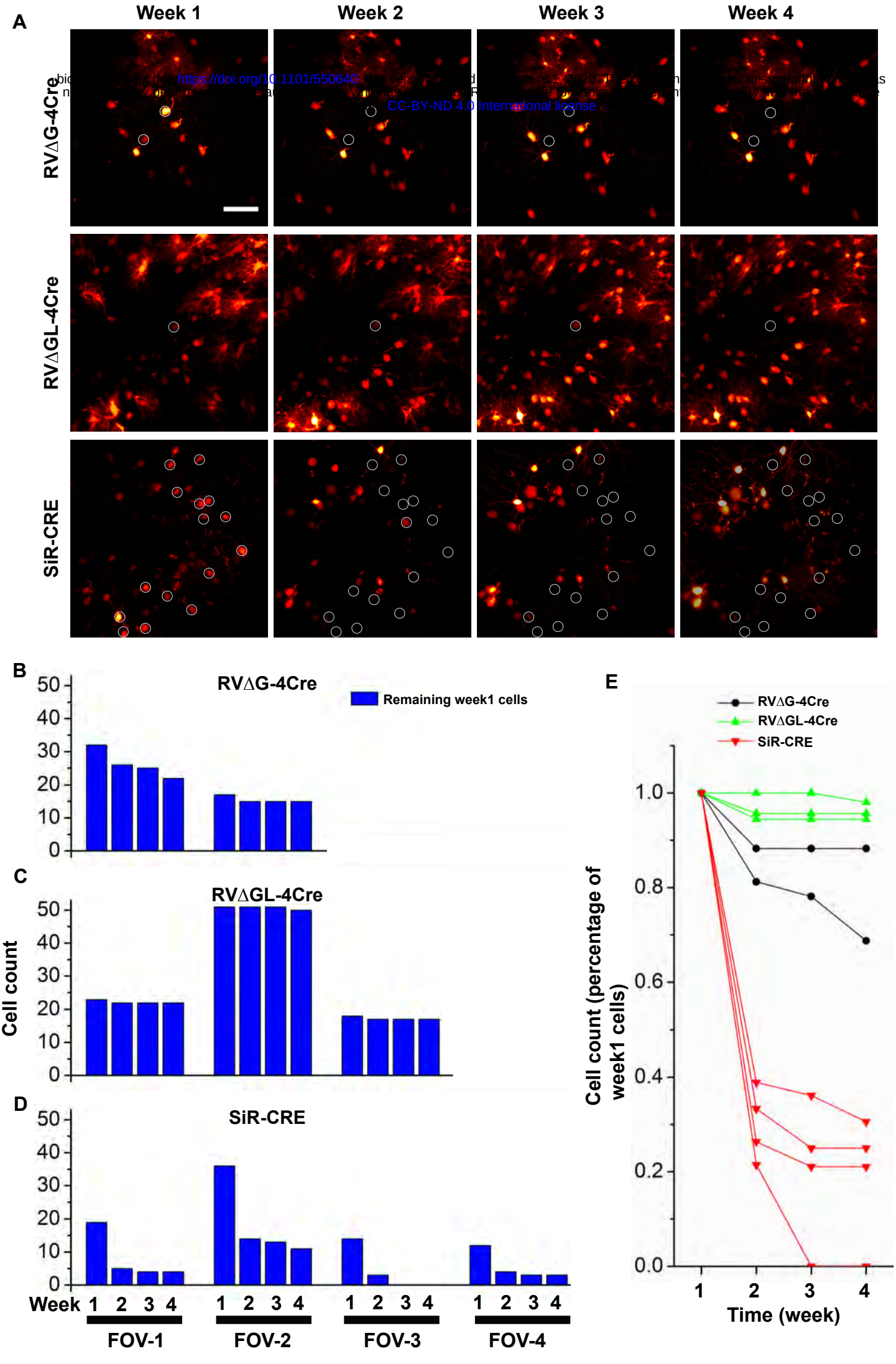
Ala88Val mutation in iCre gene of SiR-CRE (1/21 clones)

AGCTGGTT
CAACTCCAACCCTTGGGAGCAATATAACAAAAAACATGTTATGGTGCCATTAAACCGCTGCATTTCATCAAAGTCA AGTTGATTACCTTTACATTTTGATCCTCTTGGATGTGAAAAAAACTATTAACATCCCTCAAAGGACCTGCAGGTACG CGGCCGCGGTACCGCCACCATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAAC CTGCCTGCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGGGACAGGCA GGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGGGCTGCCTGGTGCAAGCTGAACA ACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGGGACTACCTCCTGTACCTGCAAGTCAGAGGCCTGGCT GTGAAGACCATCCAACAGCACCTGGGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTG ACTCCAATGCTGTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAAGCA GGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTCTGACAGATGCCAGGACA TCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCAGAATCAGAGTG AAGGACATCTCCCGCACCGATGGTGGGAGAATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAG CTGGTGTGGAGAAGGCCCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGC TGATGACCCCAACAACTACCTGTTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGCCACCTCCCAAC TGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGATCTATGGTGCCAAGGATGACTCTGG GCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGT GTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGA



Arg189Ile mutation in iCre gene of SiR-CRE





Random 8-nucleotide index
N gene, tabacco etch virus protease cleavage site, PEST, transcription stop/polyA signal, mutation

Gly453* in N-TEV-PEST of SiR-CRE (50/51 clones)

TCTCGGAA, TTATGGGC, CCTGCCGA, CGTCGTTG, ACTCTAGT, GATCCGGT, ATTATGTA, TATTATCC, ATGCTAGA, GTTGTTTC, GAGTCGAC, GTCGACCT, CAAGAGTA, CGAAGCCA, GCGTGATA, CGTCCAGA, CTTGGTCG, GGGGTTTT, GTCCATAC, GTTTAATA, CAAAATCC, GTGGAGGC, CGGAAGGT, AGGACGGG, CCTCGGGA, GCATTTGC, TGGTTCGG, ATTTAGTC, ATGGGTAA, CTACGGGG, CCTGGTAC, GCATTTCA, AGTGTAGG, CGCATAGA, GTCACAAT, GTATGCTT, TCGGGTTT, ATGTTGTC, CAGCCGTA, GAGTGGCC, GGCCCATT, GGATTTCC, GTACAAGC, GAAGATAT, AAACAGAT, AAGGCCTT, CGTTCGTC, AAAGACTA, ACTGAAAG, TGGGGATA
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGACTCAGGTTCCTGAGAGA ACCTCTACTTCCAATCGGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCT GCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGT AAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

Asp449Glufs*16 in N-TEV-PEST of SiR-CRE (1/51 clones)

AGGCTAAC
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGAACTCAGGTTCCTGAGAG AACCTCTACTTCCAATCGGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGC TGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTG TAAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

No mutations in sequenced region of N-TEV-PEST of SiR-FLPo (4/50 clones)

TCATTTAT, AAGTCGAA, TGAATACA, TATACACG
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGACTCAGGTTCCGGAGAGA ACCTCTACTTCCAATCGGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCT GCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGT AAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

Gly453* in N-TEV-PEST of SiR-FLPo (18/50 clones)

GAGGCCGC, CTTATCAC, TCAGTTTT, CTTCAGCA, TTACTTTT, AAAATAAG, CCATTTCT, GTCACATT, GCATACGG, AGTATTAA, CCGGGTTG, GGTAAAAT, CAGGGAGA, CATTTTTC, TGACAAAT, TCGTATAT, TAGGGATT, GATCCCCG
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGACTCAGGTTCCTGAGAGA ACCTCTACTTCCAATCGGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCT GCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGT AAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

Ser450* in N-TEV-PEST of SiR-FLPo (28/50 clones)

TACTAAAT, TTCAACTA, TAATAGGC, CGCGGGTC, AGGGTGCC, TTGCTGAC, CGGCCCTT, TACAGCGA, GAGTAGGA, GTAAACGC, CTATGGGT, TTGATTGA, AGGATTCT, GAGGTTTG, TACTTAGC, CCTGTAGT, GGGGTGGA, TTAAGTTG, CTCAATGA, CAGATCTG, GTCTGGTA, GCGTTGGG, TAAACGAG, TACTTGAG, CGGGGTGT, CGATCCCG, TGGGTGTC, GCCACGAT
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGACTGAGGTTCCGGAGAGA ACCTCTACTTCCAATCGGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCT GCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGT AAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

No mutations in sequenced region of RVΔG-4mCherry (51/51 clones)

CGCCCGCG, GCCAATCT, GGCGGAAT, TGTGGAAC, GGAGGGAT, CGTAGTGT, AGAATCTC, TGTCTGGC, GCCTTTTA, TGGAATCT, TTGTAATG, AGGCCTGT, CGGATATA, AATCCAAA, AAGGAAAA, GAACTCAT, GCCGCTTC, CTTTGCCG, CGCAACCT, TCGGGCAT, GGGCGACT, AACTGGAT, TGCGTCGG, TCAAAGCG, CGAGGCCC, ATGTAGGA, AGATACGT, TGGCTTCG, ATTTTCTA, TCGTCCGG, TCGGAGCG, CTGTTATA, TACCGTTC, CGAATGTC, CCCTCTTT, TGAGTGGG, TGTAAATA, AAGGAGTC, ATTACCCT, TCCCTGAC, TGGTAAAA, ACTATCTC, CCGGAAGC, CTTGGAGG, GCAACAAT, ATTACCGT, AATAGCAG, ACCATAGT, GGGTTGGT, CATTTATT, GGGTACAC
ATCATCAAGCCCGTCCAAACTCATTCGCCGAGTTTCTAAACAAGACATATTCGAGTGACTCATAAGAAGTTGAATA ACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAA

SINGLE-MOLECULE, REAL-TIME SEQUENCING RESULTS

Frameshifts and insertions

Position numbers in this file refer to the reference sequences included as Supplementary Files S4-S5. A "frameshift" is included in Tables 1a-2b if the number of deleted bases in positions 1439-1492 (the vicinity of the junction of the end of the N gene and the intended 3' addition) is not an integer multiple of 3, with insertions ignored. "Any error" includes either the apparent frameshifts, or the new TAA/TAG/TGA stop codons, or both, with insertions ignored. The number of “frameshifts” increases considerably if insertion mutations are included in the calculation, indicating that there is a much higher insertion rate as compared to that of deletion; however, previous studies have found that spurious insertions are high with SMRT (see main text), so we ignore insertions in this paper apart from summarizing the data below.

SiR-CRE				
Position	**Mutation**	**CCS3**	**CCS5**	**CCS8**
# Sequences		22205	866	239
243	**GAT (reference)**	20273	784	209
	GCT	3	0	0
	GGT	5	0	0
	GTT	1856	79	29
	DEL	68	3	1
1111	**GAA (reference)**	19807	760	197
	GAC	1	0	0
	GAG	2396	105	42
	GAT	0	0	0
	DEL	1	1	0
1457	AGA	6	0	0
	CGA	1	0	0
	GGA (reference)	140	5	1
	TGA	22032	858	237
	DEL	26	3	1
[1439, 1492]	Frame shift (#DELs not an integer multiple of 3, Insertion ignored)	507	13	3
[1439, 1492]	Any error (insertion ignored)	22104	864	239
% mutation in [1439, 1492]		100%	100%	100%

Table 1a. All mutations in the SiR-CRE sample at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8), as well as frameshift mutations found in the C-terminal region of N.

SiR-CRE				
Position	**Mutation**	**CCS3**	**CCS5**	**CC8**
[1439, 1492]	Number of frame shifts due to DEL (number of DELs not an integer multiple of 3, insertion ignored)	507	13	3
[1439, 1492]	Number of frame shifts due to either DEL or INS or both (Sequence length not an integer multiple of 3)	2504	89	14
[1439, 1492]	Any error (insertion included)	22174	865	239

Table 1b. Frameshift mutations in the C-terminal region of N in the SiR-CRE sample at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

SiR-FLPo				
Position	**Mutation**	**CCS3**	**CCS5**	**CCS8**
#Sequences		17086	695	210
1449	TAA	28	1	0
	TCA (reference)	8405	333	94
	TGA	8624	360	115
	TTA	3	0	0
	DEL	26	1	1
1457	AGA	3	0	0
	CGA	0	0	0
	GGA (reference)	11088	462	139
	TGA	5979	233	71
	DEL	16	0	0
[1439, 1492]	Frameshifts (#DELs not an integer multiple of 3, Insertion ignored)	448	11	3
[1439, 1492]	Any error (insertion ignored)	14444	589	180
% of mutation		**85%**	**85%**	**86%**

Table 2a. All mutations in the SiR-FLPo sample at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8), as well as frameshift mutations found in the C-terminal region of N.

SiR-FLPo				
Position	**Mutation**	**CCS3**	**CCS5**	**CCS8**
[1439, 1492]	Number of frame shifts due to DEL (number of DELs not an integer multiple of 3, insertion ignored)	448	11	3
[1439, 1492]	Number of frame shifts due to either DEL or INS or both (Sequence length not an integer multiple of 3)	3562	121	24
[1439, 1492]	Any error (insertion included)	15818	642	196

Table 2b. Frameshift mutations in the C-terminal region of N in the SiR-FLPo sample at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

RVΔG-4Cre				
Position	**Mutation**	**CCS3**	**CCS5**	**CC8**
# Sequences		17978	757	254
1355	ACT	1706	84	28
	CCT	3	0	0
	GCT	1	1	1
	TCT (reference)	16139	667	224
	DEL	129	5	1
Total				

Table 3. All mutations in the SiR-FLPo sample at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

Tables of all mutations above 2% frequency threshold

Table 4a to 4c list all single-nucleotide substitutions and deletions at positions mutated at >2% threshold frequency. The percentage of mutations is calculated based on the total number of single nucleotide and deletion mutations divided by the total number of reads aligned, when insertion mutations are ignored. Deletion mutations dominate in the medium-frequency range between 2% and 5%.

	Reference	Position	A	C	G	T	DEL	Un-mutated	Mutated (SNP/DEL)	% Mutation (SNP/DEL)
CCS3	G	1457	6	1	140	22032	26	140	22065	99.4
	A	1111	19807	1	2396	0	1	19807	2398	10.8
	A	243	20273	3	5	1856	68	20273	1932	8.7
	T	615	4	0	2	21229	970	21229	976	4.4
	G	1124	5	0	21356	7	837	21356	849	3.8
	A	81	21358	9	2	0	836	21358	847	3.8
	A	31	21381	0	0	0	824	21381	824	3.7
	T	713	1	2	4	21487	711	21487	718	3.2
	T	338	0	0	9	21498	698	21498	707	3.2
	A	1665	21544	1	4	0	656	21544	661	3.0
	A	53	21606	0	0	1	598	21606	599	2.7
	C	411	8	21683	0	1	513	21683	522	2.4
	A	838	21718	1	0	1	485	21718	487	2.2
CCS5	G	1457	0	0	5	858	3	5	861	99.4
	A	1111	760	0	105	0	1	760	106	12.2
	A	243	784	0	0	79	3	784	82	9.5
	A	31	828	0	0	0	38	828	38	4.4
	G	1124	0	0	831	0	35	831	35	4.0
	T	615	0	0	0	834	32	834	32	3.7
	A	81	837	1	1	0	27	837	29	3.3
	T	338	0	0	1	840	25	840	26	3.0
	T	713	0	0	1	841	24	841	25	2.9
	C	411	0	843	0	0	23	843	23	2.7
	A	53	846	0	0	0	20	846	20	2.3
CCS8	G	1457	0	0	1	237	1	1	238	99.6
	A	1111	197	0	42	0	0	197	42	17.6
	A	243	209	0	0	29	1	209	30	12.6
	A	31	228	0	0	0	11	228	11	4.6
	T	615	0	0	0	232	7	232	7	2.9
	G	1124	0	0	232	0	7	232	7	2.9
	C	411	0	233	0	0	6	233	6	2.5

Table 4a. SiR-CRE: substitutions and deletions at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

	Reference	Position	A	C	G	T	DEL	Un-mutated	Mutated (SNP/DEL)	% Mutation (SNP/DEL)
CCS3	C	1449	28	8405	8624	3	26	8405	8681	50.8
	G	1457	3	0	11088	5979	16	11088	5998	35.1
	T	615	4	0	0	16322	760	16322	764	4.5
	A	81	16410	4	2	5	665	16410	676	4.0
	G	1124	0	0	16410	3	673	16410	676	4.0
	T	713	1	0	2	16500	583	16500	586	3.4
	A	1665	16536	1	1	0	548	16536	550	3.2
	A	31	16547	0	0	0	539	16547	539	3.2
	T	338	0	1	4	16582	499	16582	504	2.9
	A	53	16672	2	2	0	410	16672	414	2.4
	C	411	6	16683	1	0	396	16683	403	2.4
	A	838	16696	0	0	0	390	16696	390	2.3
	G	208	301	0	16739	44	2	16739	347	2.0
CCS5	C	1449	1	333	360	0	1	333	362	52.1
	G	1457	0	0	462	233	0	462	233	33.5
	A	81	665	0	0	0	30	665	30	4.3
	T	615	1	0	0	668	26	668	27	3.9
	A	1665	669	0	1	0	25	669	26	3.7
	T	713	0	0	0	671	24	671	24	3.5
	T	338	0	0	0	673	22	673	22	3.2
	C	411	0	676	0	0	19	676	19	2.7
	G	1124	0	0	676	0	19	676	19	2.7
	A	838	679	0	0	0	16	679	16	2.3
	G	346	0	0	680	0	15	680	15	2.2
	A	31	681	0	0	0	14	681	14	2.0
CCS8	C	1449	0	94	115	0	1	94	116	55.2
	G	1457	0	0	139	71	0	139	71	33.8
	T	713	0	0	0	203	7	203	7	3.3
	A	53	205	0	0	0	5	205	5	2.4
	A	81	205	0	0	0	5	205	5	2.4

Table 4b. SiR-FLPo: substitutions and deletions at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

.

	Reference	Position	A	C	G	T	DEL	Un-mutated	Mutated (SNP/DEL)	% Mutation (SNP/DEL)
CCS3	T	1355	1706	3	1	16139	129	16139	1839	10.2
	A	31	16732	0	0	0	1246	16732	1246	6.9
	T	615	2	0	1	17266	709	17266	712	4.0
	A	81	17300	4	2	1	671	17300	678	3.8
	G	1124	0	0	17379	4	595	17379	599	3.3
	T	713	0	1	1	17469	507	17469	509	2.8
	T	338	1	0	2	17503	472	17503	475	2.6
	A	1506	17520	0	0	0	458	17520	458	2.5
	A	53	17579	1	2	1	395	17579	399	2.2
CCS5	T	1355	84	0	1	667	5	667	90	11.9
	A	31	703	0	0	0	54	703	54	7.1
	A	81	717	0	0	0	40	717	40	5.3
	T	615	0	0	0	729	28	729	28	3.7
	G	1124	0	0	738	0	19	738	19	2.5
	A	53	739	0	0	0	18	739	18	2.4
	T	713	0	0	0	740	17	740	17	2.2
	A	1506	740	0	0	0	17	740	17	2.2
CCS8	T	1355	28	0	1	224	1	224	30	11.8
	A	31	241	0	0	0	13	241	13	5.1
	T	615	0	0	0	248	6	248	6	2.4

Table 4c. RVΔG-4Cre: substitutions and deletions at positions mutated at >2% frequency at all stringencies (CCS3, CCS5, CCS8).

```
LOCUS       Exported                2183 bp ds-DNA     linear   SYN
07-JAN-2019
DEFINITION  synthetic linear DNA.
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 2183)
  AUTHORS   Trial User
  TITLE     Direct Submission
  JOURNAL   Exported Jan 7, 2019 from SnapGene 4.1.9
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..2183
                     /organism="PCR product"
                     /mol_type="other DNA"
     primer_bind     1..48
                     /label=Barcode5_cagc_N10_leader_fp
     primer_bind     1..20
                     /label=Barcode5_cagc_fp
     gap             21..30
                     /estimated_length=10
     5'UTR           31..100
                     /label=RV leader
                     /note="SPBN leader"
     CDS             101..1450
                     /codon_start=1
                     /locus_tag="N"
                     /label=N
                     /note="N"
                     /
translation="MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPCITLGK

APDLNKAYKSVLSGMSAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIARKGDKIT

PGSLVEIKRTDVEGNWALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKT

NIADRIEQIFETAPFVKIVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYS

AIRVGTVVTAYEDCSGLVSFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAV

PHSYFIHFRSLGLSGKSPYSSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSV

LGGYLGEEFFGKGTFERRFFRDEKELQEYEAAELTKTDVALADDGTVNSDDEDYFSGET

RSPEAVYTRIMMNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYSSDS"
     CDS             1460..1480
                     /codon_start=1
                     /locus_tag="TEV site"
```

```
                     /label=TEV site
                     /note="TEV site"
                     /translation="ENLYFQS"
     CDS             1490..1609
                     /codon_start=1
                     /locus_tag="ECFP destabilized by fusion to
residues 422-461
                     of mouse ornithine decarboxylase, giving an in
vivo(1)"
                     /product="ECFP destabilized by fusion to residues
422-461
                     of mouse ornithine decarboxylase, giving an in
vivo
                     half-life of ~2 hours"
                     /label=ECFP destabilized by fusion to residues
422-461
                     /label=ECFP destabilized by fusion to residues
422-461...
                     /label=ECFP destabilized by fusion to residues
422-461 of
                     mouse ornithine decarboxylase, giving an in
vivo(1)
                     /note="destabilization domain"
                     /note="mammalian codon-optimized"
                     /
translation="SHGFPPEVEEQDDGTLPMSCAQESGMDRHPAACASARINV"
     CDS             1703..2167
                     /codon_start=1
                     /locus_tag="P"
                     /label=P
                     /note="P"
                     /
translation="MSKIFVNPSAIRAGLADLEMAEETVDLINRNIEDNQAHLQGEPIE

VDNLPEDMGRLHLDDGKSPNHGEIAKVGEGKYREDFQMDEGEDPSFLFQSYLENVGVQI

VRQMRSGERFLKIWSQTVEEIISYVAVNFPNPPGKSSEDKSTQTTGRELKK"
     primer_bind     complement(2148..2183)
                     /label=Barcode3_P_rp
ORIGIN
        1 acacgcatga cacactcagc nnnnnnnnnn acgcttaaca accagatcaa
agaaaaaaca
       61 gacattgtca attgcaaagc aaaaatgtaa cacccctaca atggatgccg
acaagattgt
      121 attcaaagtc aataatcagg tggtctcttt gaagcctgag attatcgtgg
atcaatatga
      181 gtacaagtac cctgccatca aagatttgaa aaagccctgt ataaccctag
gaaaggctcc
      241 cgatttaaat aaagcataca agtcagtttt gtcaggcatg agcgccgcca
aacttaatcc
```

```
      301 tgacgatgta tgttcctatt tggcagcggc aatgcagttt tttgagggga
catgtccgga
      361 agactggacc agctatggaa ttgtgattgc acgaaaagga gataagatca
ccccaggttc
      421 tctggtggag ataaaacgta ctgatgtaga agggaattgg gctctgacag
gaggcatgga
      481 actgacaaga gaccccactg tccctgagca tgcgtcctta gtcggtcttc
tcttgagtct
      541 gtataggttg agcaaaatat ccgggcaaaa cactggtaac tataagacaa
acattgcaga
      601 caggatagag cagatttttg agacagcccc ttttgttaaa atcgtggaac
accatactct
      661 aatgacaact cacaaaatgt gtgctaattg gagtactata ccaaacttca
gatttttggc
      721 cggaacctat gacatgtttt tctcccggat tgagcatcta tattcagcaa
tcagagtggg
      781 cacagttgtc actgcttatg aagactgttc aggactggta tcatttactg
ggttcataaa
      841 acaaatcaat ctcaccgcta gagaggcaat actatatttc ttccacaaga
actttgagga
      901 agagataaga agaatgtttg agccagggca ggagacagct gttcctcact
cttatttcat
      961 ccacttccgt tcactaggct tgagtgggaa atctccttat tcatcaaatg
ctgttggtca
     1021 cgtgttcaat ctcattcact ttgtaggatg ctatatgggt caagtcagat
ccctaaatgc
     1081 aacggttatt gctgcatgtg ctcctcatga aatgtctgtt ctaggggct
atctgggaga
     1141 ggaattcttc gggaaaggga catttgaaag aagattcttc agagatgaga
aagaacttca
     1201 agaatacgag gcggctgaac tgacaaagac tgacgtagca ctggcagatg
atggaactgt
     1261 caactctgac gacgaggact acttttcagg tgaaaccaga agtccggagg
ctgtttatac
     1321 tcgaatcatg atgaatggag gtcgactaaa gagatctcac atacggagat
atgtctcagt
     1381 cagttccaat catcaagccc gtccaaactc attcgccgag tttctaaaca
agacatattc
     1441 gagtgactca ggttccggag agaacctcta cttccaatcg ggatccggta
gccatggctt
     1501 cccgccggag gtggaggagc aggatgatgg cacgctgccc atgtcttgtg
cccaggagag
     1561 cgggatggac cgtcaccctg cagcctgtgc ttctgctagg atcaatgtgt
aagaagttga
     1621 ataacaaaat gccggaaatc tacggattgt gtatatccat catgaaaaaa
actaacaccc
     1681 ctcctttcga accatcccaa acatgagcaa gatctttgtc aatcctagtg
ctattagagc
     1741 cggtctggcc gatcttgaga tggctgaaga aactgttgat ctgatcaata
gaaatatcga
```

```
     1801 agacaatcag gctcatctcc aaggggaacc catagaggtg gacaatctcc
ctgaggatat
     1861 ggggcgactt cacctggatg atggaaaatc gcccaaccat ggtgagatag
ccaaggtggg
     1921 agaaggcaag tatcgagagg actttcagat ggatgaagga gaggatccta
gcttcctgtt
     1981 ccagtcatac ctggaaaatg ttggagtcca aatagtcaga caaatgaggt
caggagagag
     2041 atttctcaag atatggtcac agaccgtaga agagattata tcctatgtcg
cggtcaactt
     2101 tcccaaccct ccaggaaagt cttcagagga taaatcaacc cagactactg
gccgagagct
     2161 caagaagtac tagagtagca ctc
//
```

```
LOCUS       Exported                2183 bp ds-DNA     linear   SYN
07-JAN-2019
DEFINITION  synthetic linear DNA.
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 2183)
  AUTHORS   Trial User
  TITLE     Direct Submission
  JOURNAL   Exported Jan 7, 2019 from SnapGene 4.1.9
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..2183
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     primer_bind     1..48
                     /label=Barcode9_cagc_N10_leader_fp
     primer_bind     1..20
                     /label=Barcode9_cagc_fp
     gap             21..30
                     /estimated_length=10
     5'UTR           31..100
                     /label=RV leader
                     /note="SPBN leader"
     misc_feature    89..97
                     /label=Transcription start signal (N)
     CDS             101..1450
                     /codon_start=1
                     /locus_tag="N"
                     /label=N
                     /note="N"
                     /
translation="MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPCITLGK

APDLNKAYKSVLSGMSAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIARKGDKIT

PGSLVEIKRTDVEGNWALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKT

NIADRIEQIFETAPFVKIVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYS

AIRVGTVVTAYEDCSGLVSFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAV

PHSYFIHFRSLGLSGKSPYSSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSV

LGGYLGEEFFGKGTFERRFFRDEKELQEYEAAELTKTDVALADDGTVNSDDEDYFSGET

RSPEAVYTRIMMNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYSSDS"
     CDS             1460..1480
```

```
                     /codon_start=1
                     /product="tobacco etch virus (TEV) protease
recognition and
                     cleavage site"
                     /label=TEV site
                     /translation="ENLYFQS"
     CDS             1490..1609
                     /codon_start=1
                     /locus_tag="ECFP destabilized by fusion to
residues 422-461
                     of mouse ornithine decarboxylase, giving an in
vivo(1)"
                     /product="ECFP destabilized by fusion to residues
422-461
                     of mouse ornithine decarboxylase, giving an in
vivo
                     half-life of ~2 hours"
                     /label=PEST
                     /label=ECFP destabilized by fusion to residues
422-461 of
                     mouse ornithine decarboxylase, giving an in
vivo(1)
                     /note="destabilization domain"
                     /note="mammalian codon-optimized"
                     /
translation="SHGFPPEVEEQDDGTLPMSCAQESGMDRHPAACASARINV"
     misc_feature    1661..1671
                     /label=Transcription stop/pA signal (N/P & P/M))
     misc_feature    1674..1682
                     /label=Transcription start signal (P)
     misc_feature    1703..2167
                     /locus_tag="P"
                     /label=P
                     /note="P"
     primer_bind     complement(2148..2183)
                     /label=Barcode4_P_rp
ORIGIN
        1 ctgcgtgctc tacgaccagc nnnnnnnnnn acgcttaaca accagatcaa
agaaaaaaca
       61 gacattgtca attgcaaagc aaaaatgtaa cacccctaca atggatgccg
acaagattgt
      121 attcaaagtc aataatcagg tggtctcttt gaagcctgag attatcgtgg
atcaatatga
      181 gtacaagtac cctgccatca aagatttgaa aaagccctgt ataaccctag
gaaaggctcc
      241 cgatttaaat aaagcataca agtcagtttt gtcaggcatg agcgccgcca
aacttaatcc
      301 tgacgatgta tgttcctatt tggcagcggc aatgcagttt tttgagggga
catgtccgga
      361 agactggacc agctatggaa ttgtgattgc acgaaaagga gataagatca
```

```
ccccaggttc
      421 tctggtggag ataaaacgta ctgatgtaga agggaattgg gctctgacag
gaggcatgga
      481 actgacaaga gaccccactg tccctgagca tgcgtcctta gtcggtcttc
tcttgagtct
      541 gtataggttg agcaaaatat ccgggcaaaa cactggtaac tataagacaa
acattgcaga
      601 caggatagag cagatttttg agacagcccc ttttgttaaa atcgtggaac
accatactct
      661 aatgacaact cacaaaatgt gtgctaattg gagtactata ccaaacttca
gatttttggc
      721 cggaacctat gacatgtttt tctcccggat tgagcatcta tattcagcaa
tcagagtggg
      781 cacagttgtc actgcttatg aagactgttc aggactggta tcatttactg
ggttcataaa
      841 acaaatcaat ctcaccgcta gagaggcaat actatatttc ttccacaaga
actttgagga
      901 agagataaga agaatgtttg agccagggca ggagacagct gttcctcact
cttatttcat
      961 ccacttccgt tcactaggct tgagtgggaa atctccttat tcatcaaatg
ctgttggtca
     1021 cgtgttcaat ctcattcact ttgtaggatg ctatatgggt caagtcagat
ccctaaatgc
     1081 aacggttatt gctgcatgtg ctcctcatga aatgtctgtt ctaggggct
atctgggaga
     1141 ggaattcttc gggaaaggga catttgaaag aagattcttc agagatgaga
aagaacttca
     1201 agaatacgag gcggctgaac tgacaaagac tgacgtagca ctggcagatg
atggaactgt
     1261 caactctgac gacgaggact acttttcagg tgaaaccaga agtccggagg
ctgtttatac
     1321 tcgaatcatg atgaatggag gtcgactaaa gagatctcac atacggagat
atgtctcagt
     1381 cagttccaat catcaagccc gtccaaactc attcgccgag tttctaaaca
agacatattc
     1441 gagtgactca ggttccggag agaacctcta cttccaatcg ggatccggta
gccatggctt
     1501 cccgccggag gtggaggagc aggatgatgg cacgctgccc atgtcttgtg
cccaggagag
     1561 cgggatggac cgtcaccctg cagcctgtgc ttctgctagg atcaatgtgt
aagaagttga
     1621 ataacaaaat gccggaaatc tacggattgt gtatatccat catgaaaaaa
actaacaccc
     1681 ctcctttcga accatcccaa acatgagcaa gatctttgtc aatcctagtg
ctattagagc
     1741 cggtctggcc gatcttgaga tggctgaaga aactgttgat ctgatcaata
gaaatatcga
     1801 agacaatcag gctcatctcc aaggggaacc catagaggtg gacaatctcc
ctgaggatat
     1861 ggggcgactt cacctggatg atggaaaatc gcccaaccat ggtgagatag
```

```
ccaaggtggg
     1921 agaaggcaag tatcgagagg actttcagat ggatgaagga gaggatccta
gcttcctgtt
     1981 ccagtcatac ctggaaaatg ttggagtcca aatagtcaga caaatgaggt
caggagagag
     2041 atttctcaag atatggtcac agaccgtaga agagattata tcctatgtcg
cggtcaactt
     2101 tcccaaccct ccaggaaagt cttcagagga taaatcaacc cagactactg
gccgagagct
     2161 caagaagtgt gtatcagtac atg
//
```

```
LOCUS       Exported                2024 bp ds-DNA     linear   SYN
07-JAN-2019
DEFINITION  synthetic linear DNA.
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 2024)
  AUTHORS   Trial User
  TITLE     Direct Submission
  JOURNAL   Exported Jan 7, 2019 from SnapGene 4.1.9
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..2024
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     primer_bind     1..48
                     /label=Barcode1_cagc_N10_leader_fp
     primer_bind     1..20
                     /label=Barcode1_cagc_fp
     gap             21..30
                     /estimated_length=10
     5'UTR           31..100
                     /label=RV leader
                     /note="SPBN leader"
     misc_feature    89..97
                     /locus_tag="Transcriptional start (N)"
                     /label=Transcriptional start (N)
     CDS             101..1450
                     /codon_start=1
                     /locus_tag="N"
                     /label=N
                     /note="N"
                     /
translation="MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPCITLGK

APDLNKAYKSVLSGMSAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIARKGDKIT

PGSLVEIKRTDVEGNWALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKT

NIADRIEQIFETAPFVKIVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYS

AIRVGTVVTAYEDCSGLVSFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAV

PHSYFIHFRSLGLSGKSPYSSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSV

LGGYLGEEFFGKGTFERRFFRDEKELQEYEAAELTKTDVALADDGTVNSDDEDYFSGET

RSPEAVYTRIMMNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYSSDS"
```

```
     misc_feature    1502..1512
                     /label=Transcription stop/pA signal (N/P & P/M))
     misc_feature    1515..1523
                     /locus_tag="Transcriptional start (P)"
                     /label=Transcriptional start (P)
     misc_feature    1544..2008
                     /locus_tag="P"
                     /label=P
                     /note="P"
     primer_bind     complement(1989..2024)
                     /label=Barcode2_P_rp
ORIGIN
        1 tcagacgatg cgtcatcagc nnnnnnnnnn acgcttaaca accagatcaa
agaaaaaaca
       61 gacattgtca attgcaaagc aaaaatgtaa cacccctaca atggatgccg
acaagattgt
      121 attcaaagtc aataatcagg tggtctcttt gaagcctgag attatcgtgg
atcaatatga
      181 gtacaagtac cctgccatca aagatttgaa aaagccctgt ataaccctag
gaaaggctcc
      241 cgatttaaat aaagcataca agtcagtttt gtcaggcatg agcgccgcca
aacttaatcc
      301 tgacgatgta tgttcctatt tggcagcggc aatgcagttt tttgagggga
catgtccgga
      361 agactggacc agctatggaa ttgtgattgc acgaaaagga gataagatca
ccccaggttc
      421 tctggtggag ataaaacgta ctgatgtaga agggaattgg gctctgacag
gaggcatgga
      481 actgacaaga gaccccactg tccctgagca tgcgtcctta gtcggtcttc
tcttgagtct
      541 gtataggttg agcaaaatat ccgggcaaaa cactggtaac tataagacaa
acattgcaga
      601 caggatagag cagatttttg agacagcccc ttttgttaaa atcgtggaac
accatactct
      661 aatgacaact cacaaaatgt gtgctaattg gagtactata ccaaacttca
gatttttggc
      721 cggaacctat gacatgtttt tctcccggat tgagcatcta tattcagcaa
tcagagtggg
      781 cacagttgtc actgcttatg aagactgttc aggactggta tcatttactg
ggttcataaa
      841 acaaatcaat ctcaccgcta gagaggcaat actatatttc ttccacaaga
actttgagga
      901 agagataaga agaatgtttg agccagggca ggagacagct gttcctcact
cttatttcat
      961 ccacttccgt tcactaggct tgagtgggaa atctccttat tcatcaaatg
ctgttggtca
     1021 cgtgttcaat ctcattcact ttgtaggatg ctatatgggt caagtcagat
ccctaaatgc
     1081 aacggttatt gctgcatgtg ctcctcatga aatgtctgtt ctaggggggct
atctgggaga
```

```
     1141 ggaattcttc gggaaaggga catttgaaag aagattcttc agagatgaga
aagaacttca
     1201 agaatacgag gcggctgaac tgacaaagac tgacgtagca ctggcagatg
atggaactgt
     1261 caactctgac gacgaggact acttttcagg tgaaaccaga agtccggagg
ctgtttatac
     1321 tcgaatcatg atgaatggag gtcgactaaa gagatctcac atacggagat
atgtctcagt
     1381 cagttccaat catcaagccc gtccaaactc attcgccgag tttctaaaca
agacatattc
     1441 gagtgactca taagaagttg aataacaaaa tgccggaaat ctacggattg
tgtatatcca
     1501 tcatgaaaaa aactaacacc cctcctttcg aaccatccca aacatgagca
agatctttgt
     1561 caatcctagt gctattagag ccggtctggc cgatcttgag atggctgaag
aaactgttga
     1621 tctgatcaat agaaatatcg aagacaatca ggctcatctc caaggggaac
ccatagaggt
     1681 ggacaatctc cctgaggata tggggcgact tcacctggat gatggaaaat
cgcccaacca
     1741 tggtgagata gccaaggtgg gagaaggcaa gtatcgagag gactttcaga
tggatgaagg
     1801 agaggatcct agcttcctgt tccagtcata cctggaaaat gttggagtcc
aaatagtcag
     1861 acaaatgagg tcaggagaga gatttctcaa gatatggtca cagaccgtag
aagagattat
     1921 atcctatgtc gcggtcaact ttcccaaccc tccaggaaag tcttcagagg
ataaatcaac
     1981 ccagactact ggccgagagc tcaagaagct atacatgact ctgc
//
```