SapTrap, a Toolkit for High-Throughput CRISPR/Cas9 Gene Modification in *Caenorhabditis elegans*

Matthew L. Schwartz*,† and Erik M. Jorgensen*,†,1
*Department of Biology and †Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84112-0840

ABSTRACT In principle, clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 allows genetic tags to be inserted at any locus. However, throughput is limited by the laborious construction of repair templates and guide RNA constructs and by the identification of modified strains. We have developed a reagent toolkit and plasmid assembly pipeline, called "SapTrap," that streamlines the production of targeting vectors for tag insertion, as well as the selection of modified *Caenorhabditis elegans* strains. SapTrap is a high-efficiency modular plasmid assembly pipeline that produces single plasmid targeting vectors, each of which encodes both a guide RNA transcript and a repair template for a particular tagging event. The plasmid is generated in a single tube by cutting modular components with the restriction enzyme SapI, which are then "trapped" in a fixed order by ligation to generate the targeting vector. A library of donor plasmids supplies a variety of protein tags, a selectable marker, and regulatory sequences that allow cell-specific tagging at either the N or the C termini. All site-specific sequences, such as guide RNA targeting sequences and homology arms, are supplied as annealed synthetic oligonucleotides, eliminating the need for PCR or molecular cloning during plasmid assembly. Each tag includes an embedded *Cbr-unc-119* selectable marker that is positioned to allow concurrent expression of both the tag and the marker. We demonstrate that SapTrap targeting vectors direct insertion of 3- to 4-kb tags at six different loci in 10–37% of injected animals. Thus SapTrap vectors introduce the possibility for high-throughput generation of CRISPR/Cas9 genome modifications.

KEYWORDS CRISPR/Cas9; high-throughput; gene tagging; plasmid toolkit

THE clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system has revolutionized genome editing in nearly all model systems, including *Caenorhabditis elegans* (Frøkjær-Jensen 2013; Doudna and Charpentier 2014; Xu 2015). The Cas9 protein cuts genomic DNA at sites that match an ~20-nucleotide guide RNA sequence (Gasiunas *et al.* 2012; Jinek *et al.* 2012). Error-prone repair of the break can create point mutations, small indels, or large deletions at the cut site (Cho *et al.* 2013; Cong *et al.* 2013; Friedland *et al.* 2013; Jinek *et al.* 2013; van Schendel *et al.* 2015). However, the true power of CRISPR/Cas9 for genome editing lies in the insertion of exogenous DNA sequences, such as genetically encoded protein tags (Katic and Grosshans 2013; Lo *et al.* 2013; Mali *et al.* 2013; Tzur *et al.* 2013). To insert an exogenous sequence, one must simply supply a repair template with homology arms that flank the cut site. The cell uses homology-based repair to heal the break and copy the exogenous DNA into the cut site.

Although in theory CRISPR/Cas9 makes it easy to insert exogenous sequences into a genome, practical limitations have prevented high-throughput implementation. Two critical limiting factors are (1) the time and expense required to build both the repair template and guide RNA constructs for each desired insertion and (2) the time required to screen through candidates to identify the rare, heritably modified organisms. In the simplest insertion strategy, a repair template that contains only the exogenous sequence flanked by homology arms is introduced with Cas9 and a guide RNA. Insertion events among the progeny can be identified only by PCR or by screening for the expected phenotype, which creates a large screening burden. There are two strategies to enrich for the rare animals with edited DNA: either by insertion of selectable markers at the target site or by co-CRISPR events at a second site (Arribere *et al.* 2014; Kim *et al.* 2014; Dickinson *et al.* 2015; Norris *et al.* 2015; Paix *et al.* 2015; Ward 2015).

In the first strategy, a selectable marker is included with the exogenous sequence in the repair template (Dickinson *et al.*

doi: 10.1534/genetics.115.184275
Manuscript received October 31, 2015; accepted for publication January 20, 2016; published Early Online January 29, 2016.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.184275/-/DC1.
[1]Corresponding author: Department of Biology, University of Utah, 257 S. 1400 E., Salt Lake City, UT 84112-0840. E-mail: jorgensen@biology.utah.edu

2013). Transgenic animals can then be directly identified, but it complicates the design and construction of the repair template. The position of the selectable marker within the repair template is typically chosen on a case-by-case basis to limit interference between the target gene and the selectable marker. However, the unique design and construction of each repair template is time-consuming. Recently, the Goldstein and Calarco groups have embedded selectable markers within a synthetic intron in the exogenous sequence (Dickinson *et al.* 2015; Norris *et al.* 2015). This arrangement simplifies construction because the selectable marker no longer needs to be positioned in a unique location within each repair template. However, the selectable markers are oriented in the same direction as the target gene, and transcription of the target gene is terminated in the synthetic intron until the selectable marker is removed in a later step. Thus, selectable marker strategies allow direct selection of insertions but either require complex repair template designs or compromise target gene expression.

In "co-CRISPR" strategies, a second locus is edited simultaneously with the target locus (Arribere *et al.* 2014; Kim *et al.* 2014). Simple markerless repair templates can be used at the target locus because animals are selected by the second-site edit. The phenotype produced by the second-site edit identifies worms in which Cas9 was highly active, the progeny of which are enriched for edits at the target locus. However, among selected animals, the fraction of animals edited at the target locus is highly variable, so screening is still required (Arribere *et al.* 2014; Paix *et al.* 2015; Ward 2015). Hence, co-CRISPR strategies allow simplified repair template construction and allow target gene function in primary strains but reduce selection efficiency.

Both selectable marker-based strategies and co-CRISPR have been optimized for the insertion of relatively simple sequences, such as translational GFP fusions. Classical transgene methods allow a greater variety of tag functions, including transcriptional reporting, translational fusions, conditional expression, and tissue-specific and even single-cell expression. Strategies to add these functions to native-locus tags can be envisioned, but they will generally involve adding regulatory sequences around the tag in the repair template. Adding these regulatory sequences will in turn complicate repair template design and assembly. Ideally, repair templates would be assembled in a modular fashion to simplify the addition of regulatory sequences to tags.

Here we present SapTrap, a PCR-free high-efficiency modular plasmid assembly method for high-throughput production of CRISPR/Cas9 targeting plasmids for *C. elegans*. A single-tube SapTrap assembly reaction generates a single plasmid targeting vector that encodes both a guide RNA transcript and a repair template for an individual insertion event. The guide RNA targeting sequence and homology arms are supplied as synthetic, annealed oligonucleotides, and a prebuilt plasmid library supplies the remaining repair template components: fluorescent and nonfluorescent tags, a selectable marker, and optional regulatory sequences. A novel intron-embedded selectable marker strategy obviates the need to position the marker on a gene-by-gene basis and yields concurrent expression of the tagged gene and selectable marker after genomic insertion. Nonetheless, the marker is removable to yield a scarless insertion of the tag. Finally, we provide repair templates capable of tagging a protein in a tissue-specific manner. The SapTrap toolkit reduces the expense and workload necessary to produce vectors for genome editing in the worm and will expand the experimental utility of tags inserted in the genome.

Materials and Methods

Reagents

All chemicals were purchased from Sigma-Aldrich (St. Louis). All enzymes were purchased from New England Biolabs (Beverly, MA). All synthetic DNAs were purchased from Integrated DNA Technologies.

Strains

The wild type is N2 Bristol. All strains were grown at room temperature on OP50 or HB101 bacteria.

Plasmid construction

Single guide RNA and repair template destination vectors: pMLS134 was derived from Addgene plasmid #46169 [*PU6*::*unc-119_sgRNA*, a gift from John Calarco (Friedland *et al.* 2013)]. Using PCR site-directed mutagenesis, the *unc-119* single guide RNA (sgRNA) targeting sequence was replaced with a *Sap*I insertion site (see Supporting Information, Figure S3B), and the *Sap*I site in the plasmid backbone was replaced with an *Nde*I site. pMLS256 and pMLS257 were derived from pBluescript II (Agilent Technologies). First, using PCR site-directed mutagenesis, the multiple-cloning site of pBluescript II was replaced with a *Sap*I insertion site (see Figure S3A), and the *Sap*I site in the pBluescript backbone was deleted. To make pMLS256, the *PU6:sgRNA* cassette from pMLS134 was blunt ligated into pMLS257 linearized by PCR amplification with the primers used to delete the backbone *Sap*I site. PCR site-directed mutagenesis was then used to modify the sgRNA *Sap*I insertion site to match the repair template insertion site.

Three-site destination vectors: To make pMLS234, pMLS235, and pMLS236, first a *gfp*::syntron-nested-*Cbr-unc-119* cassette containing flexible linker sequences was introduced into pDONR-221 by BP reaction (Thermo Fisher). The *PU6*::*sgRNA* cassette from pMLS134 was amplified and blunt ligated into the resultant vector linearized using the same primers as for *PU6*::*sgRNA* insertion into pMLS256 (above). The resultant vector was amplified in two pieces with primers that introduced additional *Sap*I insertion sites on the 5′ end of GFP and on the 3′ end of GFP. For pMLS234, the 5′ flexible linker was deleted by the *Sap*I insertion site, for pMLS235 the 3′ flexible linker was deleted by the *Sap*I insertion site, and for pMLS236 neither flexible linker was deleted. The final vectors were produced by blunt ligating the appropriate PCR fragments.

Tag/selectable marker donor plasmids: pMLS252, pMLS254, pMLS271, pMLS291, and pMLS292 were made in two steps. First, the tag sequences (*gfp*, *halotag*, *snaptag*, *mcherry*, or *2xNLS-mcherry*) were PCR amplified with primer-encoded *Sap*I restriction sites and attB1 and attB2 tails. PCR products were inserted into pDONR-221 by BP reaction. The resultant entry vectors were linearized by PCR, and the syntron::*Cbr-unc-119* cassette, amplified from pCFJ150 (Frøkjær-Jensen *et al.* 2008) with primer-appended syntron sequences, was inserted by Gibson assembly (New England Biolabs). pMLS286 was produced by first blunt ligating a PCR-amplified tagRFP sequence with primer-encoded *Sap*I sites into pMLS280 (described below) linearized by digestion with *Xma*I. The syntron::*Cbr-unc-119* cassette was introduced into this entry vector by Gibson assembly, as described above.

N- and C-tagging connector plasmids: pMLS268, pMLS269, and pMLS279 were constructed by PCR amplifying the insert sequence with primers containing attB1 and attB2 tails and inserting the PCR products into pDONR-221 by BP reaction. pMLS285, pMLS287, pMLS288, pMLS382, and pMLS383 were constructed by blunt ligating the insert sequences into pMLS280. All inserts were generated from ligated oligos; the oligos for the *egl-13* NLS sequence were designed to anneal with large 5′ overhangs that were subsequently filled in by incubation with Phusion polymerase (New England Biolabs). To construct pMLS272, -281, and -282, individual fragments of the insert sequence were produced by PCR or annealing oligos. The fragments were then assembled by Gibson assembly and inserted into pDONR-221 by BP reaction.

Additional plasmids: pMLS280 was made by Gibson assembly. The multiple-cloning site from pBluescript II was amplified with M13F and M13R primers and inserted between the M13F and M13R primer binding sites in pDONR221. pMLS328 was generated by PCR site-directed mutagenesis of pDD104 [*Peft-3*::*Cre*, Addgene plasmid #47551, a gift from Bob Goldstein (Dickinson *et al.* 2013)] with a primer-encoded *egl-13* NLS sequence inserted at the 3′ end of the Cre recombinase sequence. FLP-D5 was modified to include an N-terminal SV40 nuclear localization sequence and a C-terminal *egl-13* nuclear localization sequence and harbors an aspartic acid residue at position 5. FLP expression vectors were constructed by multisite LR reactions (Thermo Fisher) into either pCFJ150(pMLS260, pMLS262) or pCFJ201(pMLS359, pMLS360) (Frøkjær-Jensen *et al.* 2012) with promoters initially cloned into [4–1] entry vectors, the FLP construct cloned into a [1–2] entry vector, and the *let-858* 3′-UTR cloned into a [2–3] entry vector.

Selection of overhang sequences

The 5′ overhangs within the sgRNA insertion site, at the repair template–plasmid junctions and at the junctions of the tag and marker cassette, were designed without selection-based optimization. To select 5′ overhangs for the homology arm–connector junctions with the highest fidelity, we designed oligo pairs to occupy both homology arm slots and both connector module slots in the SapTrap assembly reaction. We screened a panel of oligo pairs encoding four different candidate overhang sequences for each junction. We ran 16 SapTrap assembly reactions containing the pMLS256 destination vector, a GFP and *Cbr-unc-119* donor plasmid, and every possible combination of overhang sequences for the connector–homology arm junctions. Each reaction included both N- and C- terminal connector modules. The reaction containing GCG (Ala) as the 5′ homology arm–C-tagging connector overhang and ACG (Thr) as the N-tagging connector–3′ homology arm overhang produced the highest rate of correctly assembled plasmids (11/12, >90%).

SapTrap assembly reactions

SapTrap enzyme mixture: A 1.25X master SapTrap enzyme mixture (1.25X NEB cutsmart buffer; 1.25 mM ATP, pH 7.6; 6.25 mM DTT; 12.5 units/μl T4 DNA ligase; 0.25 units/μl T4 polynucleotide kinase; and 1.25 units/μl *Sap*I) was prepared on ice and frozen in 2-μl aliquots at −80°. Because *Sap*I precipitates from solution, all solutions containing *Sap*I were vigorously pipetted up and down to resuspend *Sap*I before withdrawing an aliquot.

DNA preparation: Oligo pairs (see File S1 for step-by-step protocols) were annealed as follows: complementary oligo pairs were resuspended to 10 μM each in 1× oligo annealing buffer (OAB) (20 mM Tris-Cl, pH 7.5; 50 mM NaCl; and 1 mM $MgCl_2$) and heated to 95°–100° in a heat block. The heat block was then switched off and allowed to cool slowly (1–2 hr) to room temperature. Sets of three annealed oligo pairs comprising the 5′ homology arm, the 3′ homology arm, and the sgRNA targeting sequence were diluted in TE buffer (10 mM Tris-Cl, pH 8.0; 1 mM EDTA, pH 8.0) to 150 nM each pair. Destination and donor plasmids were diluted individually to 50 nM in TE, using the formula

$$[\text{plasmid}]\ (\text{nM}) = 1.5 \times \frac{[\text{plasmid}]\ (\text{ng}/\mu\text{l})}{\text{plasmid size (Kb)}}. \qquad (1)$$

SapTrap reaction: To assemble combined sgRNA and repair template vectors, equal volumes of each 50-nM plasmid stock (pMLS256 destination, connector donor, and tag and selectable marker donor) and the annealed oligo mixture were premixed. A total of 0.5 μl of this DNA mixture was mixed with 2 μl of SapTrap enzyme mixture, yielding a final reaction containing 2.5 nM of each plasmid and 7.5 nM of each annealed oligo pair. Reactions were incubated at 20°–25° overnight. Then, the T4 DNA ligase was inactivated by a 30-min incubation at 65°. Reactions then received 2.5 μl of 1× Cutsmart buffer (New England Biolabs) + 2 units/μl of a counterselection restriction enzyme and were incubated at 37° for 1 hr. A total of 1–2 μl of the reaction mixture was used to transform chemically competent TOP10 *Escherichia coli*.

Clone screening: Colonies were screened for correctly assembled plasmids by colony PCR with M13F (5′-TGTAAAACGACGGCCAGT)

and M13R (5′-CAGGAAACAGCTATGACCATG) primers. Individual bacterial colonies were sampled with a sterile P10 pipette tip, which was dipped into a 5-μl 1× PCR reaction mixture (Phusion polymerase, buffer HF; New England Biolabs). Reactions were cycled 30 times with a 10-sec, 58° annealing step and a 1-min 45-sec, 72° extension step. Plasmid DNA was isolated from colonies that produced a clear band of the correct size (~3.4–4 kb, depending on insert) in the colony PCR reaction. Properly assembled plasmids were isolated and sequenced using M13F, M13R, and oMLS471 (5′-TCCAAGAACTCGTACAAAAATGCTC) sequencing primers to confirm the 3′ homology arm, the 5′ homology arm, and the sgRNA targeting sequence, respectively.

Worm injections and strain isolation

CRISPR/Cas9 injections: To insert tags with combined sgRNA and repair template targeting vectors, young adult EG6207 [*unc-119(ed3)* III] hermaphrodites reared at 15° on HB101 bacteria were micro-injected in the gonad with an injection mixture of 65 ng/μl combined sgRNA and repair template targeting plasmid, 25 ng/μl Addgene plasmid #46168 (*Peft-3*::*cas9-SV40_NLS*::*tbb-2 3′UTR*, a gift from John Calarco) (Friedland *et al.* 2013), and 10 ng/μl fluorescent co-injection markers. For *gfp* and nonfluorescent insertions, the co-injection markers were 2 ng/μl pCFJ90 (*Pmyo-2*::*mcherry*), 4 ng/μl pCFJ104 (*Pmyo-3*::*mcherry*), and 4 ng/μl pGH8 (*Prab-3*::*mcherry*) (Frøkjær-Jensen *et al.* 2008). For *tagrfp* insertions, the co-injection markers were 2 ng/μl pCFJ420 (*Pmyo-2*::*gfp*::*h2b*) and 8 ng/μl pCFJ421 (*Peft-3*::*gfp*::*h2b*) (Frøkjær-Jensen *et al.* 2012). All plasmids were purified using a QIAGEN (Valencia, CA) miniprep kit followed by ethanol precipitation. Injected P0 worms were placed two to a plate on 6-cm NGM plates seeded with OP50 bacteria and incubated for 7–10 days at 25°. Plates were inspected under a fluorescence dissecting microscope for the presence of motile [*unc-119(+)*] animals without co-injection markers. Such animals were singled out and insertions were confirmed by PCR with primers outside the homology arms. Only a single strain was isolated from each plate.

Quantification of insertion frequency: Insertion frequency was calculated by dividing the number of independent strains containing the desired insertion by the estimated number of P0 animals that were successfully injected. A successful injection was defined as an injection producing array-positive animals in the F_2 generation. Each recovery plate was originally seeded with two P0 animals. After 7–10 days, the plates were inspected and scored as either positive for a successful injection event (presence of array-positive F_2 animals) or negative for a successful injection event (no array-positive F_2 animals) (see Table S3). Assuming the P0 animals were distributed independently of injection quality, the number of successfully injected animals was calculated as

$$u + i = 1 \tag{2}$$

$$u^2 = \frac{N}{N+P} \tag{3}$$

$$i = 1 - \left(\frac{N}{N+P}\right)^{1/2}, \tag{4}$$

where u is the fraction of worms that were not successfully injected, i is the fraction of worms that were successfully injected, N is the number of plates with two unsuccessfully injected animals, and P is the number of plates with one or two successfully injected animals.

Cre recombinase injections: To excise the LoxP-flanked *Cbr-unc-119* cassette, young adult hermaphrodites were micro-injected in the gonad with an injection mixture of 50 ng/μl pDD104 (*Peft-3*::*Cre*) or 50 ng/μl pMLS328 (*Peft-3*::*2xNLS-Cre*), 48 ng/μl pBluescript II, and 2 ng/μl pCFJ90 (*Pmyo-2*::*mcherry*). A total of 10–20 array-positive F_1 animals were placed 5 to a plate on 6-cm NGM plates seeded with OP50 and incubated at room temperature for 3–5 days. Unc-119 F_2 animals were cloned. To remove the *unc-119(ed3)* allele, *Cbr-unc-119(−)* hermaphrodites were mated with N2 males, and the F_1 male progeny were mated to either EG6207 (*unc-32* insertions) or N2 (all other insertions) hermaphrodites. F_2 progeny from the second cross were cloned to identify strains homozygous for both the tag insertion and *unc-119(+)*. *unc-32* insertions were backcrossed to EG6207 to facilitate identification of chromosomal crossovers between *unc-32* and *unc-119* (5.6 cM apart) on chromosome III.

FLP injections: Young adult hermaphrodites of the *Cbr-unc-119(−) gfp*::*FLP-on*::*rab-3* [*gfp*::*FLP-on*::*rab-3* II; *unc-119(ed3)*, III] or *snt-1*::*FLP-on*::*gfp* strain [*snt-1*::*FLP-on*::*gfp* II; *unc-119(ed3)* III] were micro-injected in the gonad with an injection mix of 5 ng/μl FLP expression vector [*unc-119(+)*] and either 95 ng/μl pBluescript II or 5 ng/μl pRJH179 (*Punc47*::*snb-1*::*tagrfp*) and 90 ng/μl pBluescript II. *unc-119*(+) F_1 animals were cloned, and strains with stably passaging extrachromosomal arrays were selected for imaging studies.

Thrashing assays: Individual young adult hermaphrodite worms reared on HB101 (*unc-119*) or OP50 (all other strains) were placed in a 75-μl drop of M9 buffer on an NGM plate. After a 1-min equilibration period, worms were observed for 1 min and the number of thrashing events was counted.

Fluorescence imaging: Worms were anesthetized in 25 mM NaN_3 on a 5% agarose pad. Once fully anesthetized, the pads were overlaid with a glass coverslip, sealed, and imaged on a Zeiss (Thornwood, NY) LSM-5 Pascal laser-scanning fluorescence microscope equipped with a 488-nm argon laser and a 543-nm helium-neon laser. Raw image stacks were converted to Z projections and adjusted using ImageJ software.

Strains are available upon request. Plasmids are available from Addgene as a single kit or as individual plasmids (https://www.addgene.org/).

Data availability

All relevant data are contained within the manuscript and supporting information files.

Results

A combined tag and selectable marker

To streamline the production of repair templates for CRISPR/Cas9-mediated insertion, we first sought to reduce the general complexity of the repair template. We opted to use a selectable marker to facilitate direct identification of modified worms. In this design, a LoxP-flanked *C. briggsae unc-119* selectable marker is positioned within a synthetic intron of the tag, such as *gfp* (Figure 1A). The LoxP sites allow the *unc-119* cassette to be excised by expression of CRE recombinase. To allow concurrent expression of both *gfp* and *unc-119*, we inserted the *unc-119* gene in the opposite orientation relative to *gfp*, a configuration that mimics naturally occurring intron-nested genes (Kumar 2009). Using this cassette, a complete repair template can be generated simply by adding homology arms to each side of the *gfp* tag.

To test whether our cassette allows concurrent expression of both the tagged gene and the selectable marker, we used CRISPR/Cas9 to insert the cassette as an N-terminal tag on UNC-17 (vesicular acetylcholine transporter) or as a C-terminal tag on UNC-32 (v-ATPase subunit). Compromising the function of either of these genes causes an uncoordinated phenotype, and complete knockout of either gene is lethal (Alfonso *et al.* 1993; Pujol *et al.* 2001). The repair templates were injected into *unc-119(ed3)* animals and rescued animals were selected in the F_2. For both loci, homozygous strains expressed GFP in the expected pattern and were phenotypically wild type (Unc-119+; Figure 1, B and C). We assayed locomotion in each strain by measuring thrashing in liquid media. The primary *unc-32*::*gfp* strain showed no locomotion defect compared to the wild type, whereas the primary *gfp:unc-17* strain exhibited a moderate thrashing defect. After excising the *unc-119(+)* selectable marker by germline CRE recombinase expression and outcrossing to remove the *unc-119(ed3)* mutation, both tagged strains exhibited thrashing behavior indistinguishable from the wild type (Figure 1C). Thus, the locomotion defect of the primary *gfp*::*unc-17* strain is not attributable to the GFP tag impairing UNC-17 protein function; rather, the embedded *unc-119* gene interferes with *unc-17* expression or vice versa. The locomotion defect of the primary *gfp*::*unc-17* strain was not obvious on solid media and did not complicate selection of insertions from the *unc-119(ed3)* parent strain. Thus, our tagging cassette allows concurrent expression of both the tagged gene and the selectable marker; however, the function of either gene may be moderately impaired in primary strains for some loci.

Modular assembly of targeting vectors

Next, we sought to streamline the production of vectors for CRISPR/Cas9 repair. Each individual targeting event requires two components: an sgRNA (Jinek *et al.* 2012) that directs Cas9 to cleave at a specific site and a repair template containing the tag flanked by homology arms of the target (Figure 1A). Typically, these components are generated independently, so that two constructs must be built for each targeting event. In addition, Cas9 must be supplied as yet another plasmid (in our case P*eft-3:cas9* is co-injected in all experiments). To reduce the plasmid assembly workload and subsequent plasmid management, we designed a single plasmid to encode both the sgRNA transcript and the repair template (Figure 2A).

To modify a single vector at multiple sites in a single step, we employed Golden Gate assembly (Engler *et al.* 2009). Golden Gate assembly uses the property of type-IIS restriction enzymes to cut outside of their recognition sequence to drive the ordered assembly of up to 10 different DNA fragments in a single-tube digest-and-ligate reaction. We designed our Golden Gate strategy, using the restriction enzyme *Sap*I, and have named our assembly method "SapTrap." We chose *Sap*I because its 7-bp recognition sequence is rare and because it produces 3-base overhangs that can be conveniently positioned to coincide with codons in the open reading frame. First, we designed a destination vector (pMLS256) that is opened by *Sap*I digestion at two sites: one site is in a U6 promoter-driven sgRNA expression cassette and accepts the ~20-bp sgRNA targeting sequence; the second site is flanked by M13 sequencing primer binding sites and accepts the repair template (Figure 2A). We divided the repair template into five separate components to be supplied independently to the assembly reaction: the 5′ and 3′ homology arms, the combined tag and selectable marker cassette, and two optional N- and C-terminal "connectors." The connector modules fit between the tag and homology arms and encode either peptide linkers or regulatory sequences that control how the tag is expressed in relation to the target gene (Figure 2B, Table S1, and Table S5). *Sap*I generates 3-base 5′ overhangs; we designed unique overhang sequences for each junction between DNA components. Overhang sequences were chosen to maximize assembly fidelity and to encode favorable amino acids (M, G, A, T) at junctions within coding regions of the repair template. The overhang sequences were designed to achieve high fidelity in reactions containing all five repair template DNA components; in most constructs the tag is inserted at the N or the C terminus of the protein and only one connector module is necessary (Figure 2).

Individual DNA components can be supplied to the SapTrap reaction in three ways: as PCR products, donor plasmids, or annealed oligos. PCR products and donor plasmids require *Sap*I digestion to produce 5′ overhangs, whereas annealed oligos are designed to contain the appropriate 5′ overhangs without digestion. Because we have successfully inserted tags using homology arms of just 50–60 bp (see below), we prefer annealed oligos for homology arms and sgRNA targeting sequences. A library of donor plasmids supplies DNA components of the three remaining types: N-tagging connectors, C-tagging connectors, and the tags (which also include the selectable marker) (Figure 2B, Table S1, and Table S5). Donor plasmids of the same type all produce the same 5′ overhangs upon *Sap*I digestion; by selecting different combinations of connector and tag donor plasmids from the library, a large number of distinct repair templates can be

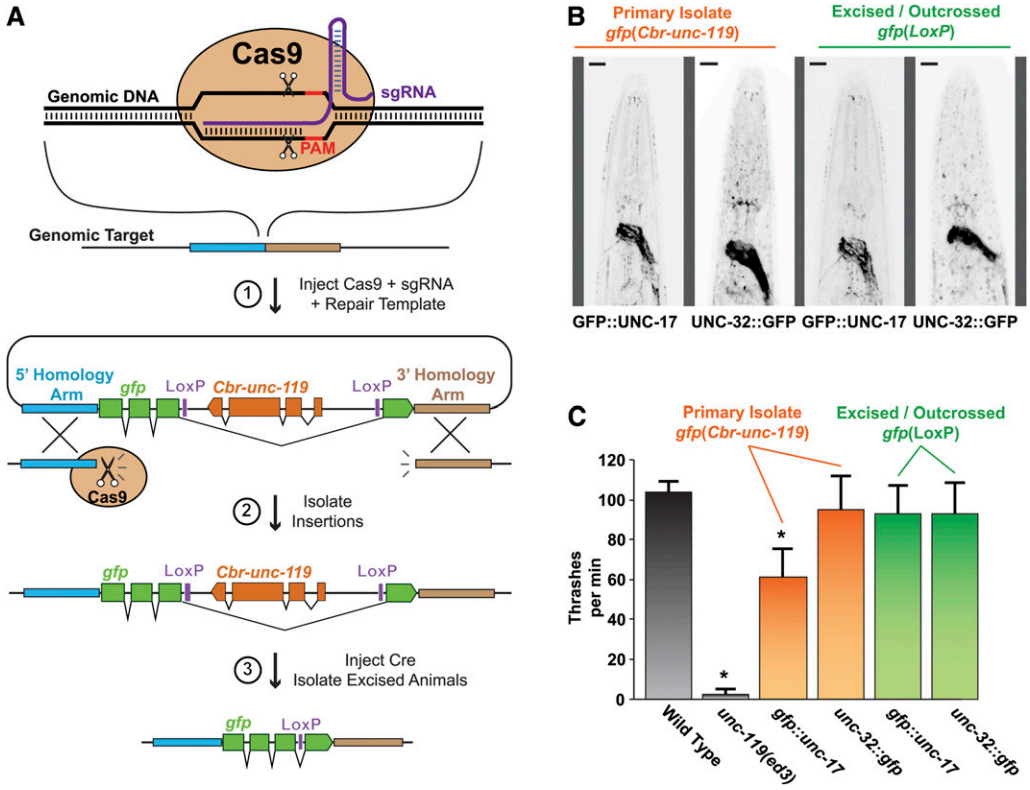


Figure 1 Combined genetic tag and removable selectable marker. (A) Cas9 cleaves the genomic DNA at a specific site, directed by the sgRNA. The repair template contains homology arms flanking the cleavage site and a genetic tag (*gfp*) with a removable *Cbr-unc-119* marker gene embedded in the intron. Insertions are selected by rescue of the *unc-119* phenotype. *Cbr-unc-119* is excised by Cre-recombinase-mediated site-specific recombination, leaving a single LoxP site. (B) Representative fluorescent micrographs of worm strains tagged with *gfp* at the native *unc-17* or *unc-32* genomic loci. Colors have been inverted to enhance contrast. Bars, 10 μm. (C) Thrashing behavior of worms in liquid media for the same strains imaged in B. Bars represent mean ± SD, n = 10. *$P < 0.01$ compared to every other genotype, Student's t-test. $P > 0.01$ for all comparisons among the strains lacking an asterisk (*).

built with a single set of homology arms and sgRNA oligos. For high-throughput projects with the goal of introducing a single tag at a large number of loci, the tag and connector donor plasmids can be eliminated from the SapTrap reaction by preassembling these components into a "three-site" destination vector (Figure S1). The three-site vector requires only synthetic oligos for assembly.

Targeting vectors are assembled by incubating the destination vector, annealed oligos, and donor plasmids with *Sap*I, T4 polynucleotide kinase (to phosphorylate annealed oligos), T4 DNA ligase, and ATP (Figure S2). The enzyme mixture can be prepared in bulk and stored at −80° in single reaction aliquots. To assemble a targeting vector, the DNA components are mixed with a thawed aliquot of SapTrap enzyme mixture and incubated overnight at 20°–25°. Background from unreacted destination vector can be eliminated by subsequently heat-inactivating the DNA ligase and digesting the assembly reaction with a counterselection restriction enzyme. Recognition sites for these enzymes exist only in the portion of the destination vector that is removed during targeting vector assembly (Figure S3). Counterselection restriction sites should be absent from homology arms, repair template, and sgRNA coding sequence; eight different counterselection sites are available so that an appropriate enzyme can be chosen for any particular plasmid. Transformation of competent *E. coli* with a single 2.5-μl SapTrap reaction yields >100 colonies in our experience. The donor plasmids do not contribute background colonies because they contain a kanamycin resistance gene, whereas the assembled targeting vector contains an ampicillin resistance gene. For 24 unique constructs assembled during this study, 49% of all colonies screened contained the correctly assembled plasmid (91/185 colony PCRs performed) with a range of 20–100% correct assemblies for individual reactions. Of 51 plasmids subjected to Sanger sequencing of the oligo-derived sequences, 86% (44/51) had the correct sequence and 14% (7/51) had a single-base deletion or point mutation in one of the oligo-derived sequences.

Short homology arms

In previous studies, homology arms as short as 500 bp for plasmid repair templates and 30 bp for linear repair templates efficiently direct insertion into the *C. elegans* genome (Paix *et al.* 2014; Dickinson *et al.* 2015). To determine the effect of homology arm length on insertion frequency for SapTrap vectors, we targeted the 5′ end of the *snb-1* gene with a *gfp* cassette flanked by homology arms 0, 44, 100, or 400 bp in length (Figure 3A and Table S3). All vectors used the same sgRNA that cleaves 7 bases upstream of the translational start site (see Table S2). As expected, no insertions were recovered using the construct lacking homology arms. However, 44-bp homology arms generated insertions in 21% of injected animals (P0s segregating insertions/successfully injected P0s), and increasing the length of the homology arms to 400 bp resulted in a marginal but insignificant increase in the insertion rate. Short homology arms are advantageous because they can be generated inexpensively by annealing pairs of synthetic oligos. Oligo-derived homology arms eliminate

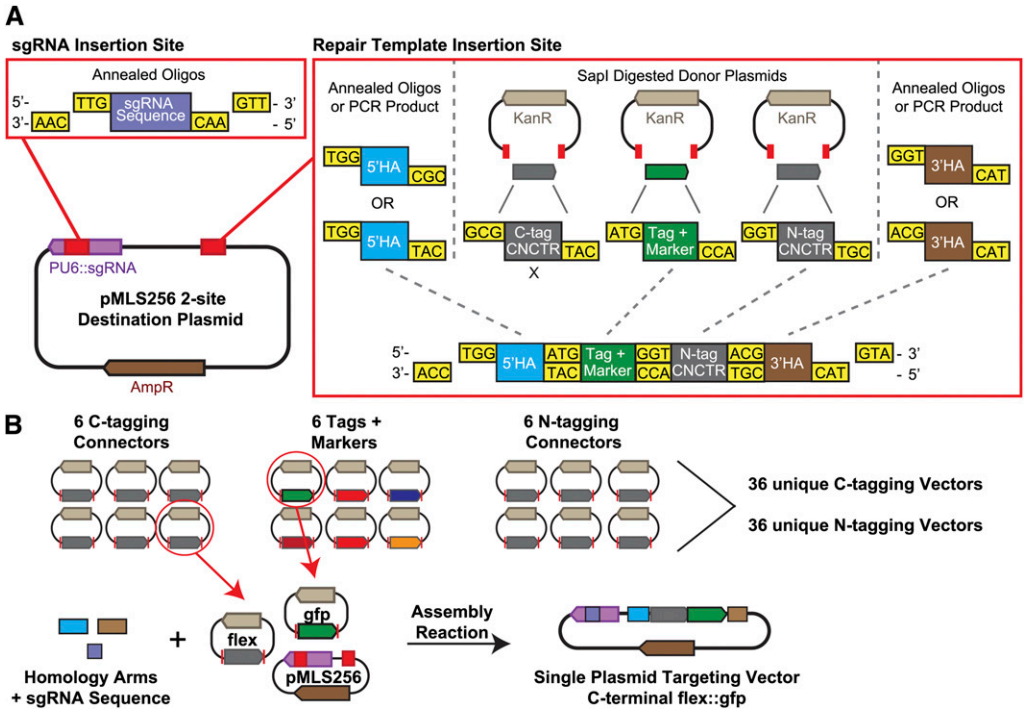


Figure 2 The SapTrap assembly method. (A) pMLS256 contains two *Sap*I insertion sites that accept the repair template and the sgRNA targeting sequence. Individual components assemble in the correct location and in the correct order by ligation of complementary 5′ overhangs. Tag and marker cassettes and connector modules (CNCTR) are liberated from donor plasmids by *Sap*I digestion; sgRNA targeting sequences and homology arms (HA) are supplied as annealed oligos. Individual positions within the repair template can be bypassed by the encoding of 5′ overhangs on adjacent fragments. (B) Modular assembly of repair templates from the SapTrap library. Specific repair templates are built by supplying a pair of homology arms and a guide RNA targeting sequence and selecting a connector module and a tag and marker cassette. The current SapTrap library contains six unique C-tagging connectors, six unique N-tagging connectors, and six unique tags. Thirty-six unique repair templates can be built using a single set of homology arm and guide RNA sequences.

the need to perform PCR during vector construction and facilitate mutation of the PAM site or the sgRNA binding site in the repair template to prevent recutting after repair. For these reasons, we used annealed oligos ("short arms") for all subsequent plasmid constructions, unless noted otherwise. Currently, 60-base oligos are the longest oligos that can be custom synthesized inexpensively with a low error rate; 3 bp must be dedicated to the SapTrap overhang, and thus homology arms are 57 bp in length. Homology arm length was extended from 44 bp tested at the *snb-1* locus to 57 bp because longer homology arms allow more flexibility in picking a guide sequence relative to the insertion site.

To validate our SapTrap vectors, we tested insertion at six different genes. We generated short-arm targeting vectors to introduce *gfp* or *tagrfp* with intron-nested *unc-119* at the 5′ end of *rab-3*, *snb-1*, and *unc-17* and at the 3′ end of *sng-1*, *snt-1*, and *unc-32*. These vectors generated insertions at all loci except the synaptotagmin locus *snt-1*, with insertion frequencies ranging from 10% to 37% (Figure 3B and Table S3). All insertions were verified by PCR and fluorescence imaging (Figure S6). We conclude that the short-arm targeting vectors direct efficient insertion at a wide range of genetic loci.

Because we were unable to insert a tag into *snt-1* at the original site, we tested a different Cas9 cut site. We identified a suitable cut site 277 bp upstream of the *snt-1* stop codon (see Figure S4). To tag *snt-1* at the 3′ end using this cut site, the 5′ homology arm needed to include the 277 bases between the upstream cut site and the insertion site, as well as a true 5′ homology arm flanking the 5′ side of the cut site. Because this homology arm is too long to assemble from annealed oligos, we ordered a synthetic 750-bp DNA fragment (gBlocks; IDT) encoding both 5′ and 3′ homology arms and the sgRNA targeting sequence. We flanked each homology arm and the sgRNA targeting sequence with *Sap*I sites so the intact gBlock sequence could be fed into a SapTrap reaction to release all three target site-specific sequences. The 750-bp length gBlock allotment allowed us to encode longer homology arms (165 bp for the 5′ homology arm and 147 bp for the 3′ homology arm) than possible using synthetic oligos. The resulting targeting vectors (*snt-1* long) directed insertion of *gfp* or *tagrfp* at the 3′ end of *snt-1* in 7% of injected P0s, a low but comparable frequency to the short-armed vectors targeting other genes. These results demonstrate that large synthetic double-stranded DNA substrates are an alternative to oligo-based assembly pipelines for SapTrap vectors.

In addition to 40 successfully tagged strains (Figure 3 and Table S3), we isolated 7 strains that lacked an insertion at the targeted locus, but were stably rescued for *unc-119*. Because these strains did not carry extrachromosomal array markers, they appear to represent off-site insertions of the *unc-119(+)* repair template and were not counted as targeted insertions in our frequency calculations. We observed an off-site insertion rate of 15% (7 false-positive strains/47 total positive strains). False-positive strains that appear to contain off-site insertions have been reported previously (Dickinson *et al.* 2015; Katic *et al.* 2015) and appear to be a general consequence of tag insertion by CRISPR/Cas9.

Tissue-specific tags

Within SapTrap's modular design, we included two optional connector modules to add sequences between the tag and

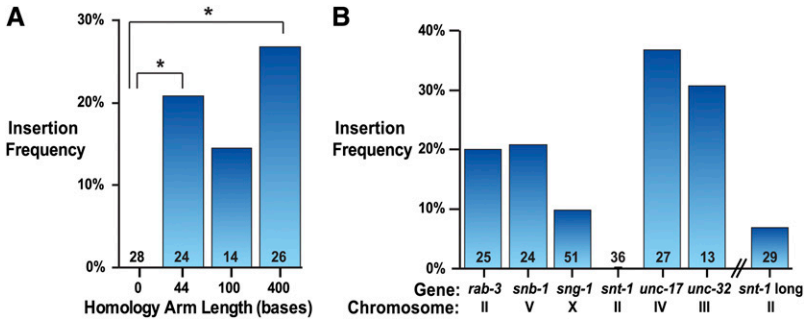


Figure 3 Insertion frequency of single plasmid targeting vectors containing a combined tag and selectable marker cassette. (A) Effect of homology arm length on insertion frequency at the *snb-1* locus. Insertion frequency is the number of independent strains containing the targeted insertion divided by the total number of successfully injected P0 animals (defined as P0s that gave rise to array-positive F_2 progeny). Number over bars indicates number of successfully injected P0 animals. $*P < 0.05$, Fisher's exact test; all other comparisons $P > 0.05$. (B) Insertion frequency at different genetic loci. All targeting vectors used 57-bp arms except for "*snt-1* long," which used 160 (5′)- and 140 (3′)-bp homology arms, and "*snb-1*," which used 44-bp arms. Insertion frequency was calculated as in A.

each homology arm (Figure 2A). For simple tagging operations, we built connectors that encode glycine-rich flexible linkers (see Table S1 and Table S5). However, the connector modules can include more complex regulatory sequences to support more complex tagging operations. For example, it is often experimentally useful to restrict expression of a tagged protein to a specific subset of tissues or even a single cell. To illustrate the utility of the connectors, we developed a set of connector modules that can confer cell type-specific expression of tags at native loci.

We developed specialized strategies for cell-specific expression of either N- or C-terminal tags. Both strategies are based on the FLP/FRT recombination system (Golic and Lindquist 1989; Hubbard 2014). FLP is a site-specific recombinase from yeast that acts at FLP-recombinase targets (FRTs); if two FRTs are in the same orientation, FLP will excise the sequence between the FRTs. For both the N- and the C-terminal tags, we built connector modules containing tandem FRT sites flanking an "off cassette" that disrupts the attachment of the tag to the protein and blocks expression of the fluorescent protein (Figure 4A). Excision of the intervening sequence by FLP couples the tag to the protein of interest, using an "FLP-on" strategy (Davis *et al.* 2008).

For N-terminal tags, the off cassette consists of a PEST degron from mouse ornithine decarboxylase (Li *et al.* 1998) and the intergenic region from the *gdp-2/gdp-3* operon (Lee *et al.* 1992). Before the off cassette is excised, the transcript is *trans*-spliced into two messages: one containing the tag fused to the PEST degron and a second containing the untagged target gene. The tag::PEST protein is translated but rapidly degraded, while the target gene is translated separately with only the FRT site (12 amino acids) appended to the N terminus. For C-terminal tags, the off cassette consists of the *let-858* 3′-UTR. This 3′-UTR sequence contains a transcriptional stop motif, terminating transcription before the tag sequence is reached (Davis *et al.* 2008). In both cases, FLP expression induces recombination between the two FRT sites, excising the off cassette and leaving a single FRT site between the native gene and the tag sequence (Figure 4A). The residual FRT site lies in frame with the target gene and tag sequences and encodes a 12-amino-acid flexible linker sequence when translated (GSSYSLESIGTS).

To validate these FLP-dependent tags, we used CRISPR/Cas9 to introduce an FLP-on N-terminal GFP tag at the *rab-3* locus and an FLP-on C-terminal GFP tag at the *snt-1* locus and assayed GFP expression in the absence and presence of FLP recombinase expression. To promote excision of the off cassette, we used a hyperactive variant of FLP recombinase [FLP-D5 (Nern *et al.* 2011)] with two nuclear localization signals appended (M. W. Davis, unpublished data). In the absence of FLP expression, neither strain produced detectable levels of GFP in the nervous system (Figure 4B and Figure S5). Upon pan-neuronal expression of FLP, both strains exhibited strong GFP expression throughout the nervous system, as expected since both of these genes are expressed in all neurons (Stefanakis *et al.* 2015). We conclude that the FLP-on connectors permit recombination-dependent tag expression for both N- and C-terminal tags.

To validate cell specificity of FLP-on GFP induction, we expressed FLP under the control of a variety of more restrictive promoters and assayed the induction of SNT-1::GFP. First, we examined conditional tagging of SNT-1 in GABA neurons. Since SNT-1 is a synaptic vesicle protein, we expressed the synaptic vesicle marker SNB-1::tagRFP in GABA neurons to confirm subcellular localization. Expression of FLP under the *unc-47* promoter (McIntire *et al.* 1997) caused SNT-1::GFP to colocalize with tagRFP at synapses along the ventral nerve cord (Figure 4C), indicating that recombination of the locus was specifically induced in the GABA neurons. In a second experiment, we induced SNT-1::GFP in the acetylcholine neurons by expressing FLP under the control of the *unc-17* promoter. In this strain, SNT-1::GFP localized to different synapses from the GABA synapses. Finally, we tested induction of SNT-1::GFP in the serotonin neurons by expressing FLP under the promoter for *tph-1*. There are only six serotonin neurons in the *C. elegans* hermaphrodite: the bilateral ADF and NSM neurons in the head (Figure 4D) and the HSN neurons flanking the vulva (Sze *et al.* 2000). Expression of FLP under the control of the *tph-1* promoter caused specific expression of SNT-1::GFP in the presynaptic regions of the ADF and NSM neurons in the head (Figure 4E, HSN not detected). Note that the untagged SNT-1 protein is still expressed in all other neurons; it is simply not tagged by GFP in these cells. In all experiments presented in Figure 4,

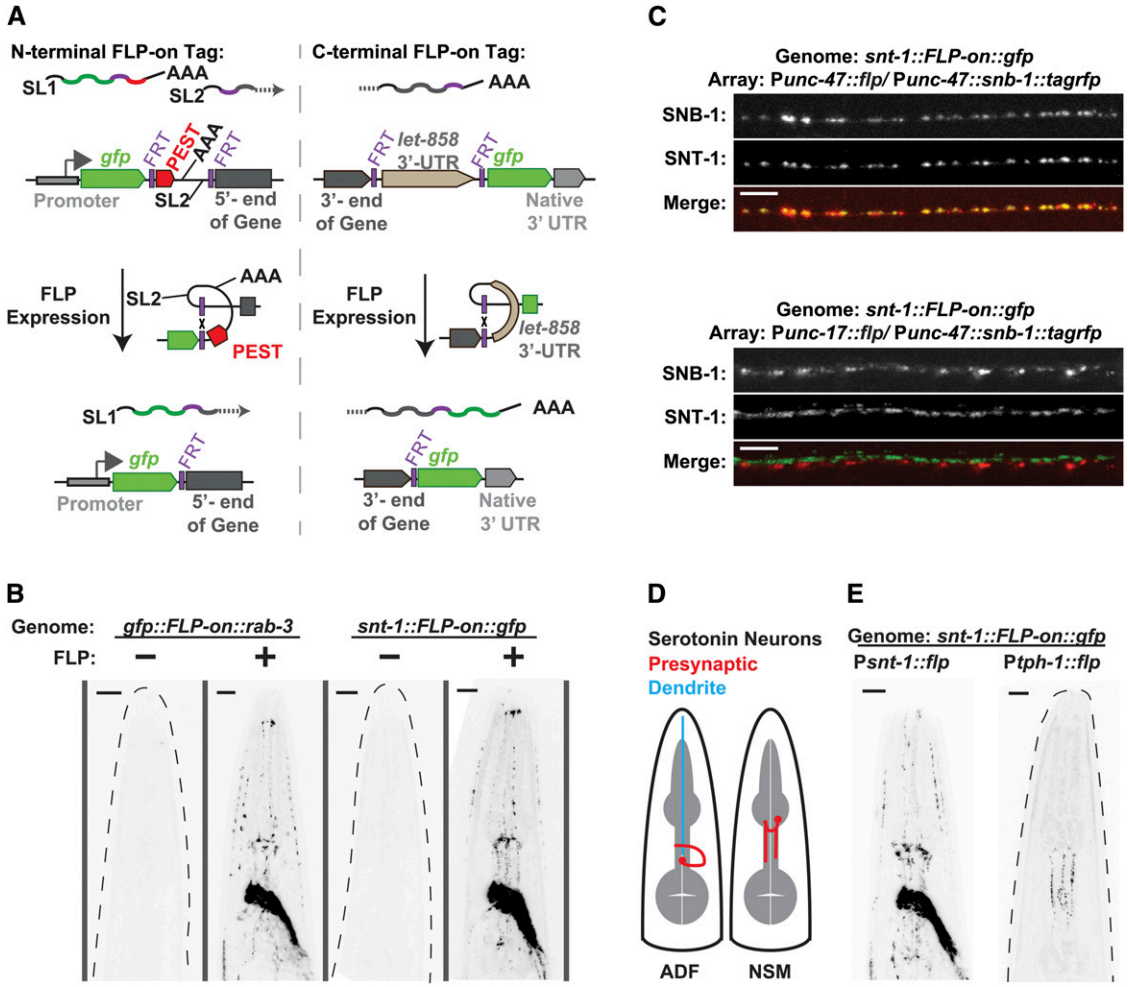


Figure 4 FLP-on expression of fluorescent tags integrated at native loci. FLP-on::GFP tags were inserted and *Cbr-unc-119* was removed as shown in Figure 1A. (A) FLP-on strategies for N-terminal and C-terminal tags. Prior to FLP-based excision, the N-terminal FLP-on connector promotes *trans*-splicing so that the fluorescent tag and PEST sequence occupy a distinct transcript from the gene transcript. The C-terminal FLP-on connector halts transcription before the fluorescent tag sequence is transcribed. In both cases, FLP expression excises the connector sequence, yielding a single transcript encoding the gene fused to the genetic tag by an FRT-encoded linker. (B) Representative fluorescence images of worms harboring *FLP-on*::*gfp* tags fused to either the N terminus of *rab-3* or the C terminus of *snt-1*. FLP was expressed from an extrachromosomal array under the control of the pan-neuronal *snt-1* promoter. Dashed gray lines outline worms that did not exhibit fluorescence. Bars, 10 μm. (C) Representative fluorescence micrographs of the ventral nerve cords of *snt-1*::*FLP-on*::*gfp* worms with FLP expressed in different neuronal subtypes. FLP was expressed in either the GABA (P*unc-47*) or acetylcholine (P*unc-17*) neurons from stable extrachromosomal arrays. To show cell specificity, *snb-1*::*tagrfp* was expressed in the GABA neurons (P*unc-47*) in both strains. Bars, 5 μm. (D) The serotonin neurons in the *C. elegans* head, ADF and NSM. Both neurons exist as bilateral pairs; a single neuron is depicted for clarity. Presynaptic regions are shown in red. ADF has a long dendritic process (blue) that does not contain presynapses. (E) FLP activation of *snt-1*::*FLP-on*::*gfp* in the ADF and NSM neurons. FLP was expressed from extrachromosomal arrays under the control of the *snt-1* or *tph-1* promoters. Bars, 10 μm.

GFP was induced in 99.6% (261/262 animals) of the worms harboring the FLP-expressing array (Table S4). Thus, these FLP-on constructs enable careful analysis of the subcellular localization of proteins without fluorescence appearing in other cells and without the overexpression that is often inherent in classical transgenes.

Discussion

SapTrap is a high-efficiency plasmid assembly protocol and component toolkit for inserting genetically encoded tags at native loci in *C. elegans*. SapTrap produces single plasmid targeting vectors that, when co-injected with a Cas9 expression plasmid, insert genetic tags with high frequency (10–37%). Four design features distinguish the SapTrap method: (1) modular assembly, (2) a library of tags and regulatory sequences, (3) an embedded selectable marker, and (4) short oligo-derived homology arms.

The primary advantage of SapTrap is it allows high-efficiency modular assembly of a single but complex targeting vector. Each targeting vector contains a guide RNA expression cassette and a repair template. The only site-specific reagents required for assembly of the SapTrap targeting construct are the sgRNA targeting sequence and the 5′ and 3′ homology

arms. These sequences are sourced from oligos or other synthetic DNAs, eliminating the need for PCR or molecular cloning during vector assembly. The non-site-specific components are tags, selectable markers, and "connector" modules that are provided as a prebuilt plasmid library. By choosing different combinations of tags and connectors from the library, a variety of functionally distinct tagging vectors can be assembled for a single insertion site. The modular design is particularly useful if target genes need to be coupled to a variety of different tags, such as different fluorescent markers or affinity tags. Finally, assembly is robust and inexpensive: the 2.5-μl reactions produce hundreds of clones, and the non-DNA reagents cost <$1 per reaction.

The library of non-site-specific elements of the repair template is divided into two functional types: the tag and the connector. Tag sequences generally encode proteins and currently include fluorescent tags (GFP, tagRFP) and non-fluorescent tags (Halo, SNAP) (Keppler *et al.* 2003; Los and Wood 2007) useful for fluorescence imaging or biochemical purification. We envision that the library will grow to include other tags such as degrons that can induce proteolytic degradation of the protein leading to functional knockout (Rakhit *et al.* 2014), proximity labeling tags for identifying neighboring proteins or interaction partners (Roux *et al.* 2012; Rhee *et al.* 2013), and MS2 stem loops for tracking mRNAs *in vivo* (Bertrand *et al.* 1998). Connectors are optional DNA modules that fit between the tag sequence and the homology arms and generally control the transcriptional and translational coupling between the target gene and the tag. The connectors include basic flexible peptide linkers for building translational fusions, *trans*-splicing elements for generating transcriptional reporters, and a transcriptional termination sequence for generating fluorescently marked null mutations (Table S5). To address an important limitation of native-locus tags, we developed a set of N- and C-terminal FLP-on connectors that prevent expression of the tag in the absence of FLP recombinase. By expressing FLP in specific cells, the tag can be coupled to the protein in only those cells. By separating the tags from the connector regulatory sequences, the SapTrap library is kept small and the combinatorial utility of the modules is maximized. For example, the single *gfp* donor plasmid in the current library can be inserted internally or at either the N or the C terminus of any gene product and can be coupled to the protein as a constitutive or conditional translational fusion or as a nuclear-localized or cytoplasmic transcriptional reporter, simply by pairing it with different connectors from the existing library.

Successful CRISPR/Cas9 insertions are identified by co-insertion of a selectable marker. Tags in the SapTrap library contain the *C. briggsae unc-119* gene nested in a synthetic intron of the tag, and worms containing SapTrap tag insertions are directly selected by screening for rescue of the injected *unc-119(ed3)* strain. A disadvantage of co-insertion of a selectable marker is that it could affect expression of the target gene. Co-CRISPR avoids selection markers and enriches for successfully injected strains by monitoring events at a second locus (Arribere *et al.* 2014; Kim *et al.* 2014). Targeted events can be identified in co-CRISPR by phenotypic selection or a secondary selection such as expression of GFP. On the other hand, a selectable marker is useful when inserting tags that are undetectable on a fluorescence dissecting microscope, such as Halo and SNAP tags or tags incorporating FLP-on connectors. A feature of the SapTrap constructs is that the *unc-119* selectable marker is inserted in the reverse orientation relative to the target gene, and thus the target gene and the selectable marker can be concurrently expressed. For genes with viable mutant phenotypes concurrent expression is not crucial (Dickinson *et al.* 2015; Norris *et al.* 2015). However, when tagging essential genes for which loss-of-function cannot be tolerated during strain construction, concurrent expression of the target gene and tag is advantageous. Although it was conceivable that this arrangement would lead to colliding RNA polymerases or silencing from double-stranded RNA production in neurons, we observed successful expression of six different neuronal gene constructs containing the inverted *unc-119* selectable marker. However, we did observe moderate interference between the tag and marker genes at one locus. Although weak, the interference suggests that the *unc-119* marker should always be removed after strain isolation. In our work, CRE-mediated excision of the *Cbr-unc-119* marker is efficient; injection of a CRE expression construct generated excisions at all six loci and nearly all successfully injected animals produce *unc-119* progeny due to successful excision. The Goldstein laboratory has recently demonstrated that a heat-shock-inducible CRE recombinase inserted at single copy in the worm genome drives effective germline CRE expression (Dickinson *et al.* 2015). We hope to incorporate this advance in future implementations of our system, negating the need for a second injection step. Finally, we note that the SapTrap assembly protocol is not limited to our novel selection cassette. By removing the *unc-119* marker from the tag donor plasmids, markerless repair templates for co-CRISPR can be built. Alternatively, other recently published syntron-embedded selectable markers (Dickinson *et al.* 2015; Norris *et al.* 2015) can be added to tag donor plasmids, combining the full utility of the SapTrap modular assembly toolkit with these alternative selection strategies, each with its own associated strengths.

Short homology arms of just 30–60 bp were pioneered for co-CRISPR applications by the Seydoux group (Paix *et al.* 2014). We have similarly observed that homology arms of just 57 bp generated insertions at high frequency at most loci tested. An advantage to short homology arms is that they can be generated inexpensively and without PCR by annealing synthetic oligonucleotides. But short arms also mean that the cut site must be very close to the targeted insertion site (in our experiments separated by only 4–20 bp), which limits selection of the guide RNA binding site. The Meyer group recently demonstrated that guide RNAs that bind to sites with a diguanine motif immediately preceding the PAM sequence (called ggNGG guides) are significantly more efficient

at generating double-stranded breaks in the worm genome *in vivo* (Farboud and Meyer 2015). Binding sites containing the ggNGG motif are significantly less common than binding sites containing only the NGG PAM sequence. Within the ranges of the short homology arms we employed, we were unable to locate suitable binding sites conforming to the ggNGG design principle (Table S2). Nonetheless, high insertion rates were achieved at most loci tested, as observed by others using short homology arms (Paix *et al.* 2014). It is possible that high injection concentrations of the targeting construct overcome inefficient cutting. Either higher levels of sgRNA expression may compensate for suboptimal sgRNAs or higher repair template levels may favor insertion events even when cutting is inefficient. In cases where a guide RNA binding site cannot be located near the desired insertion site, SapTrap can accept longer homology arms from alternative synthetic DNA sources or from PCR products. Alternatively, Cas9 variants with altered PAM specificities may increase the availability of guide RNA binding sites within these narrow windows (Bell *et al.* 2015; Kleinstiver *et al.* 2015).

Finally, in addition to the practical advantages of SapTrap for building constructs for modification of individual genes, the SapTrap single vectors will be advantageous for high-throughput applications, for example tagging of hundreds of genes in the genome. It will be simpler to build libraries with a single targeting vector for each gene than to build separate sgRNA and repair template vectors for each targeting event. For high-throughput applications seeking to introduce a specific tag at a large number of loci, the tag can be preassembled in a three-site destination vector (see Figure S1), reducing the complexity of the assembly operation. Coupled with robotic micro-injection (Gilleland *et al.* 2015), it is conceivable that SapTrap vector libraries could be used for genome-wide projects to determine the expression pattern or generate knockouts of all genes in the *C. elegans* genome.

Acknowledgments

We thank Christian Frøkjær-Jensen for suggesting the reverse-oriented *Cbr-unc-119* selectable marker, M. Wayne Davis for contributing the 2xNLS-FLP-D5 construct, and Robert Hobson for contributing the *Punc-47*::*snb-1*::*tagRFP* plasmid. We thank Jamie White, Aude Peden, Eric Bend, and Christian Frøkjær-Jensen for additional plasmid reagents. We thank Christian Frøkjær-Jensen, M. Wayne Davis, and members of the Jorgensen laboratory for critical discussions. This work was supported by National Institutes of Health grant R01 2R01GM095817 (to E.M.J.).

Literature Cited

Alfonso, A., K. Grundahl, J. S. Duerr, H. P. Han, and J. B. Rand, 1993 The Caenorhabditis elegans unc-17 gene: a putative vesicular acetylcholine transporter. Science 261: 617–619.

Arribere, J. A., R. T. Bell, B. X. Fu, K. L. Artiles, P. S. Hartman *et al.*, 2014 Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. Genetics 198: 837–846.

Bell, R. T., B. X. Fu, and A. Z. Fire, 2015 Cas9 variants expand the target repertoire in *Caenorhabditis elegans*. Genetics 202: 381–388.

Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer *et al.*, 1998 Localization of ASH1 mRNA particles in living yeast. Mol. Cell 2: 437–445.

Cho, S. W., J. Lee, D. Carroll, J. S. Kim, and J. Lee, 2013 Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 195: 1177–1180.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819–823.

Davis, M. W., J. J. Morton, D. Carroll, and E. M. Jorgensen, 2008 Gene activation using FLP recombinase in *C. elegans*. PLoS Genet. 4: e1000028.

Dickinson, D. J., J. D. Ward, D. J. Reiner, and B. Goldstein, 2013 Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. Nat. Methods 10: 1028–1034.

Dickinson, D. J., A. M. Pani, J. K. Heppert, C. D. Higgins, and B. Goldstein, 2015 Streamlined genome engineering with a self-excising drug selection cassette. Genetics 200: 1035–1049.

Doudna, J. A., and E. Charpentier, 2014 Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096.

Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet, 2009 Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4: e5553.

Farboud, B., and B. J. Meyer, 2015 Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. Genetics 199: 959–971.

Friedland, A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaiacovo, G. M. Church *et al.*, 2013 Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nat. Methods 10: 741–743.

Frøkjær-Jensen, C., 2013 Exciting prospects for precise engineering of *Caenorhabditis elegans* genomes with CRISPR/Cas9. Genetics 195: 635–642.

Frøkjær-Jensen, C., M. W. Davis, C. E. Hopkins, B. J. Newman, J. M. Thummel *et al.*, 2008 Single-copy insertion of transgenes in *Caenorhabditis elegans*. Nat. Genet. 40: 1375–1383.

Frøkjær-Jensen, C., M. W. Davis, M. Ailion, and E. M. Jorgensen, 2012 Improved *Mos1*-mediated transgenesis in *C. elegans*. Nat. Methods 9: 117–118.

Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys, 2012 Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA 109: E2579–E2586.

Gilleland, C. L., A. T. Falls, J. Noraky, M. G. Heiman, and M. F. Yanik, 2015 Computer-assisted transgenesis of *Caenorhabditis elegans* for deep phenotyping. Genetics 201: 39–46.

Golic, K. G., and S. Lindquist, 1989 The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell 59: 499–509.

Hubbard, E. J., 2014 FLP/FRT and Cre/*lox* recombination technology in *C. elegans*. Methods 68: 417–424.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816–821.

Jinek, M., A. East, A. Cheng, S. Lin, E. Ma *et al.*, 2013 RNA-programmed genome editing in human cells. eLife 2: e00471.

Katic, I., and H. Grosshans, 2013 Targeted heritable mutation and gene conversion by Cas9-CRISPR in *Caenorhabditis elegans*. Genetics 195: 1173–1176.

Katic, I., L. Xu, and R. Ciosk, 2015 CRISPR/Cas9 genome editing in *Caenorhabditis elegans*: evaluation of templates for homology-mediated repair and knock-ins by homology-independent DNA repair. G3 5: 1649–1656.

Keppler, A., S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel *et al.*, 2003 A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. Nat. Biotechnol. 21: 86–89.

Kim, H., T. Ishidate, K. S. Ghanta, M. Seth, D. Conte, Jr. *et al.*, 2014 A co-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. Genetics 197: 1069–1080.

Kleinstiver, B. P., M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen *et al.*, 2015 Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523: 481–485.

Kumar, A., 2009 An overview of nested genes in eukaryotic genomes. Eukaryot. Cell 8: 1321–1329.

Lee, Y. H., X. Y. Huang, D. Hirsh, G. E. Fox, and R. M. Hecht, 1992 Conservation of gene organization and *trans*-splicing in the glyceraldehyde-3-phosphate dehydrogenase-encoding genes of *Caenorhabditis briggsae*. Gene 121: 227–235.

Li, X., X. Zhao, Y. Fang, X. Jiang, T. Duong *et al.*, 1998 Generation of destabilized green fluorescent protein as a transcription reporter. J. Biol. Chem. 273: 34970–34975.

Lo, T. W., C. S. Pickle, S. Lin, E. J. Ralston, M. Gurling *et al.*, 2013 Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 195: 331–348.

Los, G. V., and K. Wood, 2007 The HaloTag: a novel technology for cell imaging and protein analysis. Methods Mol. Biol. 356: 195–208.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell *et al.*, 2013 RNA-guided human genome engineering via Cas9. Science 339: 823–826.

McIntire, S. L., R. J. Reimer, K. Schuske, R. H. Edwards, and E. M. Jorgensen, 1997 Identification and characterization of the vesicular GABA transporter. Nature 389: 870–876.

Nern, A., B. D. Pfeiffer, K. Svoboda, and G. M. Rubin, 2011 Multiple new site-specific recombinases for use in manipulating animal genomes. Proc. Natl. Acad. Sci. USA 108: 14198–14203.

Norris, A. D., H. M. Kim, M. P. Colaiacovo, and J. A. Calarco, 2015 Efficient genome editing in *Caenorhabditis elegans* with a toolkit of dual marker selection cassettes. Genetics 201: 449–458.

Paix, A., Y. Wang, H. E. Smith, C. Y. Lee, D. Calidas *et al.*, 2014 Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in *Caenorhabditis elegans*. Genetics 198: 1347–1356.

Paix, A., A. Folkmann, D. Rasoloson, and G. Seydoux, 2015 High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. Genetics 201: 47–54.

Pujol, N., C. Bonnerot, J. J. Ewbank, Y. Kohara, and D. Thierry-Mieg, 2001 The *Caenorhabditis elegans unc-32* gene encodes alternative forms of a vacuolar ATPase *a* subunit. J. Biol. Chem. 276: 11913–11921.

Rakhit, R., R. Navarro, and T. J. Wandless, 2014 Chemical biology strategies for posttranslational control of protein function. Chem. Biol. 21: 1238–1252.

Rhee, H. W., P. Zou, N. D. Udeshi, J. D. Martell, V. K. Mootha *et al.*, 2013 Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science 339: 1328–1331.

Roux, K. J., D. I. Kim, M. Raida, and B. Burke, 2012 A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196: 801–810.

Stefanakis, N., I. Carrera, and O. Hobert, 2015 Regulatory logic of pan-neuronal gene expression in *C. elegans*. Neuron 87: 733–750.

Sze, J. Y., M. Victor, C. Loer, Y. Shi, and G. Ruvkun, 2000 Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. Nature 403: 560–564.

Tzur, Y. B., A. E. Friedland, S. Nadarajan, G. M. Church, J. A. Calarco *et al.*, 2013 Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. Genetics 195: 1181–1185.

van Schendel, R., S. F. Roerink, V. Portegijs, S. van den Heuvel, and M. Tijsterman, 2015 Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nat. Commun. 6: 7394.

Ward, J. D., 2015 Rapid and precise engineering of the *Caenorhabditis elegans* genome with lethal mutation co-conversion and inactivation of NHEJ repair. Genetics 199: 363–377.

Xu, S., 2015 The application of CRISPR-Cas9 genome editing in *Caenorhabditis elegans*. J. Genet. Genomics 42: 413–421.

Communicating editor: O. Hobert

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.184275/-/DC1

SapTrap, a Toolkit for High-Throughput CRISPR/Cas9 Gene Modification in *Caenorhabditis elegans*

Matthew L. Schwartz and Erik M. Jorgensen

DOI: 10.1534/genetics.115.184275

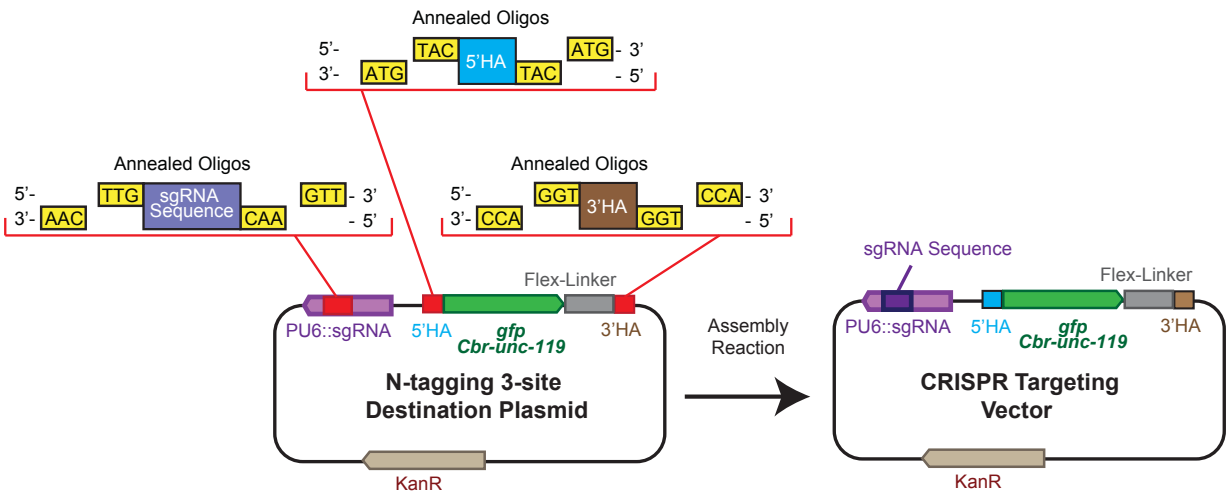


Figure S1. Three-site SapTrap destination plasmid.

For high-throughput applications in which a single tag and connector will be introduced at a large number of loci, the assembly reaction can be simplified by pre-assembling the tag and removable selectable marker cassette and connector module in the destination vector. Annealed oligos encoding the homology arms and sgRNA targeting sequence are supplied to the assembly reaction along with the 3-site destination vector to produce a single plasmid targeting vector. Note that in contrast to the two-site assembly strategy depicted in Figure 2 and Figure S2, the destination vector is kanamycin resistant. The two-site strategy uses a library of KanR plasmids as sources of DNA, and therefore AmpR must be used for selection of the final CRISPR targeting vector. The three-site SapTrap strategy only uses annealed oligos (the connector and tag are pre-assembled in the destination vector), so any antibiotic marker can be used for the destination vector.

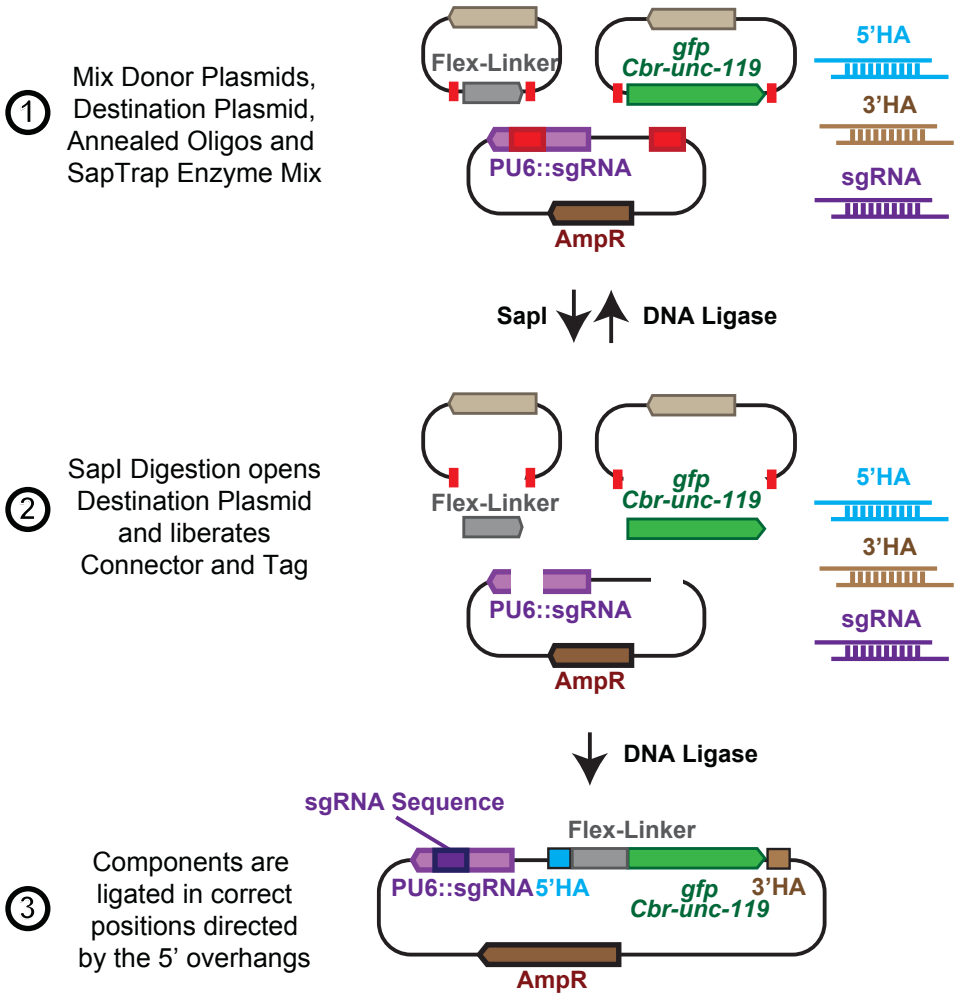


Figure S2. The SapTrap assembly.

SapI cleaves the donor and destination plasmids. These plasmids can undergo futile rounds of digestion and ligation during the assembly reaction. The desired product is "trapped" by ligation; ligation joins in the product lack SapI recognition sequences and are immune to further digestion.

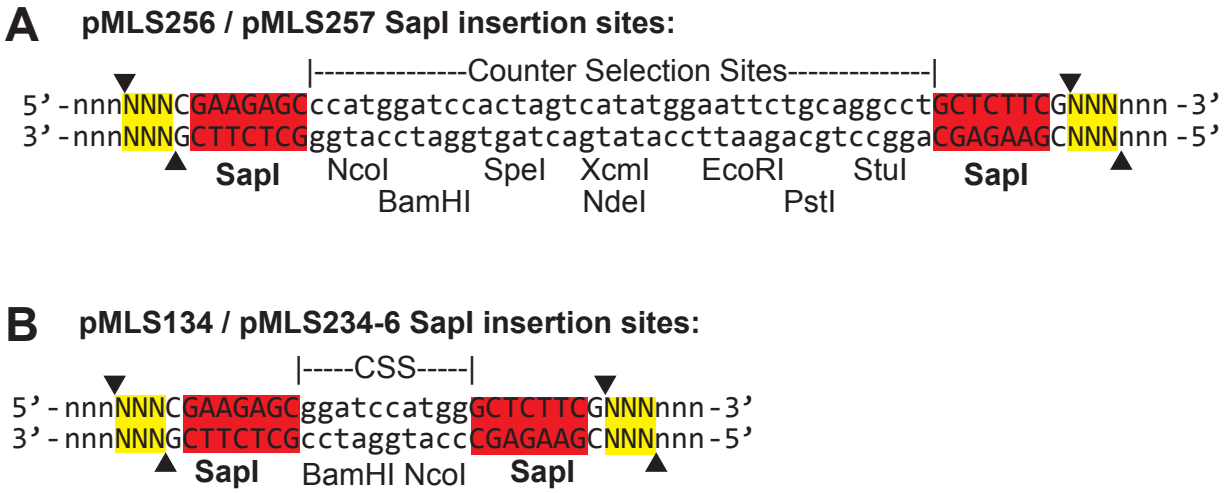


Figure S3. Sequence of the SapI insertion sites.

SapI's recognition sequence (red) is adjacent to the cleavage site (yellow). The SapI insertion sites contain tandem SapI sites oriented so that cleavage of both sites liberates the recognition sites from the plasmid backbone. A series of restriction enzyme recognition sites (counter selection sites, CSS) are encoded between the SapI sites and allow unreacted destination plasmids to be specifically digested at the end of the assembly reaction. **(A)** SapI insertion site in pMLS256 and pMLS257. **(B)** SapI insertion site in pMLS134 and pMLS234-6.

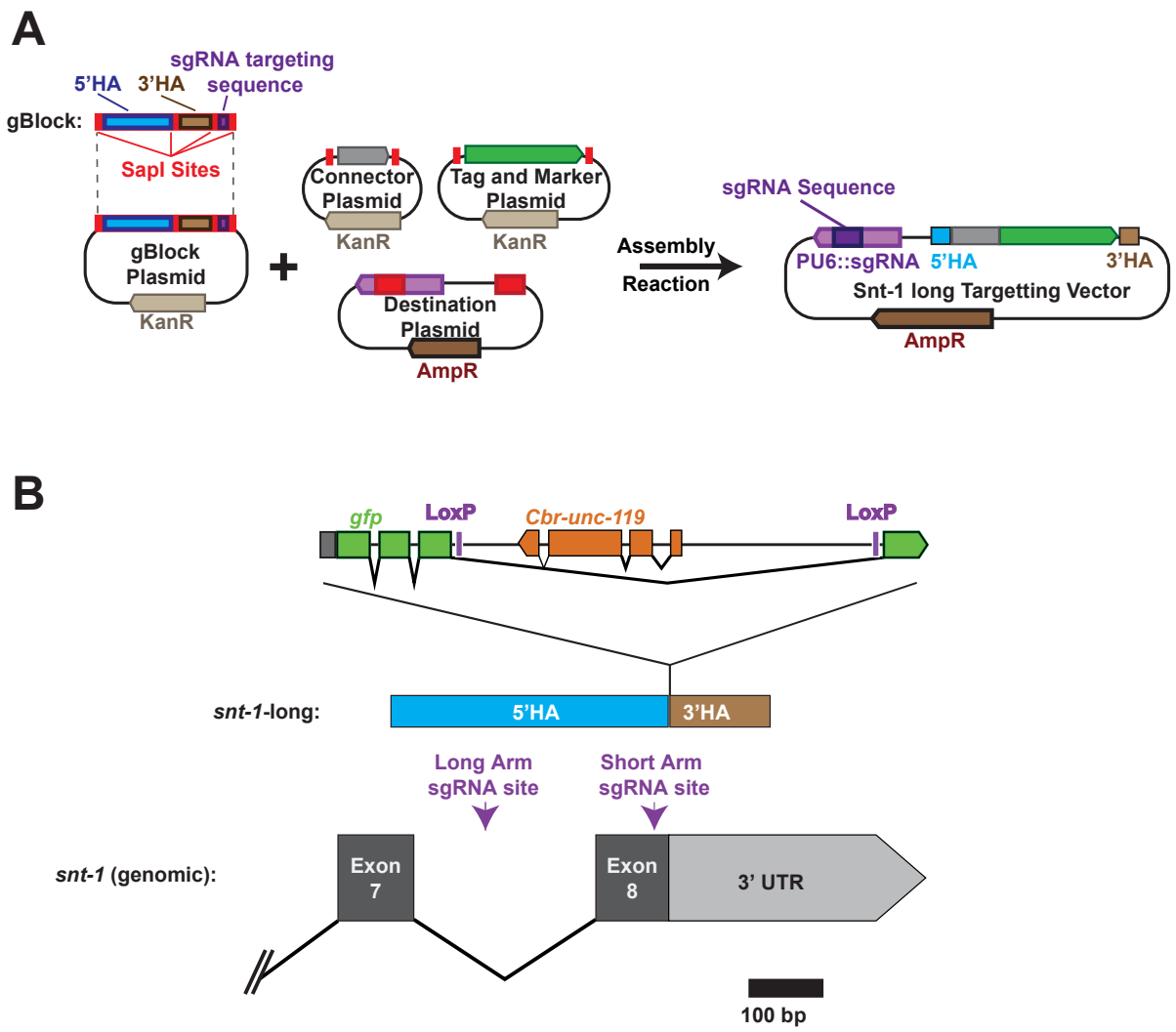


Figure S4. "*snt-1*-long" targeting constructs.

(A) A synthetic "gBlock" DNA fragment encoding the 5' and 3' homology arms and sgRNA targeting sequence, each flanked by SapI sites (red) was cloned into a donor plasmid. This plasmid was included in SapTrap assembly reactions in place of the annealed oligo components to assemble the *snt-1* long targeting vector. **(B)** The genomic structure of the 3' end of *snt-1* is shown. sgRNA cut sites are illustrated with purple arrowheads, and the 5' homology arms and 3' homology arms are illustrated with blue and brown rectangles, respectively.

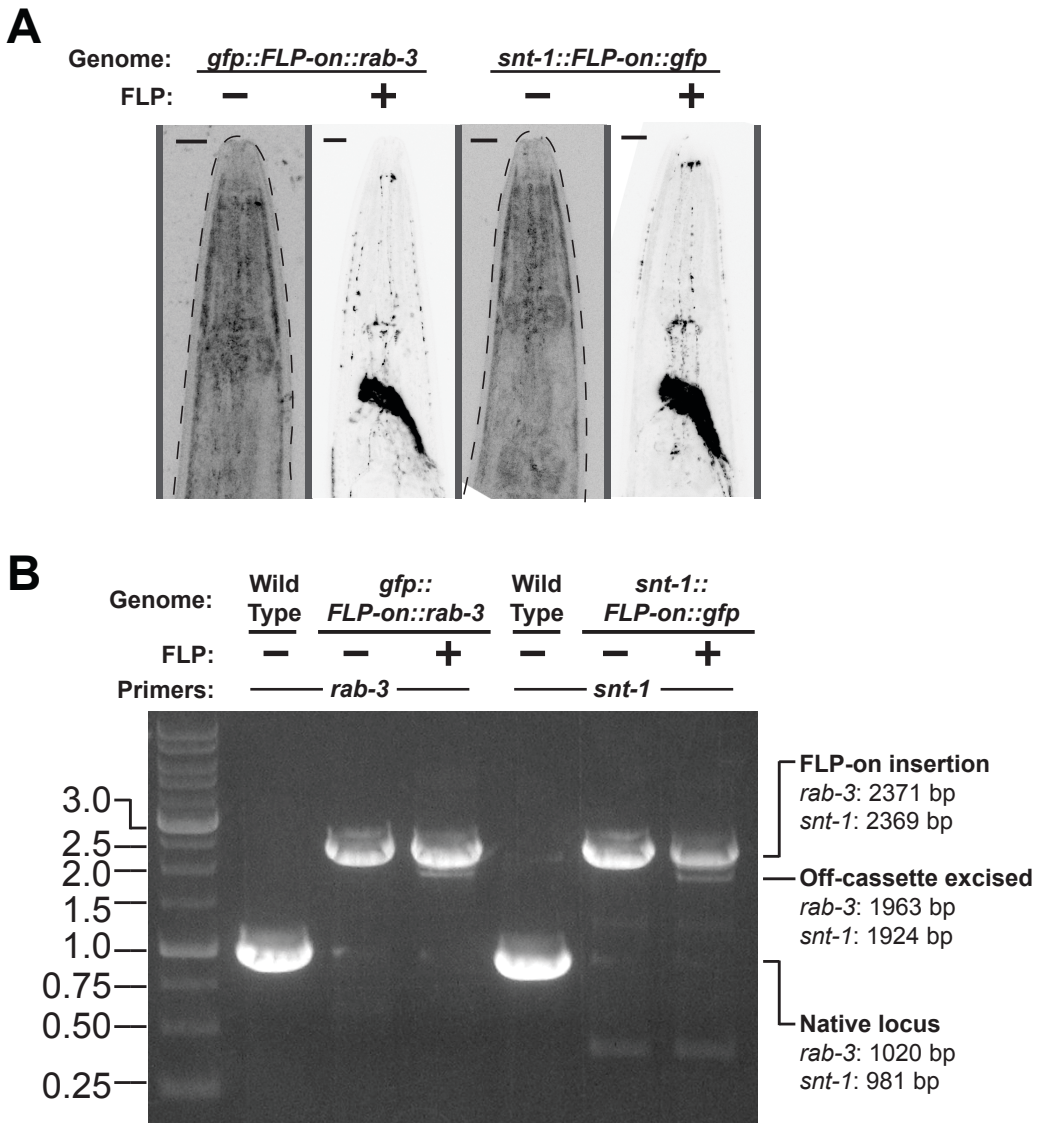


Figure S5. *FLP-on::gfp* strains with and without FLP expression.

A. *FLP-on::gfp* strains do not exhibit detectable fluorescence in the nervous system in the absence of FLP expression. Here, the same fluorescent micrographs shown in Figure 4B have been adjusted to maximize the fluorescence signal. The most densely labeled structure in the FLP expressing strains, the nerve ring, is not detectable in the absence of FLP expression. Scale bars = 10 um. **B.** PCR analysis of *FLP-on::gfp* strains with and without neuronal FLP expression driven by the P*snt-1* promoter. The insertion site was amplified with primers that bind ~500 bp

upstream and downstream of the insertion site. FLP recombination results in the excision of 400 – 450 bases from the inserted sequence. In both cases, a band corresponding to the FLP-excised insertion is only detected in the presence of FLP expression. The intact FLP-on cassette remains the predominant allele, consistent with activation in a subset of tissues.

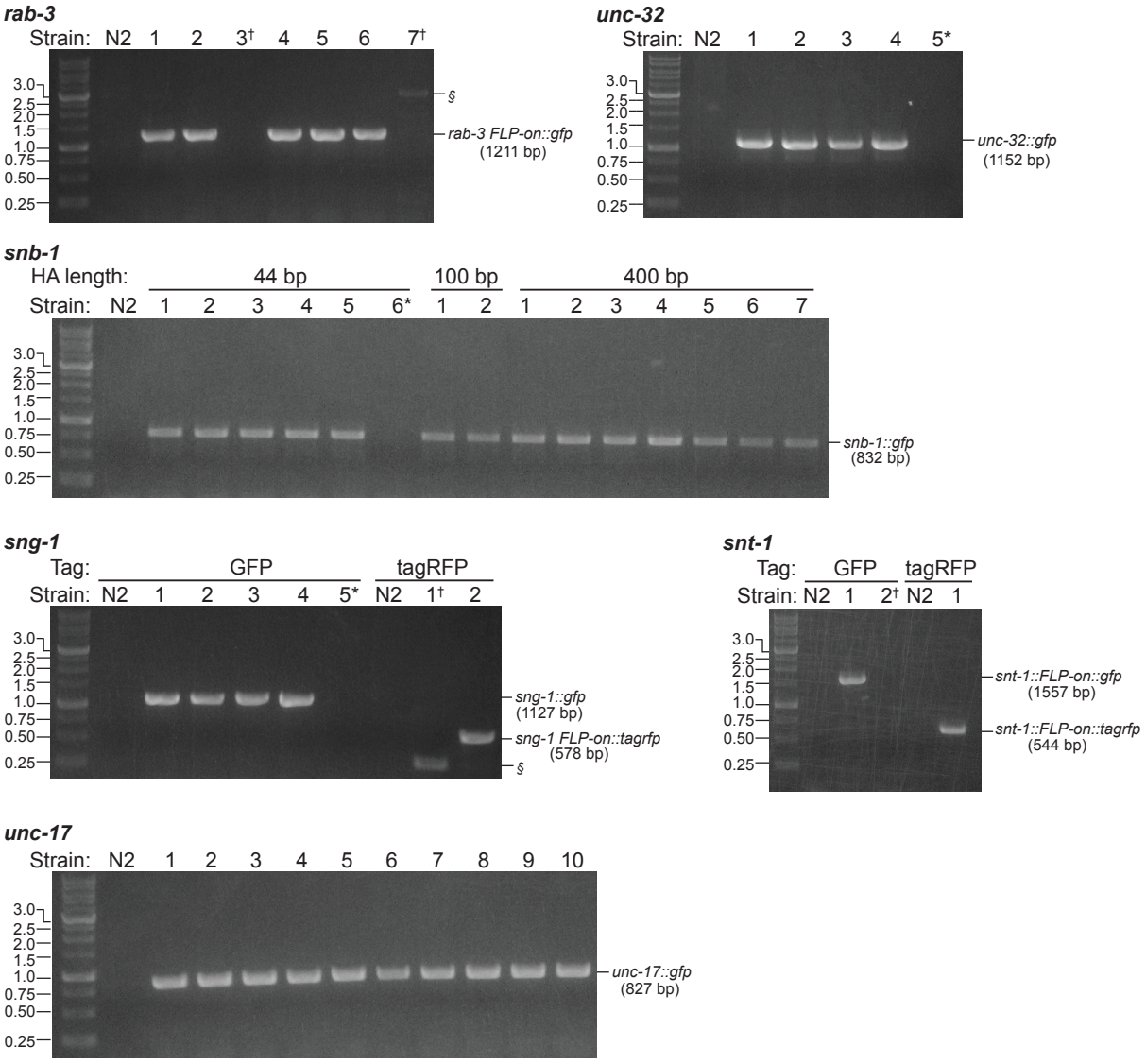


Figure S6. Insertion confirmation by PCR. For every candidate insertion strain isolated from the injections reported in this study (Figure 3 and Table S3), genomic DNA was prepared and tested for the presence of the anticipated insertion by PCR. PCRs were performed using one primer that binds within the intended insert sequence and a second primer that binds in the genome external to the homology arms used to direct insertion. Molecular weight standard sizes are listed in kb. The expected sizes of the PCR products from successful insertions are listed on the right of each gel. Unless noted, each strain exhibited the anticipated fluorescence intrinsically or upon pan-neuronal FLP expression. * = Strain did not exhibit fluorescence

anticipated from insertion allele. † = Strain did not exhibit fluorescence anticipated from the insertion allele upon pan-neuronal FLP expression. § Unanticipated product.

Table S1. Plasmids generated for this study.

Plasmid	Marker	Description
Guide RNA and Repair Template Destination Vectors		
pMLS134	Amp	sgRNA expression cassette only destination vector
pMLS256	Amp	Combined sgRNA / repair template destination vector
pMLS257	Amp	Repair template only destination vector
Tag and Selectable Marker Plasmids		
pMLS252	Kan	GFP::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS254	Kan	Halotag::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS271	Kan	SNAPtag::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS286	Kan	tagRFP::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS291	Kan	mCherry::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS292	Kan	2xNLS-mCherry::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
C-tagging Connector Plasmids		
pMLS268	Kan	SL2 trans-splicing leader
pMLS269	Kan	SL2::SV40-NLS
pMLS272†	Kan	syntron-embedded FLP-on linker (FRT::let-858 3' UTR::FRT)
pMLS279*	Kan	FLP-on linker (FRT::let-858 3' UTR::FRT)
pMLS287	Kan	Flexible Linker (-AGSGGSGGTGGSGM-)
pMLS382	Kan	Flagtag-TEV
N-tagging Connector Plasmids		
pMLS281†	Kan	syntron-embedded FLP-on linker (FRT::PEST-SL2::FRT)
pMLS282*	Kan	FLP-on linker (FRT::PEST-SL2::FRT)
pMLS285	Kan	*egl-13*-NLS
pMLS288	Kan	Flexible Linker (-GGSGGTGGSGGT-)
pMLS383	Kan	TEV-Flagtag
pMLS297	Kan	STOP codon + let-858 UTR
3-Site Destination Vectors		
pMLS234	Kan	N-terminal GFP::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS235	Kan	C-terminal GFP::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
pMLS236	Kan	Internal GFP::syntron-embedded LoxP-flanked reverse *Cbr-unc-119*
Additional Plasmids		
pMLS260	Amp	*Punc-17::2xNLS-FLP-D5::let-858 3'UTR* expression vector (pCFJ150)
pMLS262	Amp	*Psnt-1:: 2xNLS-FLP-D5::let-858 3'UTR* expression vector (pCFJ150)
pMLS359	Amp	*Ptph-1:: 2xNLS-FLP-D5::let-858 3'UTR* expression vector (pCFJ201)
pMLS360	Amp	*Punc-47:: 2xNLS-FLP-D5::let-858 3'UTR* expression vector (pCFJ201)
pMLS280	Kan	Kan-resistant cloning vector for short inserts
pMLS328	Amp	*Peft-3::2xNLS-CRE*

*These FLP-on linkers were used to generate the *FLP-on::gfp* strains presented in Figure 4.

† In these FLP-on linkers, the FRT sites are embedded in added syntron sequences, and a glycine-rich flexible linker has been added.

Table S2. sgRNA targeting sequences used in this study.

Target	Chr	Guide Sequence	PAM	Orientation	Distance	Repair Template Immunization
rab-3	II	(G)CCAGCCGC^{v}**CAT**ctgaaa	ata ggg	Antisense	9	Insertion
snb-1	V	GAGCGTC^{v}**CAT**gtcgtca	aga tgg	Antisense	10	Insertion
sng-1	X	(G)ATCAAACCCATACACTC	AGT CGG$_{14}$^{v}	Sense	20	CGG>CCG (silent)
snt-1	II	^{v}$_{16}$(G)TTATCACCTTCTTCTTC	AAC TGG	Antisense	33	CCA>CTA (P>T)
unc-17	IV	(G)ccaggggagagagaga	g**AT** G^{v}GG	Sense	4	Insertion
unc-32	III	(G)AGAATCTC^{v}**TAA**gatca	cct cgg	Sense	8	Insertion
***snt-1*-long**	II	(G)gcaagtgaccacagaaa	tat ggg$_{271}$^{v}	Sense	277	gggΔGAATTC

Chr = chromosome. Guide targeting sequences are printed 5' – 3'. (G) indicates a non-native G residue was appended to the 5' end of the guide sequence to facilitate U6 promoter transcriptional initiation. Coding residues are printed in capital letters; native start and stop codons are highlighted in green and red, respectively. The PAM sequence is shown separated by a space; PAM sequences were not included in the sgRNA construct. "^{v}" indicates the insertion site; subscripted numbers denote the number of additional bases between the insertion site and the last depicted base. The dashed line shows the predicted cut site at -3 bases from the 3' end of the target sequence. "Distance" refers to the distance, in bases, between the predicted cut site and the insertion site. Repair template immunization indicates the mechanism by which the sgRNA binding site is disrupted in the repair template. In cases labeled "insertion," the tag sequence is inserted into the sgRNA binding site. In other cases, the PAM site was mutated in the manner indicated.

Table S3. Injection and insertion frequencies.

Gene	Tags inserted	Total P0 injected	Successful Injections	Independent Strains	Frequency per P0 Injected	Frequency per Successful Injection	95% CI
***snb-1* : 0 bp HA**	*gfp*	48	28	0	0.0%	0.0%	0.0 – 14.3%
***snb-1* : 44 bp HA**	*gfp*	36	24	5	13.9%	20.8%	8.8 – 40.9%
***snb-1* : 100 bp HA**	*gfp*	34	14	2	5.9%	14.3%	2.8 – 41.2%
***snb-1* :400 bp HA**	*gfp*	42	26	7	16.7%	26.9%	14.5 – 46.3%
rab-3	*FLP-on gfp*	52	25	5	9.6%	20.0%	8.4 – 39.6%
sng-1	*gfp; FLP-on tagrfp*	128	51	5	3.9%	9.8%	3.8 – 21.4%
snt-1	*gfp*	80	36	0	0.0%	0.0%	0.0 – 11.5%
unc-17	*gfp*	60	27	10	16.7%	36.9%	21.5 – 55.8%
unc-32	*gfp*	42	13	4	9.5%	30.7%	12.4 – 58.0%
snt-1 long	*FLP-on gfp; FLP-on tagrfp*	60	29	2	3.3%	6.9%%	0.9 – 23.0%

Successful injections represent the number of P0s that generated *unc-119*-rescued progeny, usually as extrachromosomal arrays. Frequency is the percent of P0s that gave rise to genome insertion events. HA = homology arms. CI = confidence interval.

Table S4. Flp Activation Efficiency.

Strain	Figure	Fluorescent Animals / Number Observed	Activation Efficiency
Genome: *gfp::FLP-on::rab-3* Array: P*snt-1::flp*	4B	53/53	100%
Genome: *snt-1::FLP-on::gfp* Array: P*snt-1::flp*	4B	65/65	100%
Genome: *snt-1::FLP-on::gfp* Array: P*unc-47::flp*	4C	41/41	100%
Genome: *snt-1::FLP-on::gfp* Array: P*unc-17::flp*	4C	66/66	100%
Genome: *snt-1::FLP-on::gfp* Array: P*tph-1::flp*	4E	36/37	97.3%

Array(+) animals were selected by rescue of the *unc-119* phenotype. Animals were anesthetized in 25 mM NaN_3 on an agarose pad and observed on a compound fluorescence microscope. Animals were scored for the presence of GFP fluorescence in the nerve ring and nerve cords (*snt-1*, *unc-47* and *unc-17* promoters) or in the NSM neurons (*tph-1* promoter). We did not detect fluorescence in the HSN neurons in any animals containing the P*tph-1::flp* construct.

Table S5. Annotated list of connector plasmids by function.

Connector	Plasmid	Use
C-terminal tag connectors		
SL2 leader	pMLS268	**Transcriptional reporter.** The SL2 leader directs *trans*-splicing so that the native gene and tag will be spliced into distinct mRNAs and translated independently. A fluorescent tag attached with this connector will mark the cells in which the native gene is expressed.
SL2::SV40	pMLS269	**Nuclear-localized transcriptional reporter.** This connector is similar to pMLS268 but also appends an SV40 NLS to the tag. Used with pMLS285, two NLS sequences can be added to any tag sequence for use as a nuclear localized transcriptional reporter. In our experience, the SV40 NLS alone is insufficient for nuclear localization at native expression levels.
standard FLP-on	pMLS279	**Tissue-specific C-terminal tag.** The tag will only be fused to the native protein in tissues in which FLP is expressed.
syntron-embedded FLP-on	pMLS272	**Tissue-specific C-terminal tag.** This connector is functionally equivalent to pMLS279, but the design is compatible with variant FRT sites and recombination sites for other recombinases (such as LoxP sites for use with Cre recombinase). Many of these alternate recombination sites encode stop codons in all potential reading frames or encode amino acids that are undesirable for use in a peptide linker. In this connector, the FRT sites are embedded in a syntron and are not translated.
Flexible Linker	pMLS287	**Simple tagging.** This encodes a simple flexible linker; the tag will be fused to the gene product in every tissue in which the gene is natively expressed.
Flagtag-Tev	pMLS382	**Affinity isolation.** This connector will be useful for biochemical purification of proteins and complexes using the tag as an affinity handle. Isolated proteins and complexes can be eluted from the affinity column by TEV protease; the flagtag will remain with the native protein after TEV-site cleavage and can be used to track the protein by western blotting.
N-terminal tag connnectors		
Stop Codon + *let-858* 3'-UTR	pMLS297	**Fluorescently marked null.** This connector will add a stop codon and transcriptional stop sequence after the tag. Placed at the 5' end of a

		gene, this will result in transcription of the tag only, producing a null allele that expresses the isolated tag in the tissues in which the native gene is expressed.
egl-13 NLS	pMLS285	**Transcriptional reporter.** This connector adds the *egl-13* NLS to any tag. Used with pMLS269, any tag can be expressed as a nuclear localized transcriptional reporter.
standard FLP-on	pMLS282	**Tissue-specific N-terminal tag.** The tag will only be fused to the native protein in tissues in which FLP is expressed.
syntron-embedded FLP-on	pMLS281	**Tissue-specific N-terminal tag.** This connector is functionally equivalent to pMLS282, but the design is compatible with variant FRT sites and recombination sites for other recombinases. See above description of pMLS272.
Flexible Linker	pMLS288	**Simple tagging.** This encodes a simple flexible linker; the N-terminal tag will be fused to the gene product in every tissue in which the gene is natively expressed.
TEV-Flagtag	pMLS383	**Affinity isolation.** This connector will be useful for biochemical purification of proteins and complexes using the tag as an affinity handle. See above description of pMLS382.

Supplemental Protocol I: Oligo Design for SapTrap Vectors

1. **Identify an insertion site.** For N- and C-terminal tags, insert the tag immediately after the start codon or immediately before the stop codon, respectively. Figure P1 depicts oligo design for a C-terminal tag on *sng-1*. Figure P2 depicts oligo design for an N-terminal tag on *unc-17.*

2. **Obtain the genomic DNA sequence flanking the insertion site.** Minimally, obtain 60 bases on each side of the insertion site. If there are no suitable sgRNA binding sites near the insertion site, a larger region will need to be considered. Check the genomic sequence for SapI recognition sites. The genomic sequence of the sgRNA binding site must not contain a SapI recognition site; SapI recognition sites outside of the sgRNA binding site can be removed from the homology arms by point mutation.

3. **Identify candidate sgRNA binding sites.** Every "GG" and "CC" sequence represents an "NGG" PAM site and corresponds to a candidate sgRNA binding site. Cas9 cleaves the genomic DNA within the sgRNA binding site 3 bases before the PAM sequence. Ideal sgRNAs for use with SapTrap cleave the genomic DNA within 25 bases of the insertion site. For *sng-1*, there are 4 candidate sgRNA binding sites that direct cleavage within 25 bases of the insertion site, labeled 1 – 4 (Fig. P1A). For *unc-17*, there are 8 candidates (1-8) that direct cleavage within 25 bases of the desired insertion site (Fig. P2A).

4. **Select a high quality sgRNA binding site.** We use the following criteria to select among candidate sites, in this order:
   - **(a) Location:** Select sgRNA binding sites that direct cleavage within 25 bases of the desired insertion site, if possible. Bases between the insertion site and the cleavage site do not serve as part of the homology arm for repair. If you must use an sgRNA binding site that directs cleavage further than 25 bases from the insertion site, the homology arm that contains the binding site should be extended by using longer oligos, synthetic DNA fragments or by generating the homology arm by PCR (described below). We recommend a minimum of 35 bases of homology extend past the cleavage site on each side.
   - **(b) Ability to mutate:** We always ensure that the repair template cannot be targeted by the selected sgRNA. Ideally, select an sgRNA binding site that straddles the insertion site. These sgRNA binding sites will be disrupted by the insert sequence, so no additional mutation is required. If a binding site that straddles the insertion site is not available, mutate the PAM sequence or sgRNA binding site in the repair template. Mutation or deletion of the PAM sequence will render the binding site inactive; if this cannot be done with silent mutations, introducing several silent mutations within the sgRNA binding site should render the site inactive.
   - **(c) Specificity:** We evaluate sgRNA specificity using the Zhang lab's sgRNA design tool at http://crispr.mit.edu/. This tool accepts input sequences of up to 500 bases and reports all available sgRNA binding sites within the input sequence. Additionally, the sgRNAs are compared to a selected reference genome and evaluated for specificity within that genome. Individual sgRNAs are given a specificity score of 0 – 100; higher scores indicate guides with fewer potential off-target binding sites. We have had success with sgRNAs scoring as low as 74. In Figs. P1A and P2A, these specificity scores are printed in parentheses after each sgRNA candidate.

   To target *sng-1*, we chose sgRNA candidate 1 (Fig. P1A). Candidate 4 was rejected for its low specificity score. Candidate 1 was selected from the remaining 3 candidates because the PAM site for this binding site could be eliminated by a silent G-to-C mutation. To target *unc-17*, we chose sgRNA candidate 1 (Fig. P2A). Candidate 1 was selected because the insertion of the tag sequence separates the PAM sequence from the sgRNA binding site – no mutation is required to render the repair template immune to sgRNA binding.

5. Design oligos. (Figs. P1B and P2B). We prefer 57 base pair homology arms because these can be generated from 60-mer oligos; longer oligos are more expensive and are more likely to contain errors. However, we have successfully used up to 80-mer oligos (77 base pair homology arms). Below, (b) and (c) describe the production of 57 base pair homology arms. To design longer oligos, simply extend the homology arm selection beyond 57 bases, as needed. (e) describes the design of homology arms too long to be produced by annealing oligos.

(a) Introduce necessary mutations to the genomic sequence file. These include PAM site or sgRNA binding site mutations required to immunize the repair template against the sgRNA. If the genomic region contains SapI recognition sites, disrupt them with silent mutations. In the rare case that the selected sgRNA binding site contains a SapI recognition site, select a different sgRNA. For *sng-1*, we disrupted the PAM sequence with a silent GtoC mutation (Fig. P2B). For *unc-17*, no mutation was required to disrupt sgRNA binding in the repair template (Fig. P2B). Neither of these regions contain SapI sites.

(b) Design the 5' homology arm oligos. Copy the 57 bases immediately preceding the insertion site. The "top" oligo is "TGG" followed by this 57 base sequence. The "bottom" oligo is either "CGC" (C-terminal tag) or "CAT" (N-terminal tag) followed by the reverse complement of the 57 base homology arm sequence.

(c) Design the 3' homology arm oligos. Copy the 57 bases immediately following the insertion site. The "top" oligo is either "GGT" (C-terminal tag) or "ACG" (N-terminal tag) followed by this 57 base sequence. The "bottom" oligo is "TAC" followed by the reverse complement of the 57 base homology arm sequence.

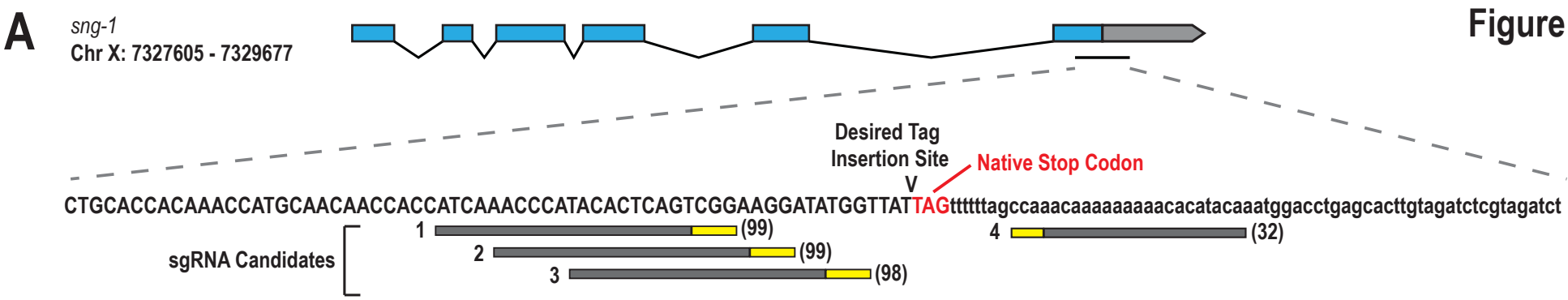
NOTE: All SapTrap tags contain an initial ATG in the 3-base overhang sequence but lack a stop codon. For N-terminal tags, the native "ATG" start codon should not be included in either homology arm. For C-terminal tags, you must include the native stop codon as part of the 3' homology arm.

(d) Design the sgRNA oligos. Copy the 20 bases immediately preceding the "NGG" PAM sequence. The "top" oligo is "TTG" followed by this 20 base sequence. The "bottom" oligo is "AAC" followed by the reverse complement of the 20 base sequence. The PAM sequence is not included in the sgRNA construct.

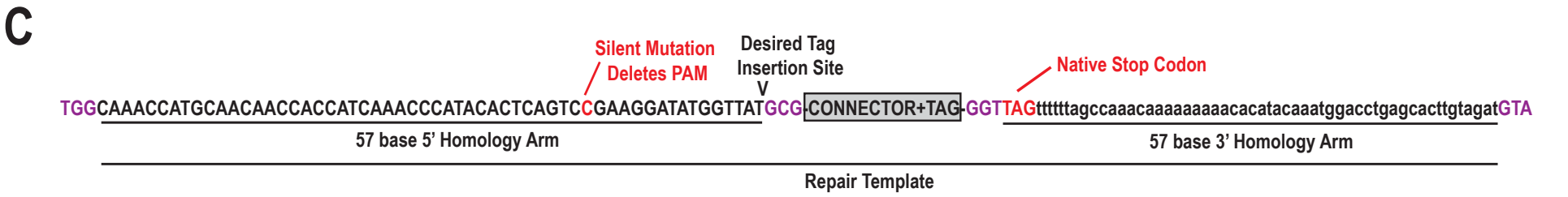
(e) Design longer homology arms. If you are forced to use a homology arm longer than ~80 bases, you will need to supply the homology arm as a purified PCR product or as synthetic double-stranded DNA. These types of DNA will require the addition of SapI recognition sequences on each side of the homology arm oriented so that SapI digestion produces the necessary 5' overhangs. Fig. P1D depicts a set of primers that would produce the 57 base pair 5' homology arm for *sng-1* by PCR, with the appended SapI recognition sites. Techniques such as overlap PCR may be needed to introduce PAM site or sgRNA binding site mutations to longer homology arms.

6. Check the oligo sequences. Make sure that the oligo pairs will anneal to produce a product of 57 base-pairs flanked by the correct 3-base 5'-overhangs on each end (Figs. P1B and P2B). Oligos will assemble with connector segments and tag segments from the SapTrap donor plasmid library to generate the desired repair templates (Figs. P1C and P2C).

Figure P1


A
sng-1
Chr X: 7327605 - 7329677
Desired Tag
Insertion Site
V
Native Stop Codon
CTGCACCACAAACCATGCAACAACCACCATCAAACCCATACACTCAGTCGGAAGGATATGGTTATTAGttttttagccaaacaaaaaaaaacacatacaaatggacctgagcacttgtagatctcgtagatct
sgRNA Candidates
1 (99)
2 (99)
3 (98)
4 (32)


B
20 base sgRNA binding site
Silent Mutation
Deletes PAM
Desired Tag
Insertion Site
V
CTGCACCACAAACCATGCAACAACCACCATCAAACCCATACACTCAGTCCGAAGGATATGGTTATTAGttttttagccaaacaaaaaaaaacacatacaaatggacctgagcacttgtagatctcgtagatct
57 base 5' Homology Arm
57 base 3' Homology Arm
Appended 3-base 5' overhang sequences
Annealed Product
5' Homology Arm Oligos
Top oligo: 5' - TGGCAAACCATGCAACAACCACCATCAAACCCATACACTCAGTCCGAAGGATATGGTTAT
Bottom oligo: 5' - CGCATAACCATATCCTTCGGACTGAGTGTATGGGTTTGATGGTGGTTGTTGCATGGTTTG
TGG 5' HA CGC
3' Homology Arm Oligos
Top oligo: 5' - GGTTAGttttttagccaaacaaaaaaaaacacatacaaatggacctgagcacttgtagat
Bottom oligo: 5' - TACatctacaagtgctcaggtccatttgtatgtgtttttttttgtttggctaaaaaaCTA
GGT 3' HA CAT
sgRNA Oligos
Top oligo: 5' - TTGATCAAACCCATACACTCAGT
Bottom oligo: 5' - AACACTGAGTGTATGGGTTTGAT
TTG sgRNA CAA


C
Silent Mutation
Deletes PAM
Desired Tag
Insertion Site
V
Native Stop Codon
TGGCAAACCATGCAACAACCACCATCAAACCCATACACTCAGTCCGAAGGATATGGTTATGCG-CONNECTOR+TAG-GGTTAGttttttagccaaacaaaaaaaaacacatacaaatggacctgagcacttgtagatGTA
57 base 5' Homology Arm
57 base 3' Homology Arm
Repair Template

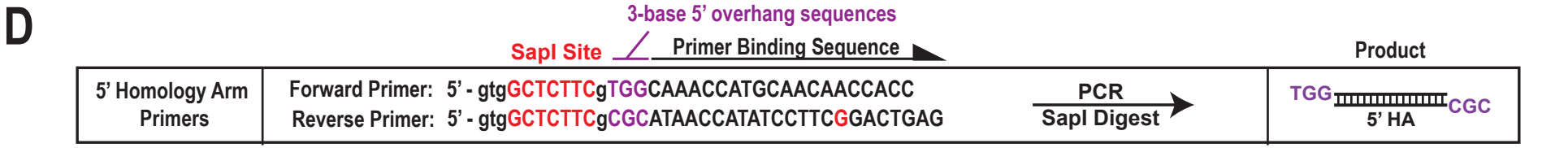
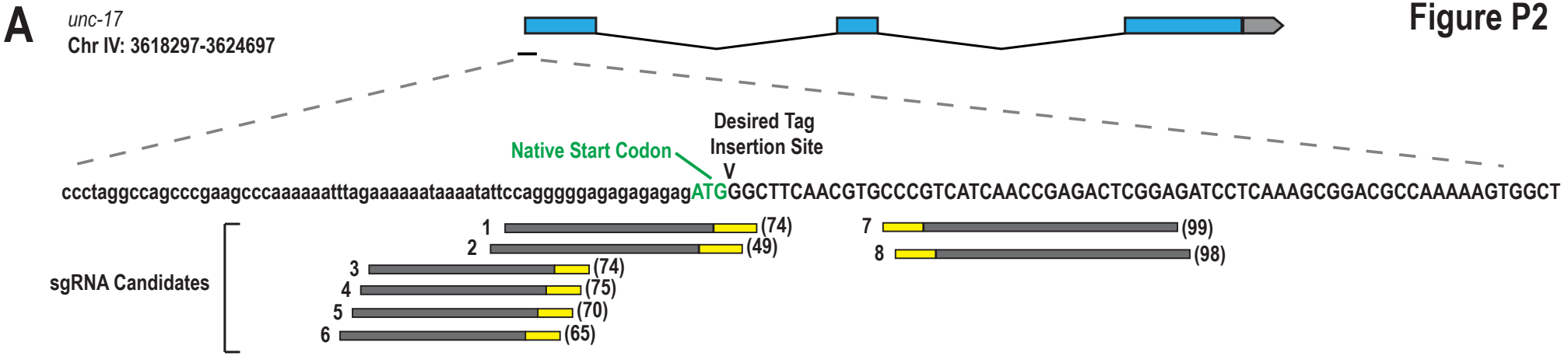
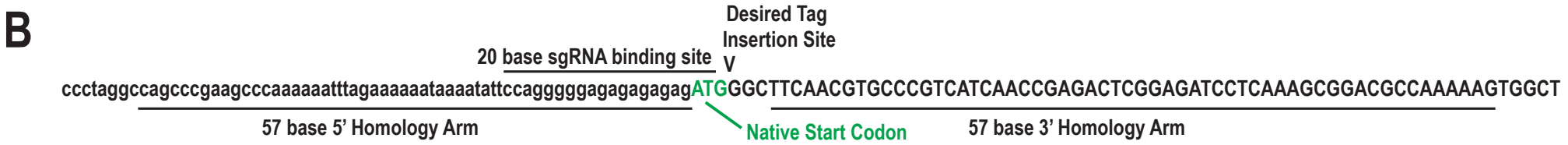

D
3-base 5' overhang sequences
SapI Site
Primer Binding Sequence
Product
5' Homology Arm Primers
Forward Primer: 5' - gtgGCTCTTCgTGGCAAACCATGCAACAACCACC
Reverse Primer: 5' - gtgGCTCTTCgCGCATAACCATATCCTTCGGACTGAG
PCR
SapI Digest
TGG 5' HA CGC

Figure P2





Appended 3-base 5' overhang sequences

		Annealed Product
5' Homology Arm Oligos	Top oligo: 5' - TGGcagcccgaagcccaaaaaatttagaaaaaataaaatattccaggggagagagagag Bottom oligo: 5' - CATctctctctcccccctggaatattttatttttttctaaatttttttgggcttcgggctg	TGG 5' HA TAC
3' Homology Arm Oligos	Top oligo: 5' - ACGTTCAACGTGCCCGTCATCAACCGAGACTCGGAGATCCTCAAAGCGGACGCCAAAAAG Bottom oligo: 5' - TACCTTTTTGGCGTCCGCTTTGAGGATCTCCGAGTCTCGGTTGATGACGGGCACGTTGAA	ACG 3' HA CAT
sgRNA Oligos	Top oligo: 5' - TTGccaggggagagagagagAT Bottom oligo: 5' - AACATctctctctcccccctgg	TTG sgRNA CAA

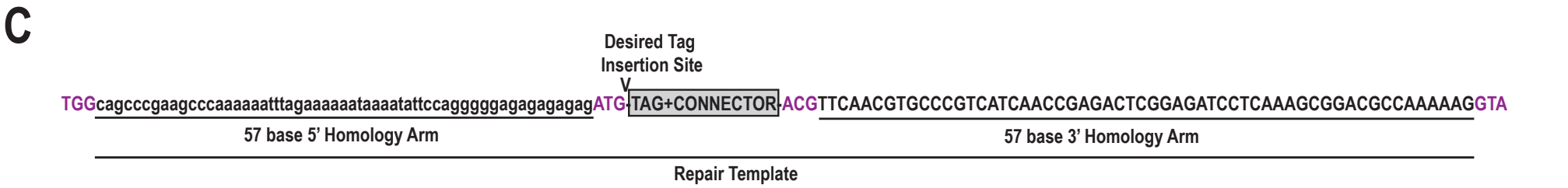


Figure P1. A. *sng-1* C-terminal insertion site. The native genomic sequence is shown for the desired insertion site flanked by 65 bases on each side. sgRNA candidates are modeled with gray bars representing the 20-base binding site and yellow bars representing the NGG PAM sequence. The number on the left of the sgRNA model provides a unique identifier for the sgRNA candidate, the number in parentheses on the right reports the specificity score from the Zhang lab CRISPR design tool (http://crispr.mit.edu/). **B.** Oligo design. The same genomic region depicted in A is shown with the addition of a PAM-site disrupting mutation for the selected sgRNA. The sequences used to generate the homology arm and sgRNA oligos are marked. **C.** Sequence of the repair template generated from the oligos depicted in B. The sgRNA cassette produced by insertion of the sgRNA oligos is not depicted. **D.** Example primers for building a homology arm by PCR. The depicted primers will produce the same 5' homology arm depicted in B and C.

Figure P2. A. *unc-17* N-terminal insertion site. The native genomic sequence is shown for the desired insertion site flanked by 65 bases on each side. sgRNA candidates are modeled with gray bars representing the 20-base binding site and yellow bars representing the NGG PAM sequence. The number on the left of the sgRNA model provides a unique identifier for the sgRNA candidate, the number in parentheses on the right reports the specificity score from the Zhang lab CRISPR design tool (http://crispr.mit.edu/). **B.** Oligo design. The same genomic region depicted in A is shown. The sequences used to generate the homology arm and sgRNA oligos are marked. **C.** Sequence of the repair template generated from the oligos depicted in B. The sgRNA cassette produced by insertion of the sgRNA oligos is not depicted.

Supplemental Protocol II: SapTrap Assembly

1. Buffers and Reagent list

- *10x Oligo Annealing Buffer (OAB):* 200 mM Tris-Cl pH 7.5, 500 mM NaCl, 10 mM $MgCl_2$.
- *SapTrap Donor and Destination Plasmids:* Dilute all plasmid preps to 50 nM. To convert a plasmid concentration in ng/uL to a molar concentration, use the formula:

$$[\text{plasmid}]\ (\text{nM}) = 1.5 * \frac{[\text{plasmid}]\ (\frac{\text{ng}}{\text{uL}})}{\text{plasmid size (Kb)}}$$

- *SapTrap Enzyme Mix:* Prepare in advance and store at -80C in 2 uL aliquots in PCR tubes.

Table P1. SapTrap Enzyme Mix Recipe

Component	20 Reactions	50 Reactions	100 Reactions
10x Cutsmart Buffer:	5 uL	12.5 uL	25 uL
H_2O:	22.5 uL	56.3 uL	112.5 uL
10 mM ATP:	5 uL	12.5 uL	25 uL
1 M DTT:	0.25 uL	0.63 uL	1.25 uL
400 U / uL T4 DNA ligase:	1.25 uL	3.1 uL	6.25 uL
10 U / uL T4 polynucleotide kinase:	1 uL	2.5 uL	5 uL
10 U / uL SapI*:	5 uL	12.5 uL	25 uL

* Thoroughly resuspend SapI before withdrawing from stock tube. SapI settles from solution during storage.

- *Restriction enzymes for counter-selection of unreacted destination vector.*
- *Oligos*

  - M13F: 5' – TGTAAAACGACGGCCAGT
  - M13R: 5' – CAGGAAACAGCTATGACCATG
  - oMLS471: 5' – TCCAAGAACTCGTACAAAAATGCTC

2. Anneal Oligos

- Resuspend oligo stocks to 100 uM in TE or water.
- Mix complmementary oligo pairs to a final concentration of 10 uM each in 100 uL of 1x OAB.
- Heat oligo mixtures to >95°C in heat block; incubate for 2 - 5 minutes.
- Turn off heat block and allow block to slowly cool to room temperature.
- 10 uM annealed oligo stocks can be stored at -20C for future use, or diluted and used immediately.
- Dilute the annealed oligos in TE to 150 nM. Dilute all 3 annealed oligo pairs for a given reaction in a single aliquot of TE, such that the final solution contains 150 nM of each of the 3 annealed oligo species.

3. SapTrap Assembly

- Combine and thoroughly mix by pipetting:
  - 1 uL of 50 nM Destination Vector (generally pMLS256)
  - 1 uL of 50 nM Connector Vector
  - 1 uL of 50 nM Tag and Marker Plasmid
  - 1 uL of annealed, diluted oligo mixture (150 nM each oligo species)
  - 1 uL of dH_2O

  For internal tags including 2 connectors, replace the 1 uL of dH_2O with 1 uL of the 50 nM stock of the second connector plasmid.
- Thaw a 2 uL aliquot of SapTrap Enzyme Mix.
- Add 0.5 uL of the DNA mixture to the 2 uL SapTrap Enzyme Mix aliquot.
- Mix thoroughly by pipetting up and down with a 2 uL pipette.
- Incubate the reaction overnight at 25°C.
- The next day, heat inactivate the T4 DNA ligase by incubation at 65°C for 30 minutes.
- Prepare a solution of 1x Cutsmart buffer + 1 – 2 U / uL of an appropriate counter-selection restriction enzyme. Choose an enzyme that does not cut the desired final product.
- Add 2.5 uL of the 1x Cutsmart + Enzyme solution to each 2.5 uL SapTrap reaction and mix.
- Incubate at 37°C for 1 hour.
- Use 1 – 2 uL of the reaction solution to transform chemically competent *e. coli*.
- Plate 25 – 50% of the transformed *e. coli* on selective media.

4. Identification of successful clones

- You may culture individual colonies and screen plasmid clones by restriction digest; however, because the efficiency of proper assemblies is variable and sometimes low (~20% – 90%), we find it is more efficient to initially screen colonies by colony PCR.
- Setup 5 – 10 uL colony PCR reactions with M13F and M13R primers. This will amplify the entire repair template region. The product will be ~3 - 4 Kb for most tagging constructs. The PCR may be inhibited if too many *E. coli* cells are added to the reaction; sample colonies sparingly.
- For the colony PCRs we use Phusion polymerase (NEB) with buffer HF + 3% DMSO, 58°C annealing and 1:45 extension at 72°C for 30 cycles.
- Successful clones produce a strong product at 3 – 4 Kb. Occasionally, unsuccessful clones will produce a weak product near 3Kb accompanied by a strong band at a lower size (Fig. P3). Score these as unsuccessful.
- Prepare plasmid DNA from 2 - 3 successful clones. If the M13F + M13R PCR gives a positive result, >90% of the time the sgRNA sequence has also been correctly incorporated.
- Sequence plasmids with M13F, M13R, and oMLS471. Together, these three sequencing reactions will cover both homology arms and the sgRNA cassette.

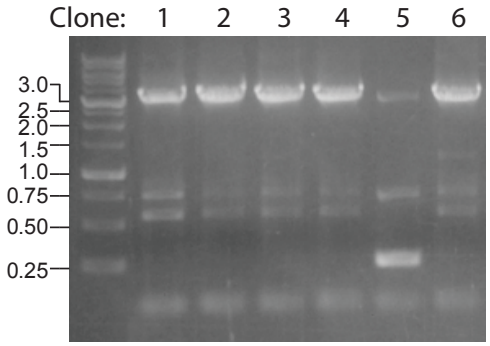


Fig. P3. Example of M13F + M13R colony PCR gel. DNA ladder sizes are reported in Kb. All colonies except colony #5 are correct. Note that there is a faint product of ~3 Kb in the colony 5 PCR. It is weak compared to the true positives and is accompanied by a strong product at a smaller size.

Supplemental Protocol III: Injection and Isolation

1. Injection animals. Culture *unc-119(ed3)* animals at 15°C on HB101 bacteria. The night before injection, pick L4 animals to a new HB101 plate and incubate at 15°C – 22°C overnight.

2. CRISPR Injection. Inject 30 to 60 P0 animals with the following injection mix. Following injection, place 2 – 3 P0 animals each on OP50 plates and incubate at 25°C for 7 – 10 days.

Table P2. Insertion Injection Mix Recipe

Component	Concentration	Note
Combined repair template and sgRNA expression vector	65 ng / uL	-
Cas9 expression vector	25 ng / uL	Addgene plasmid #46168 (Peft-3::cas9-SV40_NLS::tbb-2 3'UTR from John Calarco)
Fluorescent co-injection markers*	10 ng / uL	**For GFP and non-fluorescent insertions**, we use 2 ng / uL pCFJ90 (Pmyo-2::mCherry, Addgene plasmid #19327), 4 ng / uL pCFJ104 (Pmyo-3::mCherry; Addgene Plasmid #19328) and 4 ng / uL pGH8 (Prab-3::mCherry; Addgene plasmid #19359) **For tagRFP and mCherry insertions**, we use 2 ng / uL pCFJ420 (Pmyo-2::GFP::H2B; Addgene plasmid #34877) and 8 ng / uL pCFJ421 (Peft-3::GFP::H2B; Addgene plasmid #34876)
Total	100 ng / uL	-

* We avoid array negative selection markers; we identify array(-) animals by the lack of fluorescence from the co-injection markers. If desired, the recipe can be amended to include an array negative selection marker such as P*hsp::peel-1*.

3. Select inserts. Starting 7 days post-injection, screen the plates under a fluorescence dissecting microscope to identify motile (*unc-119(+))* animals lacking the co-injection markers.

- Insertions arise as early as post-injection day 7, but more arise through post-injection day 10. Some plates lack obvious insertions on day 7 but harbor them by day 10.

- Insertions arise on plates that also contain array(+) animals. On successful plates, the fraction of motile animals on the plate that are array(-) varies widely: putative insertions may be rare (a few animals per plate) or may be the most prevalent phenotype on the plate. It is most efficient to screen plates quickly to identify plates with high numbers of putative inserts and to more thoroughly screen the plates only if needed.

- The insertion itself can be used as a secondary phenotype in many cases. Because of the short homology arms, the tag is generally not expressed from the repair template. Tag expression will only occur upon successful insertion in most cases.

- Single and propagate putative insertion strains. Treat all putative insertions originating from a single plate as a single strain, as it is impossible to know if they arose from independent repair events. Primary isolates can be homozygous or heterozygous for the insertion; marker removal is best performed with a homozygous strain.

4. Remove the marker.

- Inject homozygous young adult animals with an injection mix containing 50 ng / uL Cre expression plasmid (pDD104; Addgene plasmid #47551 or pMLS328) and 50 ng / uL stuffer + a fluorescent co-injection marker. We usually use 2 ng / uL pCFJ90 (Pmyo-2::mCherry) as the co-injection marker and 48 ng / uL pBluescript II as stuffer DNA for a final injection mix concentration of 100 ng / uL.

- Select array(+) F1 animals. Place 5 – 10 array(+) animals each on up to 4 OP50 plates.

- Screen the F2 and select *unc-119(-)* animals. Generally, there will be several to dozens of *unc-119(-)* animals per plate. Both pDD104 and pMLS328 are *unc-119(+)*, so *unc-119(-)* animals lack the array and have undergone marker excision.

- The resultant strain can be outcrossed against wild-type to remove the ed3 allele. Alternatively, the ed3 allele can be retained for selective purposes. For example, all of the FLP expression vectors presented in this study are *unc-119(+).* This provides a convenient means to select and maintain FLP expression arrays for FLP-on experiments.