Modular safe-harbor transgene insertion (MosTI) for targeted single-copy and extrachromosomal array integration in *C. elegans*.

Sonia El Mouridi[1], Faisal Alkhaldi[1], and Christian Frøkjær-Jensen[1*]

Abstract

Efficient and reproducible transgenesis facilitates and accelerates research using genetic model organisms. Here we describe a modular safe harbor transgene insertion (MosTI) for use in *C. elegans* which improves targeted insertion of single-copy transgenes by homology directed repair and targeted integration of extrachromosomal arrays by non-homologous end-joining. MosTI allows easy conversion between selection markers at insertion site and a collection of universal targeting vectors with commonly used promoters and fluorophores. Insertions are targeted at three permissive safe-harbor intergenic locations and transgenes are reproducibly expressed in somatic and germ cells. Chromosomal integration is mediated by CRISPR/Cas9, and positive selection is based on a set of split markers (*unc-119*, *hygroR,* and *gfp*) where only animals with chromosomal insertions are rescued, resistant to antibiotics, or fluorescent, respectively. Single-copy insertion is efficient using either constitutive or heat-shock inducible Cas9 expression (25 - 75%) and insertions can be generated from a multiplexed injection mix. Extrachromosomal array integration is also efficient (7 - 44%) at MosTI landing sites or at the endogenous *unc-119* locus. We use short-read sequencing to estimate the plasmid copy numbers for eight integrated arrays (6 to 37 copies) and long-read Nanopore sequencing to determine the structure and size (5.4 Mb) of one array. Using universal targeting vectors, standardized insertion strains, and optimized protocols, it is possible to construct complex transgenic strains which should facilitate the study of increasingly complex biological problems in *C. elegans*.

The ability to modify the genome of cells and model organisms with high precision has made great advances in the past decade, primarily due to the ease of genetic engineering with CRISPR/Cas9 (Jinek *et al.* 2012; Cong *et al.* 2013; Mali *et al.* 2013). In *C. elegans*, early experiments demonstrated that Cas9 can be expressed from plasmids (Friedland *et al.* 2013; Tzur *et al.* 2013; Waaijers *et al.* 2013; Chen *et al.* 2013; Dickinson *et al.* 2013), mRNA (Lo *et al.* 2013; Chiu *et al.* 2013; Katic and Großhans 2013), or injected directly as a protein (Cho *et al.* 2013; Paix *et al.* 2015) and that templated repair is efficiently mediated from injected plasmids. Subsequent improvements have led to the use of short repair homology (Paix *et al.* 2014), partially single-stranded repair templates (Dokshin *et al.* 2018), improved sgRNA design (Farboud and Meyer 2015), and novel genetic (e.g., Kim *et al.* 2014; Arribere *et al.* 2014; Ward 2015; Farboud *et al.* 2019) and transgenic selection strategies (e.g., Dickinson *et al.* 2015; Schwartz and Jorgensen 2016). These advances have led to the routine use of CRISPR/Cas gene editing to create custom alleles and tagging genes at their endogenous loci.

However, in some cases the goal is to create stable expression of transgenes to determine the effect of regulatory elements on gene expression (Merritt *et al.* 2008), visualize cell populations (Yemini *et al.* 2021), measure cell activity (Kerr *et al.* 2000), or mark genomic regions (Frøkjær-Jensen *et al.* 2014). Initial transgenesis in *C. elegans* was based on transgene expression from extrachromosomal arrays (Stinchcomb *et al.* 1985; Mello *et al.* 1991; Mello and Fire 1995). Extra-chromosomal arrays are formed by injecting linear or supercoiled DNA into the germline syncytium resulting in large hereditary DNA structures that contain many copies of the injected DNA (Stinchcomb *et al.* 1985) propagated as one or more ring chromosomes (Woglar *et al.* 2020). Arrays have the advantage that they are easy to generate and can drive high transgene expression in somatic cells but have the disadvantages that inheritance is unstable with loss occurring during mitosis (Lin *et al.* 2021), expression is mosaic (Okkema *et al.* 1993), and arrays are frequently silenced in germ cells (Kelly *et al.* 1997). Some of these problems can be solved by integrating arrays into the genome by X-ray or gamma irradiation but integration is inherently mutagenic, labor intensive, and does not fully solve the problem of variable expression (Mello and Fire 1995). More recently, alternative methods for integrating arrays have been developed: reactive oxygen species generated by high-intensity blue light can trigger random array integration (Noma and Jin 2018) and Cas9 can be used for targeted array integration into the *dpy-3* locus (Yoshina *et al.* 2016). However, neither methodology has become widely adopted, possible due to the requirement for a specialized light-source or the mutant phenotype of integrated strains, respectively.

In some cases, expression at near-endogenous levels or expression in the germline is desired, which can be achieved by reducing the number of plasmids. Low-copy number plasmid integrations are primarily generated by biolistic transformation or by using a Mos1 transposon. Biolistic transformation works by bombarding animals with DNA-coated particles, a method that is scalable and can overcome germline silencing (Praitis *et al.* 2001). Biolistic transformation has the disadvantages that a variable number of plasmids are integrated, the insertion site is difficult to determine, and integration may results in chromosomal aberrations (Tyson *et al.* 2018). Targeted single-copy insertions can be generated based on the exogenous Mos1 DNA transposon from *Drosophila* which can transpose in *C. elegans* (Bessereau *et al.* 2001; Robert and Bessereau 2007; Vallin *et al.* 2012). The first widely used single-copy insertion method, Mos1-mediated single-copy insertion (MosSCI), relied on excising transposons in neutral genomic environments and transgene insertion by homology-directed repair from an extra-chromosomal array (Frøkjær-Jensen *et al.* 2008). Biolistic insertion and MosSCI were developed using *unc-119* selection (Maduro and Pilgrim 1995) and subsequently expanded to antibiotic resistance markers (Giordano-Santini *et al.* 2010; Semple *et al.* 2010; Radman *et*

[1]King Abdullah University of Science and Technology (KAUST), Biological and Environmental Sciences and Engineering Division (BESE), Thuwal 23955–6900, Kingdom of Saudi Arabia. [*]Correspondence: cfjensen@kaust.edu.sa

al. 2013). Later iterations of Mos1-mediated transgenesis included additional insertion sites and negative selection (Frøkjær-Jensen *et al.* 2012), and random single-copy insertion by transposition enabling a set of universal MosSCI landing sites (Frøkjær-Jensen *et al.* 2014). Single-copy transgene insertions only rely on double-strand DNA breaks and can, therefore, also be generated by CRISPR/Cas9 (Dickinson *et al.* 2013). More recent iterations of single-copy insertion techniques have incorporated specialized landing sites that contain promoters (Silva-García *et al.* 2019), self-selection based on a split antibiotic selection marker (Stevenson *et al.* 2020), or recombinase-mediated cassette exchange (Nonet 2020).

Each of the above methods have their advantages and several are continuously improved (El Mouridi *et al.* 2020; Nonet 2021) yet there is still considerable room for improvement. For example, CRISPR/Cas9-based insertion of transgenes requires different targeting vectors for every location and rely on selection protocols with substantial screening (Dickinson *et al.* 2013). The universal MosSCI sites have the disadvantage that landing sites were generated at random, and the insertion frequency is lower (Frøkjær-Jensen *et al.* 2014). Positive and negative selection markers facilitate the identification of insertions but still require a considerable amount of manual screening and rarely completely avoid false positives (McDiarmid *et al.* 2020; El Mouridi *et al.* 2021). The split antibiotic selection marker facilitates screening but is only compatible with a single antibiotic and genomic location (Stevenson *et al.* 2020). Frequently, it is also not possible to excise the selection marker used to insert transgenes which limits the complexity of transgenic strains that can be generated.

Here, we develop a method we have named Modular safe-harbor Transgene Insertion ("MosTI") that improves single-copy and array integration by synthesizing many incremental improvements developed across laboratories and by developing a collection of standardized and validated reagents. MosTI is based on identifying safe-harbor insertion sites located at autosome centers in intergenic regions characterized by permissive chromatin modifications (Ho *et al.* 2014). The method was designed to be highly modular: insertion sites can be rapidly generated and existing sites are easily converted to a different selection marker. MosTI landing sites are compatible with a collection of universal targeting vectors that contain common promoters and fluorophores. The insertion sites were engineered to use a set of split selection markers that obviate the need for negative selection markers as only insertions result in rescue (*unc-119*), fluorescence (*gfp*), or antibiotic resistance (*HygroR*). Finally, MosTI is compatible with both single-copy transgene and targeted array insertion at high frequency. We have validated and characterized several landing sites and developed protocols that allow many single-copy insertions by heat-shocking animals from a single injection. Finally, we use MosTI to resolve plasmid copy-number in integrated arrays and the detailed structure of one array by long-read sequencing.

Results

Engineered landing sites and split selection markers enable easy identification of insertions.

We set out to develop a modular transgenesis system where insertions are easily identified, new insertion sites can be readily generated, and the selection scheme can be modified. Furthermore, we wanted the ability to integrate single-copy transgenes and extra-chromosomal arrays at well-defined locations. As a first step, we developed a strategy for inserting single-copy transgenes into "safe-harbor" landing sites based on rescuing *unc-119*(*ed3*) animals using a split *cbr-unc-119*(+) selection scheme (**Figure 1A**), an approach that is similar to the split antibiotic selection used by Stevenson *et al*. (2020). The first part of the rescue marker is at the landing site and the second part is located on the target plasmid adjacent to the transgene; neither part can in isolation rescue the *unc-119* phenotype. Injecting plasmids expressing Cas9 and an efficient sgRNA (Moreno-Mateos *et al.* 2015) generates a DNA double-strand break (DSB) at the landing site. The target plasmid contains homology to the DSB site, and homology-directed repair reconstitutes a full rescuing *cbr-unc-119(+)* marker, while also inserting a single copy of the transgene. If desired, the reconstituted split *cbr-unc-119* selection marker can be excised after transgene insertion by expressing Cre recombinase. This approach greatly facilitates identification of single-copy insertions: we have observed no false positives and just a few rescued animals are easily visible among a population of Unc animals (see next section). Although not required, we typically enrich for non-rescued transgenic animals with extrachromosomal arrays using fluorescent markers and hygromycin selection.

The use of engineered landing sites is useful for several reasons. First, synthetic landing sites allowed us to generate three MosTI landing sites on different autosomes that are compatible with the same target vector and sgRNA (**Figure 1B**). One landing site is near the commonly used *ttTi5605* site on Chr. II (Frøkjær-Jensen et al., 2008), whereas the other sites were chosen for permissive chromatin marks (Ho *et al.* 2014) and minimal disruption to endogenous genes. We inserted landing site in a two-step process: an initial landing site was inserted using a short synthetic *unc-119(+)* selectable marker flanked by LoxP sites; a second injection removed the selection marker by expressing Cre recombinase, leaving behind the final landing site with the split *cbr-unc-119* selection (**Supplementary Figure S1**). This two-step process is efficient and transgenes to generate novel landing sites are easily generated in a one-pot Golden-Gate reaction with a synthetic gene fragment (Engler *et al.* 2008) (**Supplementary Figure S2**). Second, the engineered landing sites are all compatible with a single targeting plasmid. This allowed us to generate standard target vectors compatible with restriction cloning, Golden-Gate cloning, and three-fragment Gateway cloning (**Figure 1C**). We also generated an expanded set of MosTI-compatible cloning vectors: four target vectors with fluorophores (*gfp* and *mScarlet)* for transcriptional and translational fusions (El Mouridi *et al.* 2017), four vectors with a *gpd-2*::fluorophore cassette for transgene co-expression, and three vectors with promoters for ubiquitous (P*eft-3*) (Wheeler *et al.* 2016), germline (P*mex-5*) (Zeiser *et al.* 2011), and neuronal (ultra-pan-neuronal, UPN promoter) (Yemini *et al.* 2021) expression (**Figure 1C** and **Supplementary Figure S3**). We performed whole-genome sequencing on two injection strains (Chr. I: CFJ77 and Chr. II: CFJ42) to identify possible passenger mutations present in the initial *unc-119*(*ed3*) strain or generated by landing site insertion (**Supplementary Table 1**). All plasmids are available from Addgene, and the universal target vector has been deposited with a gene synthesis company for direct gene synthesis instead of standard

cloning, which is becoming increasingly convenient and cost-efficient.

We tested MosTI efficiency for single-copy insertions of a plasmid encoding a ubiquitously expressed GFP (P*eft-3*::*gfp*). We chose the *eft-3* promoter because single-copy insertions are bright in somatic cells whereas germline expression is sensitive to silencing in repressive chromatin domains (Frøkjær-Jensen *et al.* 2016). We generated single-copy insertions using constitutive Cas9 expression (P*smu-2::Cas9*) at relatively high frequencies (25-75%)(**Figure 2A**). We also inserted a different transgene (P*rpl-7*::*gfp*) by inducible expression of Cas9 (P*hsp-16.21::Cas9*) using an improved heat-shock protocol (Nonet 2020). These efficiencies are similar to MosSCI and other single-copy transgenesis methods (Frøkjær-Jensen *et al.* 2012; Dickinson *et al.* 2013; Nonet 2020) but do not capture the ease of identifying transgene insertions: we readily observed *unc-119* rescued animals in the F2 or F3 generation (for both protocols) or for many generations after injection using the heat-shock protocol. All rescued animals behaved as expected: the inserted fluorescent marker was expressed, transgenic animals could be rendered homozygous, the extra-chromosomal array was rapidly lost when animals were grown without selection, PCR and Sanger sequencing validated insertion at the intended MosTI landing site, and the *cbr-unc-119* selection marker could be removed by Cre expression without losing the inserted transgene. All P*eft-3::gfp* transgenes inserted at the three MosTI landing sites were expressed at similarly high levels, and all transgenic strains showed GFP expression in the germline (13 of 13

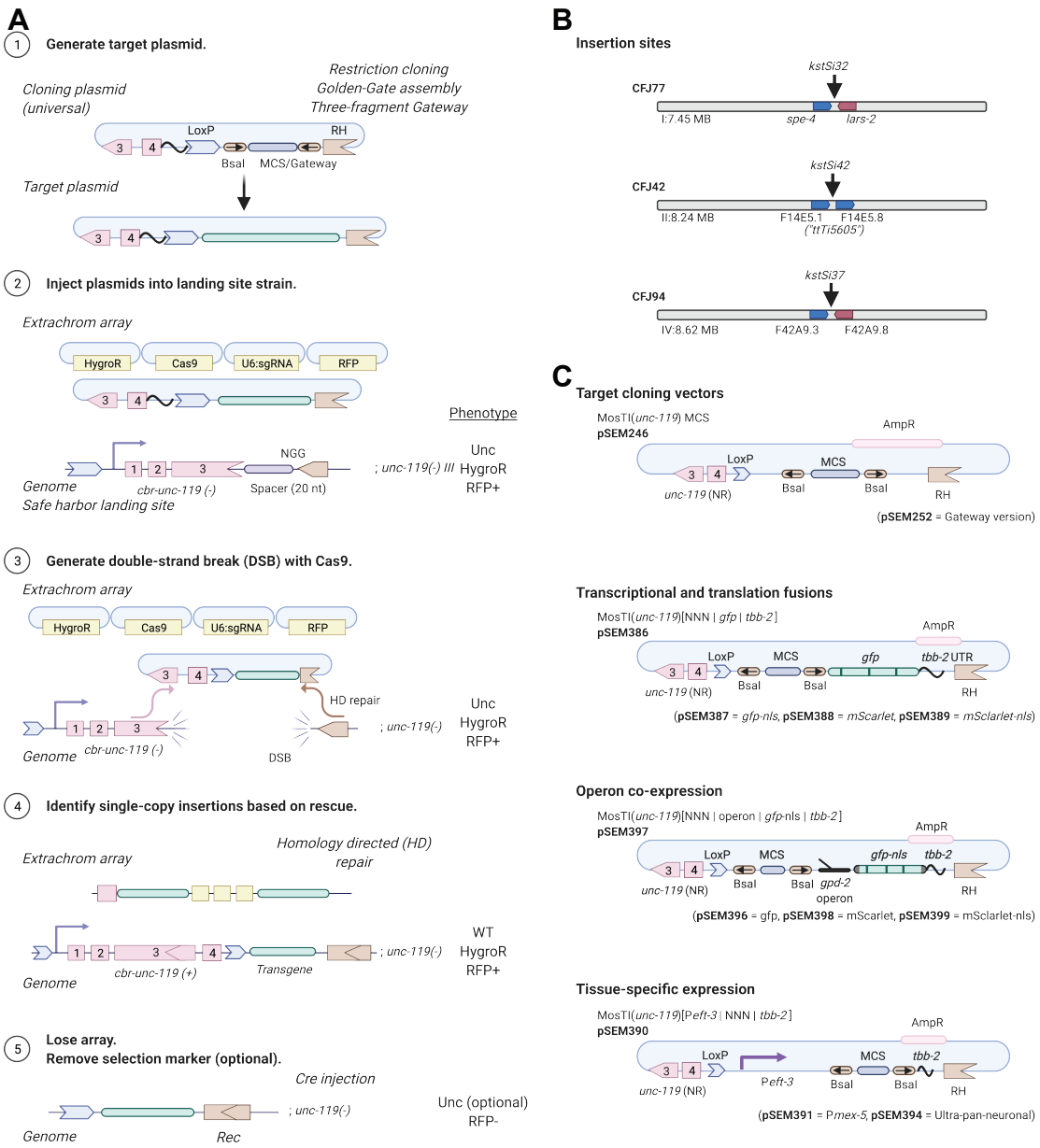


Figure 1 | Schematic of Modular Safe-harbor Transgene Insertion (MosTI) protocol
A. 1. Target transgenes are generated by cloning (standard restriction-enzyme, Gateway, Golden Gate, or gene synthesis) into a MosTI-compatible plasmid containing a non-rescuing *cbr-unc-119* rescue fragment. 2. Target plasmids are co-injected with Cas9 and sgRNA plasmids into *unc-119(ed3)* animals containing a MosTI landing site. Antibiotic or fluorescent markers can be used to enrich the population for non-rescued transgenic array animals (optional). 3. Cas9-induced double-strand breaks are repaired by homology-directed repair from the target plasmid, leading to transgene insertion and reconstitution of a functional *cbr-unc-119(+)* gene. 4. Animals with single-copy insertions are identified by phenotypic rescue (N2-like animals on plates with Unc animals). 5. The extrachromosomal array is lost, and the *cbr-unc-119* selection can be removed by Cre expression (optional). **B**. MosTI-compatible insertion sites in safe-harbor landing sites were selected for permissive chromatin environment (Ho *et al.* 2014). **C**. A collection of MosTI-compatible cloning vectors for tissue-specific gene expression, transcriptional and translational fluorescence expression, and fluorescence co-expression with a *gpd-2* operon has been deposited with Addgene (see **Supplementary Figure 3** for all plasmids).

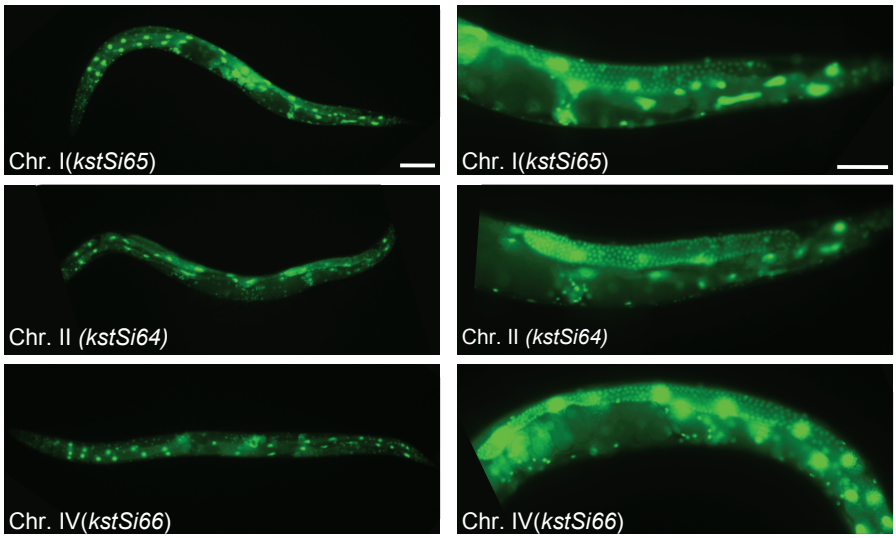
strains), confirming the permissive chromatin environment (**Figure 2B-C**).

The heat-shock protocol (Nonet 2020) requires an additional heat-shock step relative to constitutive expression of Cas9. However, heat-shock induction has the advantage of being highly scalable and requires injection of fewer animals (*i.e.*, a few transgenic animals with extra-chromosomal arrays are sufficient to generate many independent single-copy insertions). The heat-shock protocol may be particularly useful for the injection of collection of different plasmids simultaneously ("multiplexed MosTI") and inserting a single copy from the extra-chromosomal array (**Figure 3A**). Conceptually, such an approach could be used to screen many regulatory elements, for example, a library of enhancers or 3' UTRs coupled to a fluorescent reporter, as previously demonstrated by Kaymak *et al*. (2016). To test the feasibility of multiplexed MosTI, we injected a pool of three target plasmids with visually distinct expression patterns (P*rpl-7A::gfp*, P*eft-3::mMaple*, and P*eft-3::mScarlet*) together with a heat-shock inducible Cas9 plasmid (**Figure 3A**). We established four independent transgenic strains with extra-chromosomal arrays and heat-shocked a single NGM plate of each strain for seven successive generations. In most generations, we were able to generate single-copy insertions form three of the four strains. We generated a total of 29 independent single-copy insertions from these three strains and observed no obvious decline in efficiency over time (**Figure 3B**). In some cases, we were able to isolate animals with every type of transgene insertion from a single NGM plate (**Figure 3C**), demonstrating the feasibility of scaling the approach further. Several observations were notable: first, only one of 29 transgene insertions was non-fluorescent, probably caused by an indel in the transgene, as previously observed (Frøkjær-Jensen *et al.* 2008). Second, one of the four transgenic strains yielded no insertions in any generation. We did not determine the exact cause but speculate that the target spacer at the MosTI site was destroyed by NHEJ in an early generation or that this array was somehow not permissive for Cas9 expression. Third, not all transgenes were inserted at similar frequencies, possibly due to differences between the injected plasmids (the *gfp* plasmid was linearized whereas the *mMaple* and

A

MosTI site	Cas9	Transgene	sgRNA concentration	Injected animals	F2 array transgenics	Single-copy insertions	Efficiency (inserts/transgenic)
Chr. I	P*smu-2*	P*eft-3::gfp*	15 ng/ul	10	8	2	25%
Chr. II	P*smu-2*	P*eft-3::gfp*	15 ng/ul	10	8	6	75%
Chr. II	P*smu-2*	P*eft-3::gfp*	15 ng/ul	15	11	4	36%
Chr. II	P*hsp-16.21*	P*rpl-7A::gfp*	10 ng/ul	15	10	5	50%
Chr. II	P*hsp-16.21*	P*rpl-7A::gfp*	10 ng/ul	15	10	5	50%
Chr. IV	P*smu-2*	P*eft-3::gfp*	15 ng/ul	15	10	5	50%

B



C

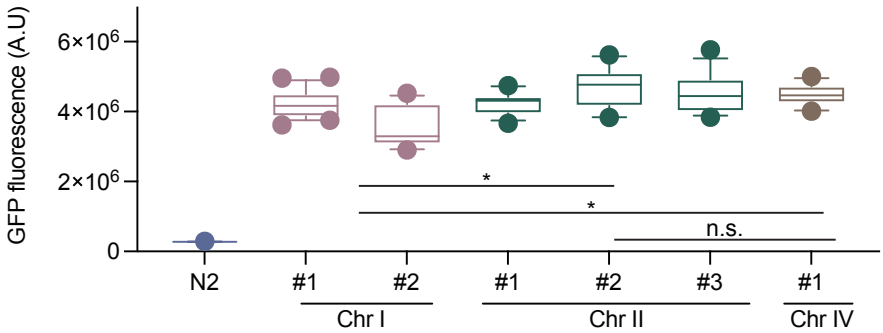


Figure 2 | MosTI insertion frequency and transgene expression
A. Table with MosTI insertion efficiencies using constitutive (P*smu-2::Cas9*) or heat-shock inducible (P*hsp-16.41::Cas9*) plasmids for Cas9 expression. **B.** Single-copy transgenes (P*eft-3::gfp*) are widely expressed, including the germline (13/13 inserts expressed). Left panels 10X magnification, right panels 40X magnification. Scale bars: 20 microns. **C.** GFP quantification by wide-field fluorescence microscopy (L4 stage) from single-copy P*eft-3::gfp* insertions at three MosTI landing sites. Statistics: Kruskal-Wallis's test of fluorescence intensity grouped by insertion site (Chr. I: 35 images, n = 2 ; Chr. II: 41 images, n = 3; Chr. IV: 13 images, n = 1 insert, * P < 0.05).

mScarlet plasmids were supercoiled). We recently showed that linearized plasmids were incorporated into arrays more efficiently than supercoiled plasmids (Priyadarshini *et al.* 2022); the integration frequency may simply reflect the relative abundance of plasmids in the array. We note that a conceptually similar approach was developed by Kaymak *et al*. (2016) using MosSCI to assay 3' UTR gene regulation but their approach was not scalable in the same way as MosTI.

In sum, MosTI is an efficient method to insert single-copy transgenes into well-defined genomic locations using several protocols. All landing sites are compatible with robust somatic and germline expression. From a practical perspective, the split selection markers make it easier to isolate animals with single-copy transgenes and the heat-shock protocol is a simple method to generate many inserts from only a few injections.

Exchanging MosTI selection markers is simple

The use of a strong phenotypic selection marker, such as *unc-119*, has the benefit that insertions are easy to identify, and insertions are easily rendered homozygous. However, *unc-119* animals are slow growing and harder to inject than wild type animals (Maduro and Pilgrim 1995). We designed MosTI to be modular and have developed a standardized strategy to convert selection markers at MosTI landing sites. Conversion is simple: we use MosTI to insert a transgene containing a second split selection marker and a new spacer sequence (**Figure 4A**). In a second (optional) step, the *cbr-unc-119* marker is removed by Cre-mediated recombination. We validated this conversion strategy using a split hygromycin resistance gene, similar to Stevenson *et al*. (2020), and a split fluorescent marker that gives bright muscle fluorescence (P*mlc-1*::*gfp*) (El Mouridi *et al.* 2020). We converted MosTI sites to HygroR (Chr. II) and *gfp* (Chr. I and

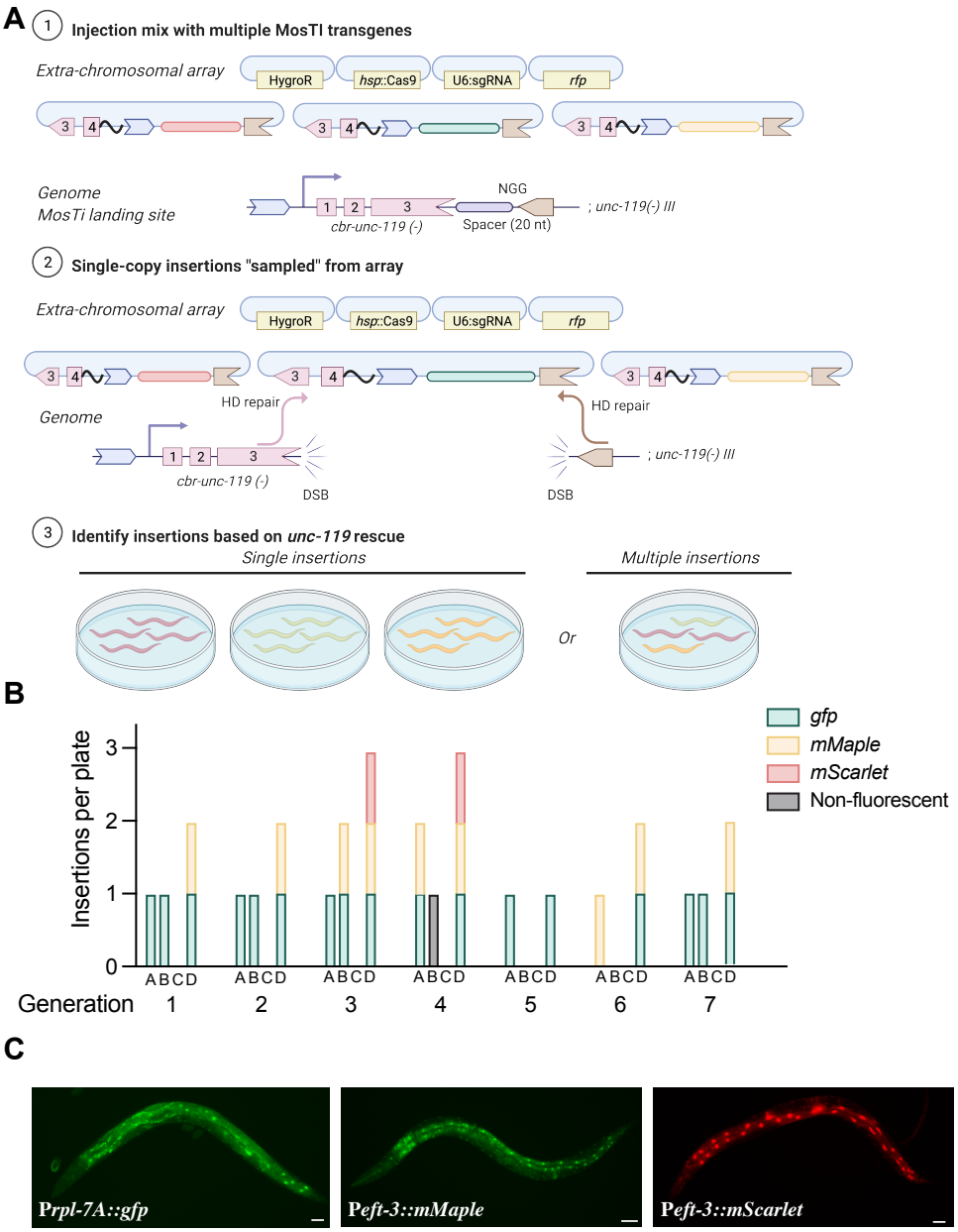


Figure 3 | Single-copy insertions from multiplex transgene injections
A. Schematic of multiplex insertions using MosTI. Multiple transgenes are pooled in one injection mix. During homology-directed repair, a single transgene from the array is inserted into the MosTI site. **B.** Quantification of single-copy MosTI insertions over seven generations from four independent transgenic animals with extra-chromosomal arrays containing three target transgenes (P*rpl-7A*::*gfp*, P*eft-3*::*mMaple3*, P*eft-3*::*mScarlet*) and a heat-shock inducible Cas9 (P*hsp-16.21*::*Cas9*). In each generation, a new population of animals was generated from a non-rescued (Unc) animal carrying an extra-chromosomal array and these animals were heat-shocked one or two days before exhausting the bacterial lawn. **C.** Fluorescence microscopy showing expression from single-copy insertions of each of the three target transgenes. 20X magnification, scale bar = 20 microns.

Chr. II) and generated single-copy P*mex-5*::*gfp* insertions (**Figure 4B**). Both selections were efficient and resulted in GFP expression in the germline from the *mex-5* promoter. However, we found that identifying fluorescent animals (four inserts from 15 injected animals) (**Figure 4C**) was considerably easier than isolating animals with antibiotic resistance (one insert from 20 injected animals). MosTI insertions most frequently happen in the F2 generation when food is almost exhausted: it was relatively easy to identify moving (*unc-119* rescued) or fluorescent (*gfp* selection) L1 animals on a starved plate. In contrast, it was harder to identify healthy animals on hygromycin, which is most effective at preventing larval growth and takes several days to kill animals (Radman *et al.* 2013).

Regardless of the preferred selection, we show that it is relatively easy to change a MosTI site to fluorescence or antibiotic selection. The conversion vector, targeting vectors, and sgRNA can be cheaply generated by gene synthesis with minimal cloning. By modifying existing designs and following a few simple design guidelines, additional safe-harbor landing sites are compatible with the standard collection of targeting vectors. We also note that *cbr-unc-119* selection can be "re-used" by inserting a second split *unc-119* fragment. This could, for example, be used to iteratively generate strains with multiple, tightly linked single-copy transgene insertions.

Split selection markers allow targeted integration of extra-chromosomal arrays

In some cases, simultaneous expression of many transgenes is needed. For example, a method for automated neuron identification relies on expression from 41 different fluorescent promoter constructs (Yemini *et al.* 2021). In other cases, high expression of a single transgene may be necessary, for example, for genetically encoded sensors (e.g., Kerr *et al.* 2000; Flytzanis *et al.* 2014; Hashemi *et al.* 2019) or for spatiotemporal control of gene expression with bi- or tri-partite effectors (Wei *et al.* 2012; Wang *et al.* 2018). Transgene arrays, which contain several hundred plasmids (Stinchcomb *et al.* 1985), are useful for this purpose, especially when arrays are chromosomally integrated to stabilize expression and avoid mitotic loss in cell division (Mello and Fire 1995). We have adapted the split MosTI selection for a CRISRP/Cas9-based method that uses non-homologous end-joining (NHEJ) to target arrays for integration (Yoshina *et al.* 2016) (**Figure 5A**). The adaptation required a short non-rescuing *unc-119* fragment ("array target fragment") to mediate array integration at MosTI landing sites by NHEJ. In this scheme, both the array target

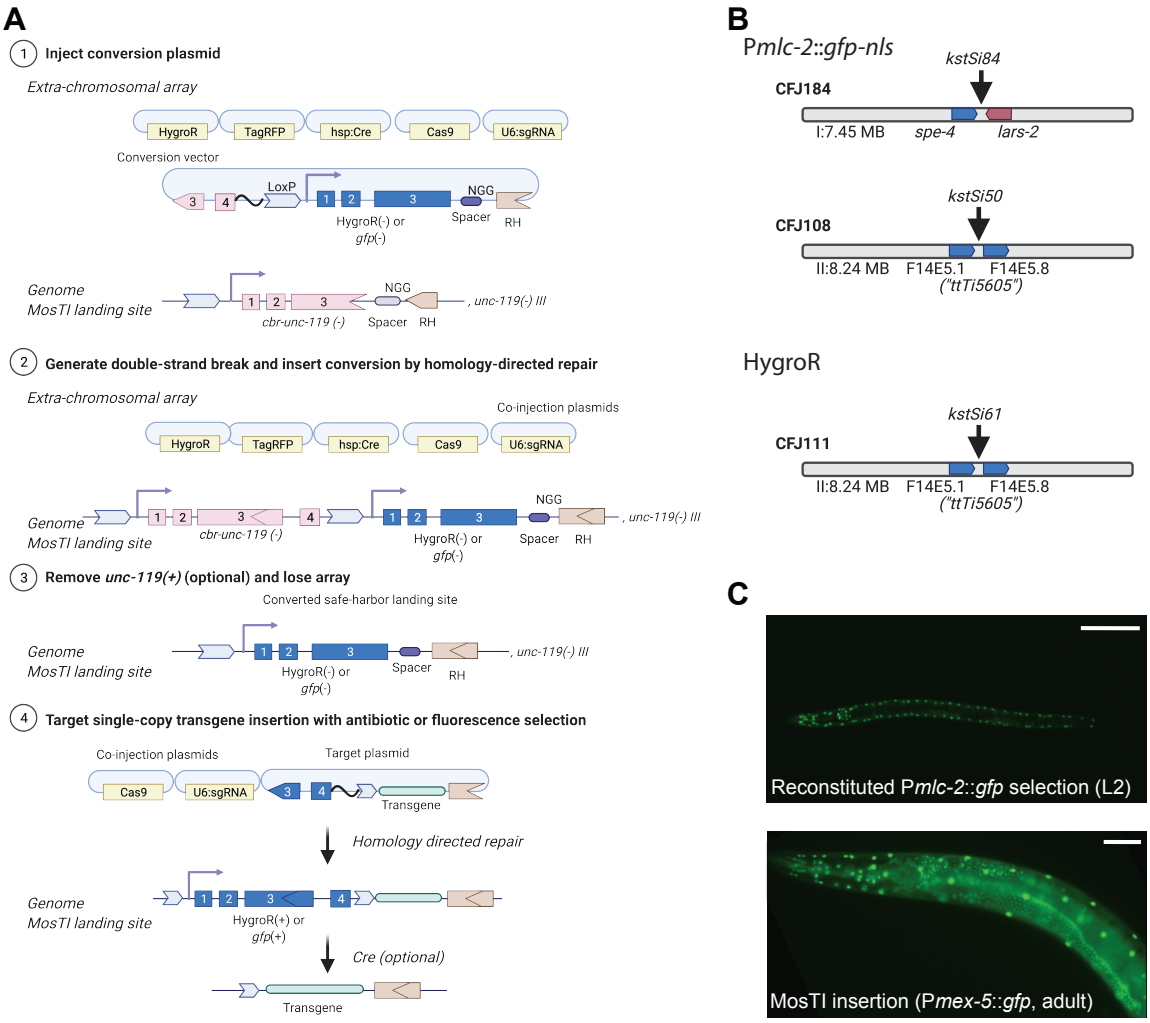
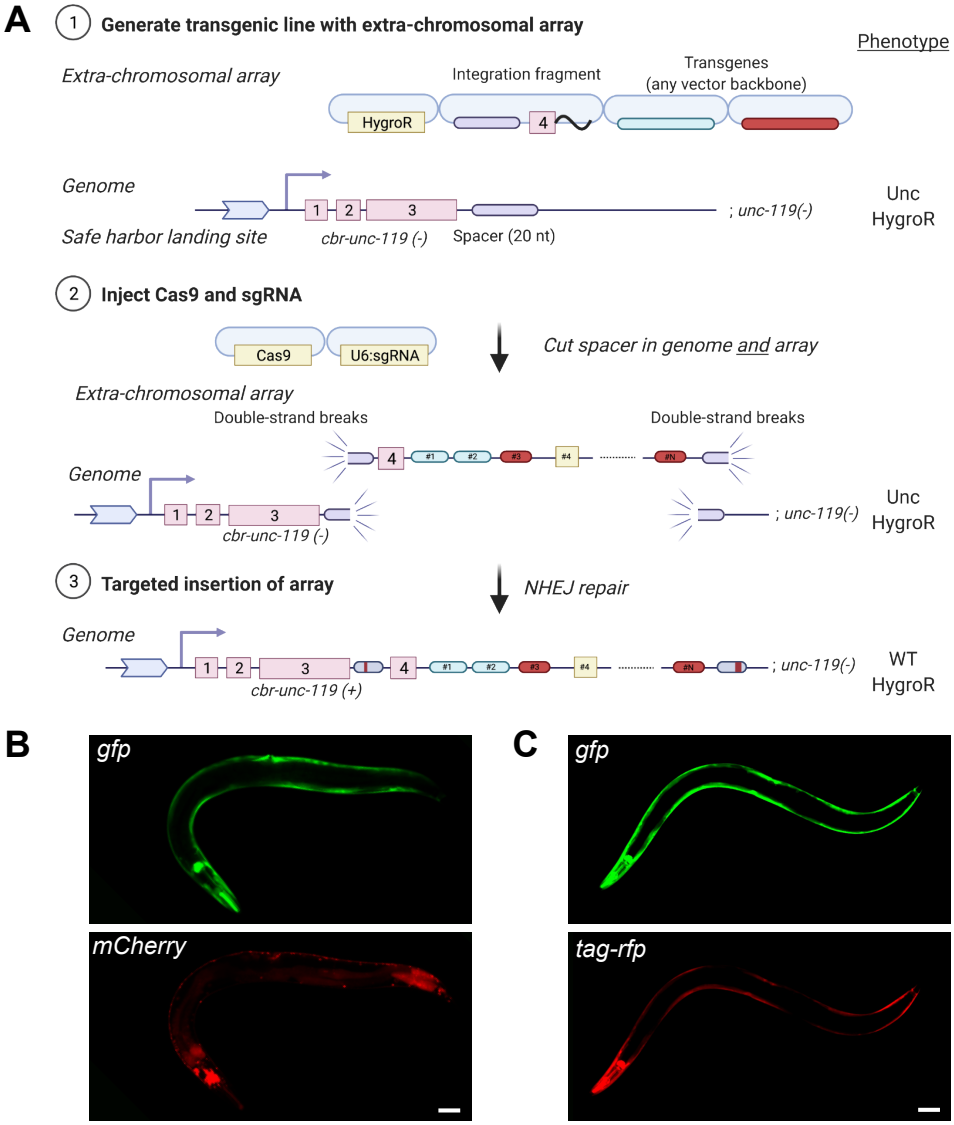


Figure 4 | MosTI sites are modular and can be converted to other split selection markers
A. Schematic of the conversion of an *unc-119* MosTI site into a MosTI site that uses a different split selection marker (here, HygroR or P*mlc-2*::*gfp*). 1. Injection of a conversion plasmid (*e.g.*, generated by gene synthesis) into *unc-119*(*ed3*) mutants with a MosTI landing site. 2. Insertion of the conversion plasmid generates a new MosTI landing site using a different split selection marker. 3. Loss of the extra-chromosomal array and (optional) excision of the reconstituted *cbr-unc-119* selection marker. 4. Injection of a modified MosTI target vector uses the novel selection marker to insert a single-copy transgene. **B.** Available MosTI sites for split P*mlc-2*::*gfp-nls* and HygroR selection. **C.** Example of a single-copy insertion of a P*mex-5*::*gfp* (germline) transgene into a converted MosTI site using a split pan-muscular *gfp* selection marker (P*mlc-2*::*gfp-nls*). Left panel: Pan-muscular nuclear GFP expression (L2 stage). Right panel: Nuclear GFP expression in muscles (from pan-muscular selection) and in the germline (from the P*mex-5*::*gfp-nls*) transgene. Left panel 20X magnification, right panel 40X magnification. Scale bar: 20 microns.

fragment and the split *cbr-unc-119* selection marker at the MosTI landing site are cut by the same sgRNA. The DSBs are generated in an intron, so NHEJ between the array and insertion site can reconstitute the full *unc-119* selection marker even if the repair generates short indels (**Figure 5A**). We first tested array integration efficiency at a MosTI site (Chr. II) using a two-step protocol ("two-step plasmid") which avoided co-integrating the Cas9 and sgRNA plasmids. We first generated transgenic MosTI animals with an extra-chromosomal array containing two fluorescent markers (P*mlc-1::gfp,* P*mlc-1::tagRFP*), a plasmid encoding antibiotic resistance (HygroR), and the array target fragment. Like the protocol for single-copy insertions, these transgenic array animals expressed the injected fluorophores and were resistant to hygromycin but were not rescued for the *unc-119* phenotype. We injected these transgenic animals with a second mix containing plasmids that expressed Cas9 and an sgRNA to generate DSBs. We successfully identified integrated arrays based on *unc-119* rescue at frequencies ranging from 20 to 33% of injected animals (**Figure 5D**). As expected, the rescued animals with the integration expressed all fluorophores from the initial extra-chromosomal array (**Figure 5B**) and could be rendered



D

Integration site	Strategy	sgRNA	Cas9	No. integrated fluorophores	sgRNA concentration	Injected animals	Transgenics	Array integrations	Efficiency (injected/integrations)
MosTI Chr. II	One-step plasmid	Plasmid sgRNA 1	P*smu-2*	2	2.5 ng/ul	15	6	4	27%
	One-step plasmid	Plasmid sgRNA 1	P*smu-2*	2	0.5 ng/ul	15	3	3	20%
	Two-step plasmid	Plasmid sgRNA 1	P*smu-2*	3	15 ng/ul	20	nd	4	20%
	Two-step plasmid	Plasmid sgRNA 1	P*hsp-16.21*	2	15 ng/ul	15	nd	3	20%
	Two-step plasmid	Plasmid sgRNA 1	P*hsp-16.21*	2	15 ng/ul	15	nd	5	33%
unc-119(*ed3*)	One-step plasmid	Plasmid sgRNA 1	P*smu-2*	2	15 ng/ul	10	6	4	40%
	Two-step plasmid	Plasmid sgRNA 1	P*smu-2*	2	15 ng/ul	15	nd	2	13%
	Two-step plasmid	Plasmid sgRNA 2	P*smu-2*	2	15 ng/ul	15	1	1	7%
	Two-step protein	crRNA 1 + tracrRNA	Protein	2		18	nd	8	44%
	Two-step protein	crRNA 2 + tracrRNA	Protein	2		17	nd	5	29%

Figure 5 | Targeted integration of extrachromosomal arrays at MosTI sites and at the endogenous *ce-unc-119* locus
A. Schematic of "two-step plasmid" integration of extrachromosomal arrays at a MosTI landing site. 1. An extrachromosomal array is formed by injecting transgenes, including a plasmid containing the fourth intron of *cbr-unc-119* ("integration fragment") into *unc-119(ed3)* animals containing a MosTI landing site. 2. Plasmids expressing Cas9 and an sgRNA injected into transgenic animals cause double-strand breaks (DSBs) at the MosTI landing site and in the array (in integration fragments). 3. DSBs are repaired by non-homologous end-joining (NHEJ) between the array and the MosTI site, resulting in *unc-119* rescue. The sgRNA target is in an intron to allow rescue even when NHEJ repair causes short indels. Note, arrays can also be inserted by a similar strategy at the *ce-unc-119(ed3)* locus, or by injecting Cas9 protein and crRNA/tracrRNA. **B.** Fluorescence microscopy of an integrated extra-chromosomal array containing transgenes expressed in muscles (P*mlc-1*::*gfp* and P*rab-3::mCherry*) inserted into a MosTI site on Chr. II . Scale bar = 20 microns. **C.** Fluorescence microscopy of an integrated array (P*mlc-1*::*gfp* and P*mlc-1::tagRFP*) inserted into the endogenous *ce-unc-119(ed3)* locus (Chr. III). Scale bar = 20 microns. **D.** Table showing the efficiency of extra-chromosomal array insertion using one-step and two-step protocols at MosTI landing sites and at *ce-unc-119*(*ed3*).

homozygous. Most strains with array insertions were fully rescued for the Unc-119 phenotype except for one strain with particularly bright expression that showed variable rescue. This integration was likely an example of somatic position effect variegation since both rescued and non-rescued animals segregated a mix of rescued and non-rescued animals over several generations (**Supplementary Figure S4**).

Although array insertion into a MosTI site may be preferable, the same integration method should work at any genetic locus with a strong loss-of-function phenotype caused by a late mutation. Many laboratories have the commonly used *unc-119(ed3)* strain and we, therefore, tested insertions into the *unc-119* locus using a modified integration fragment and sgRNA. Using the same two-step plasmid protocol, we were able to integrate arrays into the *unc-119* locus at similar frequencies to MosTI sites (7 to 40% efficiency) (**Figure 5C-D**). We tested three additional protocols for generating array insertions. In one protocol ("one-step plasmid"), we injected all plasmids in a single mix (fluorophores, Cas9, sgRNA, and integration fragment). Rescued integrants from this technique were obvious already in the F1 generation and were generally dimmer than integrants from the "two-step plasmid" protocol (see next section). These integrants may correspond to smaller arrays that would not have become stable inherited arrays (Mello *et al.* 1991; Lin *et al.* 2021). In a third protocol, we first generated an intermediary transgenic injection strain with plasmids encoding a heat-shock inducible Cas9 (*hsp-16.2*::*Cas9*) and hygromycin selection. We used this strain for a second injection with array plasmids (fluorophores, integration fragment, and sgRNA) and generated animals with two arrays. We heat-shocked these animals and were able to generate array insertions at reasonably high frequencies (20 to 33%) (**Figure 5D**). Finally, we also tested the initial two-step plasmid protocol using Cas9 protein and pre-complexed crRNA/tracrRNA guides ("two-step protein") to integrate arrays in *unc-119(ed3)* animals. The protein-based protocol resulted in similar or higher efficiencies depending on the cut-site and crRNA used (29 to 44%)(**Figure 5D**).

In sum, MosTI can be used to target extra-chromosomal arrays for integration using *unc-119* selection and a variety of protocols with similar efficiencies. We have not attempted integrations at MosTI sites using other selection markers, but the approach should work for hygromycin or fluorescence selection. The choice of protocol will depend on the desired size of integrations, how many integrations are needed, and whether continued expression of Cas9 may cause a problem for subsequent experiments. We note that we have not observed arrays changing over time, despite integration of Cas9 and sgRNA plasmid (see next section).

Conceptually similar protocols can generate array integrations at the endogenous *unc-119* locus and can presumably be extended to most other endogenous genes with a strong loss-of-function phenotypes.

Molecular characterization of targeted array insertions

An early study demonstrated that arrays formed from super-coiled plasmids contain sequences organized in tandem, with some plasmids containing short duplications and deletion (Stinchcomb *et al.* 1985). The size and repetitive nature of simple arrays formed from plasmid DNA preclude full assembly using Illumina short-read sequencing. However, long-read sequencing methods, such as Oxford Nanopore and PacBio sequencing, facilitate assembly of complex and repetitive genomic regions (Logsdon *et al.* 2020). A recent study assembled a complex extra-chromosomal array formed from yeast genomic DNA and linear transgene fragments by Nanopore sequencing (Lin *et al.* 2021). The array was mainly formed by non-homologous end-joining of yeast genomic DNA and spanned 50 contigs with a total estimated length of 11 Mb. In another study, Tyson *et al.* (2018) characterized a low-copy plasmid integration generated by biolistic bombardment using Nanopore sequencing. The biolistic integrant contained 30 kb DNA but none of the transgenes (P*pie-1*::*gfp*) were fully intact and the integration resulted in a 2 Mb duplication of the flanking region.

To better characterize integrated arrays formed from circular plasmid DNA and to detect large-scale genomic changes, we first performed short-read Illumina whole-genome sequencing (WGS). We characterized arrays integrated at the *unc-119* locus by the one-step plasmid protocol ("plasmid-integrated strains") and the two-step protein method ("protein-integrated strains"). All injections contained supercoiled plasmids encoding two fluorophores (P*mlc-1*::*gfp* and P*mlc-1*::*tagRFP*), the integration fragment, and linear stuffer DNA (1kb plus DNA ladder, Invitrogen). Injections with Cas9 plasmid were supercoiled (P*smu-2::cas9::mCherry*) and protein strains contained a supercoiled plasmid encoding an antibiotic resistance marker (HygroR) (**Figure 6A**).

We used WGS data to estimate the copy number of all plasmids in the injection mix based on average read coverage (**Figure 6A**). We detected at least one copy of all injected plasmid (from two to 23 copies) in every strain and the integrated arrays contained between six and 37 plasmids in total (**Figure 6A**). Our plasmid copy-number estimates are likely correct within a factor of two: our analysis estimated the copy number of two endogenous genes to 0.97 ± 0.14 for *tbb-2* (47% GC) and 0.48 ± 0.06 for *smu-2* (30% GC). Plasmid-integrated strains were less bright compared to protein-integrated strains, and this correlated with lower estimated plasmid copy-number (**Figure 6A**). All arrays contained multiple copies of the integration fragment with two strains (one plasmid-integrated and one protein-integrated) containing intact spacer sequences, showing that Cas9 cutting efficiency was sometimes incomplete using either integration protocol (**Supplementary Figure S5**). A couple of observations are worth noting: in the plasmid-based protocol, both sgRNA and a germline-expressed Cas9 are integrated which could possibly delete parts of the array over time. However, we propagated a plasmid-integrated strain with several intact copies of the integration plasmid for ten generations and observed no change in fluorescence (**Supplementary Figure S6**). Also, all four protein strains were derived from a single transgenic extra-chromosomal array strain. Given the variation in plasmid-copy number it is likely that only some fragments of the original array were integrated. In this model, Cas9 protein injection "catastrophically" cuts the array into fragments at most spacer sequences and the genomic location; NHEJ between array fragments and the genomic locus results in partial array integration. If this model is correct, then the relative copy-number of the integration fragment in the injection mix could be used to tune the size of the integrated array.

The *unc-119* selection scheme tolerates indels at the cut-site because the spacer is in an intron. However, large indels or duplications near the cut-site could be problematic by perturbing nearby gene expression. We analyzed the WGS data to detect small and large-scale indels near the insertion site based on read coverage (**Supplementary Figure 7**). None of the integrated strains contained short or large-scale deletions and we found only one instance of a putative short duplication (approximately 6 - 7 kb of *unc-119*) (**Supplementary Figure 7**).

To determine the structure of a simple array (*i.e.*, an array composed mainly of plasmid DNA and repetitive DNA ladder) we performed long-read Oxford Nanopore MinION sequencing on a strain with an integrated array containing a relatively high copy number (CFJ157). We were able to assemble the array into 40 contigs with an N50 value of 489 kb. We estimated the total array size at 5.5 Mb (**Figure 6C**), which is within the range of estimates based on microscopy (1 to 2.2 Mb) (Woglar *et al.* 2020) and long-read assembly (11 Mb) (Lin *et al.* 2021). Plasmid-copy number estimates based on Illumina and nanopore sequencing were in close agreement (**Figure 6D** and **Supplementary Figure 5**), validating our approach for estimating plasmid copy number and suggesting that the long-range sequencing assembly for CFJ157 is near-complete.

We determined structural features from the longest contigs and observed short tandem arrays (commonly 3-4 plasmids) interrupted by repetitive DNA sequences, presumably from the 1 kb ladder used as carrier DNA (Invitrogen) (**Figure 6B**). The sequence of the DNA ladder sequences is proprietary but are derived from phage DNA and *E. coli* DNA, including a ColE1 replication origin and an ampicillin resistance gene with homology to the injected plasmids. We observed structures that demonstrate array assembly by a combination of homologous

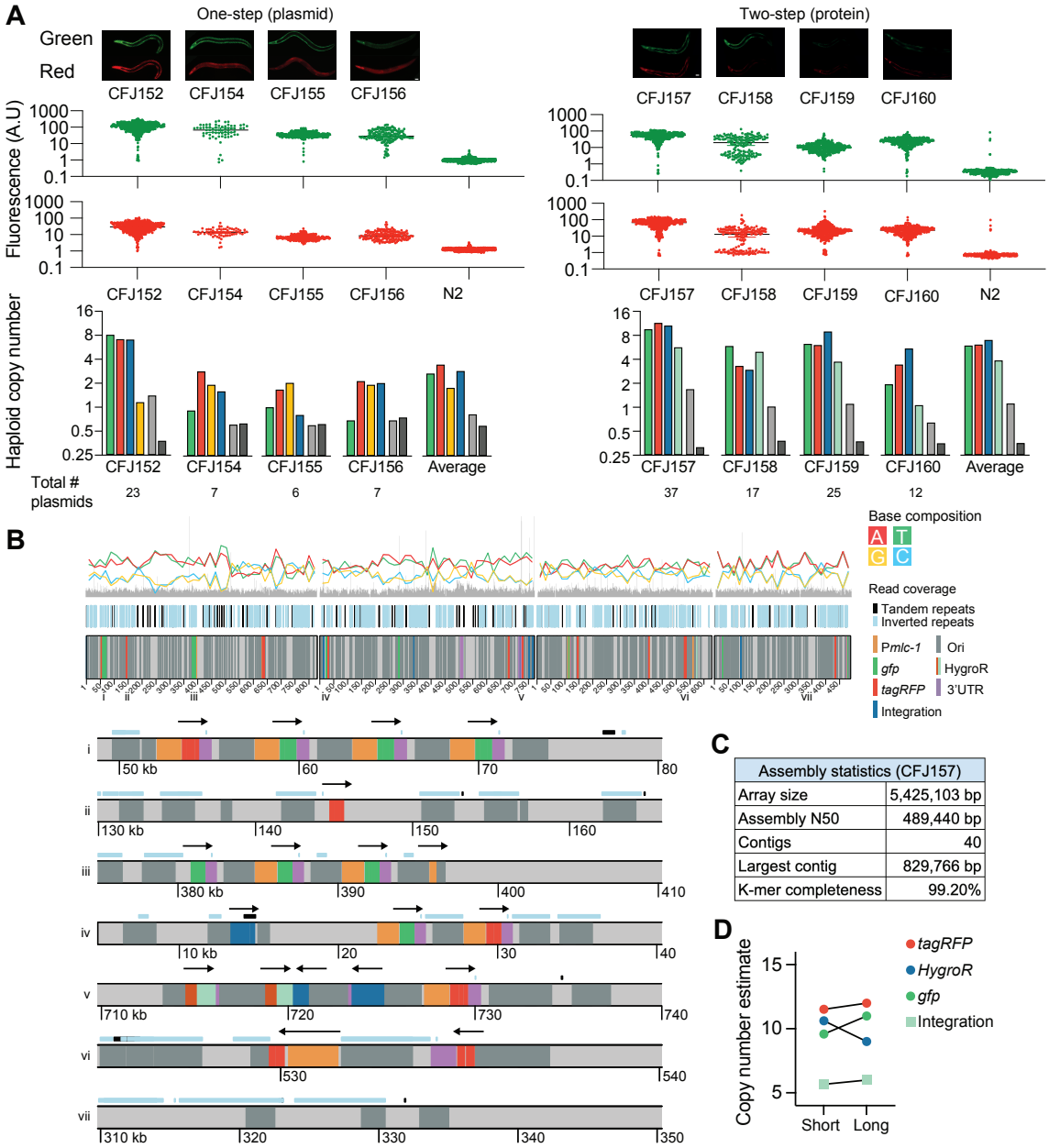


Figure 6 | Molecular characterization of targeted array integrations
A. Analysis of arrays integrated at the *ce-unc-119* locus using a "one-step plasmid" protocol (left) and a "two-step protein" protocol (right). Top: Fluorescence images of strains with array integrations. Scale bar = 20 microns. Middle: Quantification of total fluorescence in adult animals using a COPAS large-particle fluorescence sorter. Bottom: Plasmid copy-number estimates based on whole-genome sequencing (Illumina short read sequencing). **B**. Integrated array assembly from long-read Oxford Nanopore sequencing of CFJ157 (two-step protein array integration containing *gfp*, *tagRFP*, *hygroR*, integration fragment, and 1 kb plus ladder). Top: Schematic overview of four long contigs. Bottom: Individual areas from contigs (location indicated by numbers above). **C**. Assembly statistics from the strain CFJ157. **D**. Comparison of copy-number estimates based on Illumina short-read sequencing and Oxford Nanopore long-read sequencing of CFJ157.

recombination and NHEJ. In many cases, plasmid DNA was joined to ladder DNA sequences by homologous recombination in the amp or ori sequences. In another example, a tandem array of four plasmids (three *gfp* and one *tagRFP*) contained a breakpoint in the P*mlc-1* promoter. The two "parts" of the promoter were joined to the flanking ladder DNA with no obvious homology. We observed a similar structure where the tagRFP in a single plasmid (P*mlc-1*::tagRFP) was split into two fragments, consistent with prior results demonstrating that some plasmids reisolated from extra-chromosomal arrays contain truncations (Stinchcomb *et al.* 1985).

In conclusion, MosTI array insertion results in a diversity of fluorescence expression that is consistent with the plasmid-copy number in the integrated array. Integration does not generally cause wide-spread chromosomal aberrations (indels or duplications) near the insertion site, although some insertions influence the rescue marker, and possibly nearby genes, by position effect variegation. Arrays formed from a mix of supercoiled plasmids and linear carrier DNA contain short tandem plasmid arrays with some plasmids containing breakpoints in regulatory regions (promoter) or in coding regions.

Discussion

We have described protocols and a set of reagents that enable modular safe-harbor transgene insertion (MosTI) in the *C. elegans* genome. The insertion frequencies are high, and the isolation of transgenic strains with targeted insertions is facilitated by mutant rescue, antibiotic resistance, or fluorescence. MosTI is versatile by allowing insertion of both single-copy transgenes and extra-chromosomal arrays. Finally, the reagents are modular with easy conversion between selection markers or generation of novel insertion sites.

MosTI in comparison to commonly used insertion methods

In the past decade, several novel methods for transgene insertion have been described. We were inspired by these methods to develop MosTI that incorporate and improve on several of these advances. Advances have largely been driven by CRISPR/Cas9, with Dickinson *et al.* (2013) first showing that Cas9 induced double-strand DNA breaks could mediate single-copy transgene insertions at similar frequencies to MosSCI. Subsequently, improvements by many laboratories have increased the versatility and efficiency of CRISPR/Cas9 by, for example, using ribonucleoproteins (Paix *et al.* 2015; Au *et al.* 2019), oligos or modified repair templates (Dokshin *et al.* 2018), improved cloning strategies (Schwartz and Jorgensen 2016), and efficient protospacer sequences (Farboud and Meyer 2015). These improvements have, understandably, primarily focused on editing or tagging endogenous loci.

Some efforts have also focused on improving the ability to insert transgenes into safe-harbor locations. Aram *et al*. (2019) improved MosSCI by allowing the removal of selection markers. Silva-Garcia et al. (2019) used CRISPR/Cas9 to convert a set of MosSCI sites for use with *dpy-10* co-CRISPR editing to enrich for insertions (Arribere *et al.* 2014). This method, "SKI LODGE" is optimized for inserting PCR amplified coding regions with well-defined tissue specific promoters and protein tags present at the landing sites. The use of PCR products might be expected to allow multiplexed and high-throughput transgenesis but the low insertion efficiency (0.4 - 2%) of moderate size transgenes (2 kb) is currently a limitation. For this reason, we developed MosTI using plasmid templates but incorporated their use of designer landing sites with efficient protospacers.

Recently, Nonet (2020) demonstrated that recombination mediated cassette exchange (RMCE) using Flp/*FRT* or Cre/*LoxP* could be adapted for use in worms. In this case, landing sites with a GFP were randomly inserted into the genome by transposition (Frøkjær-Jensen *et al.* 2014). To insert a transgene by RMCE, the GFP is exchanged for a self-excising cassette (Dickinson *et al.* 2015) by screening for Rol animals. Although the use of RMCE is promising, particularly for inserting large transgenes, the method has a few limitations: First, insertions are generated at relatively low frequency (one in three injected animals). Second, co-insertion of the large self-excising cassette (7.7 kb) limits the size of transgene that can be inserted. Third, attempts by Nonet (2020) to integrate arrays were unsuccessful, possibly because constitutive expression of Flp recombinase "destroyed" arrays. Here, we have favored generating targeted landing sites by CRISPR/Cas9 and shorter selection markers. If preferred, the flexibility of modular MosTI landing sites can easily be converted to RMCE-compatible landing sites.

Finally, another research group has previously described a very similar use of a split antibiotic selection marker (HygroR) to insert single-copy transgenes (Stevenson *et al.* 2020). In this case, the authors generated a single landing site at the *ttTi5605* site on Chr. II and tested the efficiency of six protospacer sequences. Like our results, the authors note that single-copy insertions are easily identified based on hygromycin rescue although they report considerably lower insertion efficiencies (1 - 10%). Although insertion frequencies are difficult to compare across laboratories, one likely difference is our use of a recently developed germline-optimized Cas9 transgene (Aljohani *et al.* 2020). Stevenson *et al.* (2020) mainly focused on using *in vivo* recombineering of co-injected PCR fragments, which was first used for GFP tagging genes by Paix *et al*. (2016). *In vivo* recombineering has the significant advantage that cloning is minimized but comes at the expense of reduced insertion frequency and fidelity (Stevenson *et al.* 2020). We have preferred the use of clonal or synthetic transgenes but note that all MosTI landing sites and split selection markers are compatible with the *in vivo* recombineering strategies described by Stevenson *et al*. (2020).

Advantages of MosTI for single-copy insertions

MosTI has several advantages for single-copy insertions, in particular compared to MosSCI (Frøkjær-Jensen *et al.* 2008, 2012). First, the use of split selection markers makes it considerably easier to identify true transgene insertions, obviating the need for complex co-injection mixtures and time-consuming screens for insertions. Each selection marker has its pros and cons. For *unc-119*, the selective pressure is very strong, it is easy to generate homozygous strains, and no specialized plates or fluorescence microscopes are required. The disadvantages include the difficulty of maintaining and injecting *unc-119* animals, no selection for the transgene in genetic crosses, and possible adverse effects of partial gene rescue. The advantages of antibiotic markers include easier

strain handling, injection into wildtype animals, and strong selection in genetic crosses. The disadvantages include the cost and inconvenience of making plates with antibiotics and the difficulty of distinguishing homozygous from heterozygous inserts because non-rescued animals are "invisible" on the selective media. To our knowledge, fluorescent markers have not previously been used as the only selection marker for single-copy transgene insertions (in contrast to tagging endogenous genes), likely because fluorescence from arrays and insertions are difficult to distinguish. Split fluorescent selection markers are convenient because phenotypically normal animals are injected, homozygous and heterozygous inserts can be distinguished based on segregation, no special plates are needed, fluorescence protein expression is generally innocuous, and insertions are easily crossed to other strains by "following" the fluorescence. Minor limitations include the requirement for a fluorescence dissection microscope and possible interference between transgene and selection fluorophores. In this case, the fluorescent selection marker can be removed by Cre-mediated recombination.

Also, in contrast to MosSCI, the new MosTI insertion sites were selected based on chromosome location and local permissive chromatin marks. All the insertion sites show a high frequency of germline expression and consistently bright somatic expression. We have not attempted to do an in-depth phenotypic characterization of the strains but have not observed any obvious adverse phenotypes in animals with insertions. We have furthermore performed whole-genome sequencing on several of the insertion strains to detect mutations that were not outcrossed after the initial isolation of *unc-119*(*ed3*) or off-target effects from CRISPR/Cas9. We detected only a few mutations that cause amino-acid substitutions in genes with subtle phenotypes defined in **Supplementary Table 1**.

Using MosTI for targeted array insertion

Targeted array integration by MosTI may be useful when high expression or co-expression of many plasmids is necessary. One striking example is the use of 41 different fluorescence drivers with overlapping expression that allow automated cell-specific identification and monitoring neuronal activity "NeuroPAL" (Yemini *et al.* 2021). To develop the NeuroPAL methodology, Yemini *et al.* (2021) tested many different plasmid combinations and generated integrated lines for many "landmark" fluorophores and a genetically encoded calcium sensor (GCaMP6). Each integrant required considerable effort, including integration, mapping, outcrossing, and phenotyping. Given the continuous development of improved fluorophores and optogenetic tools it is likely that tools such as NeuroPAL exemplify circumstances where rapid and reproducible array insertion would allow "prototyping" new functionalities (*e.g.*, optogenetic inhibition or excitation of defined populations of cells). Similarly, somatic control of expression by Gal4 driver lines (Wang *et al.* 2017), FRT mediated gene activation (Davis *et al.* 2008; Voutev and Hubbard 2008), auxin-mediated degradation (Zhang *et al.* 2015), CRISPR inhibition (Long *et al.* 2015), or CRISPR activation (Gilbert *et al.* 2014) are being actively developed in *C. elegans* and often require sustained high transgene expression.

Our approach for inserting arrays using MosTI differ from the most commonly used methodology. Traditionally, arrays have been inserted by irradiation (Mello and Fire 1995; Evans 2006). Integrating arrays require specialized equipment, careful candidate screening, and is relatively slow. Furthermore, array insertion is random and inherently mutagenic. For these reasons, array integration is not standard practice despite mitotic stability and increased stability of expression in integrants (Evans 2006). In the past few years, two alternative array integration methods were developed. In one method, Noma and Jin (2018) adapted a genetically encoded reactive oxygen generator ("miniSOG") to integrate arrays by light stimulation in just two weeks at high frequency (~10%). Although this method is efficient, specialized equipment is still required, arrays are integrated randomly, and background mutations remain a concern. We have instead adapted a second method developed by Yoshina *et al*. (2016) where arrays are integrated by CRISPR/Cas9. In this method, CRISPR/Cas9 is used to generate double-strand DNA breaks at a specific genomic location (*dpy-3*) and in the array (typically, the AmpR marker). Arrays are integrated by non-homologous end-joining at relatively high frequencies (3 - 10%). The disadvantage is that arrays are inserted into a coding region causing a phenotype (Dpy) with no easy strategy for later rescue. Here, we have developed a modified CRISPR/Cas9 based integration strategy that includes a positive selection marker which is compatible with the same safe-harbor locations used for MosTI insertions or with the *unc-119*(*ed3*) mutant strain. MosTI allows quick isolation (~2 weeks) of integrated arrays at high frequency (7 to 40%) with no requirements for extensive screening or genetic mapping.

We characterized several array integrations using short-read sequencing and one array in detail by long-read Oxford Nanopore sequencing. From eight array integrations, we found no evidence for off-target insertions, large-scale chromosomal re-arrangements, or large indels at the integration sites, suggesting that targeted integrants are generally well-behaved. We observed unexpected effects in only one array integration (out of 39 in total), where the strain showed the hallmarks of position effect variegation of the *unc-119* rescue marker. Although we have not characterized this strain in detail, it is possible that repressive chromatin may partially silence *unc-119* at the junction between the genome and the integrant. To our knowledge, there are few published examples of somatic position effect variegation in *C. elegans*; targeted insertion of a large repetitive transgene structure could perhaps be used to understand heterochromatin spreading in worms that do not encode canonical genome organizers, such as CTCF (Heger *et al.* 2009).

Our long-read sequencing confirmed the tandem array structure and partial plasmid deletions in arrays, first demonstrated by Stinchcomb et al. (1985), and was in agreement with size estimates based on microscopy (Woglar *et al.* 2020) and sequencing a complex array (Lin *et al.* 2021). Although we have not sequenced arrays generated by linear DNA fragments, we and others have demonstrated better transgene expression in the soma (Etchberger and Hobert 2008) and germ line (Aljohani *et al.* 2020) from PCR products. Some of this difference may be driven by increased incorporation of linear fragments in arrays (Priyadarshini *et al.* 2022) but the presence of truncated promoters and fluorophores in the array also suggests that the cell may identify unusual DNA structures that are likely to lead to active silencing (Kelly *et al.*

1997; Leyva-Díaz *et al.* 2017). We envision that the ability to routinely insert arrays into specific locations will be useful for understanding basic characteristics of chromosomes, such as how chromosome size influences large-scale genome domains (Liu *et al.* 2010), as well as enabling the initial steps in engineering synthetic *C. elegans* chromosomes.

Multiplexed transgene insertion

CRISPR/Cas9 allows multiplexed transgene insertion because a single transgenic strain with an extra-chromosomal array can carry different repair templates. Kaymak *et al*. (2016) first demonstrated multiplexed repair ("library MosSCI") to study the effect of 3' UTRs, which largely determine the expression within the germline (Merritt *et al.* 2008). Similar to their observations, a complex mix of repair templates did not impair the MosTI repair process. However, library MosSCI was not inducible and required a relatively large number of injections to generate unique inserts (269 injected animals resulted in 11 unique transgene insertions) (Kaymak *et al.* 2016). Here, we show that coupling improved germline expression from arrays (Aljohani *et al.* 2020), an efficient sgRNA (Moreno-Mateos *et al.* 2015), and a split selection marker has allowed us to generate many insertions from a single injection by driving Cas9 expression using heat-shock. Our results were limited to proof-of-principle experiments using three plasmids that could be distinguished based on fluorescent protein expression. However, we propose that the method can be scaled up significantly: the rescuing part of the MosTI repair template (*cbr-unc-119* and the right homology region) is relatively short (<1 kb). Thus, a ~5 Mb array can in principle carry approximately 2000 different copies of a short 1.5 kb transgene, more than enough room to encode a fluorophore with regulatory elements (*e.g.*, promoters or 3' UTRs). It should be possible to develop genetic screens that combine MosTI with large-scale gene synthesis, for example using oligo pools, and high-throughput sequencing to understand gene regulatory elements at the base-pair level (e.g., de Boer *et al.* 2020).

MosTI curation and continuous development

Finally, standardized resources enable reproducibility across a scientific field, for example, by allowing direct comparison between different transgenes inserted at the same location or sharing of compatible reagents. Reagents for MosTI are freely available from the *Caenorhabditis* Genetics Center (CGC) and Addgene. To facilitate the uptake of transgenic methodology methods, we maintain a website (www.wormbuilder.org) with detailed descriptions of reagents and updated protocols. We designed MosTI to be highly modular and plan to continuously develop the technique, including new insertion sites, strains with multiplexed insertion sites each targeted by a different sgRNA, and new split selection markers. We hope that MosTI will be a resource that is of use to the many different types of experiments done by the *C. elegans* community.

Acknowledgements

We thank Mohammed Aljohani for sharing unpublished reagents and Ramatoulaye Balde for excellent lab support. Research in C.F.-J.'s laboratory is supported by KAUST core funding and KAUST's Office of Sponsored Research (OSR-CRG2020-4388). Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

Contributions

Conceptualization: CFJ
Methodology: CFJ, SEM, FA
Software: FA
Formal analysis: SEM, FA
Investigation: SEM
Writing - original draft: CFJ
Writing - review & editing: CFJ, SEM, FA
Visualization: CFJ, SEM, FA
Supervision: CFJ
Funding acquisition: CFJ

Competing interests

The authors declare no competing interests.

ORCID iD

Sonia El Mouridi: 0000-0002-3674-3151
Faisal Alkhaldi: 0000-0002-9967-4982
Christian Frøkjær-Jensen: 0000-0002-3178-0906

Materials & Methods

Strains

Strains were maintained using standard methods (Brenner 1974) and were grown at 20°C on OP50 or HB101 bacteria. Strains containing MosTI landing sites have been deposited with the *Caenorhabditis* Genetics Center (CGC) (see **Supplementary Table 1**).

Sequencing data

Short-read (Illumina) and long-read (Oxford Nanopore) sequencing data was deposited with NCBI with the Bio-Project number: PRJNA827598

Commercial software

Some figures were created with BioRender.com. Figures were designed using Adobe Illustrator 2022 (v26.0.2). Statistical analysis was performed using GraphPad Prism 9 for macOS (v9.3.1). The manuscript was written in Microsoft Word for Mac (v.16.60) and tables were generated using Microsoft Excel for Mac (v. 16.60). In silico molecular biology was performed using ApE (A plasmid Editor)(Davis and Jorgensen 2022) which is freely available at
https://jorgensen.biology.utah.edu/wayned/ape/

Molecular biology

Plasmids were generated by standard molecular techniques, including three-fragment multisite Gateway reactions (Invitrogen cat. no. 12538200), Gibson assembly (Gibson *et al.* 2009), Golden-Gate cloning (Engler *et al.* 2009), or by gene synthesis (Twist Bioscience, CA, USA). All PCRs were performed using a high-fidelity DNA polymerase (Phusion, New England Biolabs, F530S) and constructs generated by PCR were sequence verified by Sanger sequencing. Most plasmids have been deposited with Addgene (Cambridge, MA, USA) (see **Supplementary Table 1**). Annotated GenBank files of all plasmids are included in **Supplementary File 1**.

MosTI landing sites

Please see **Supplementary Table 1** for the exact insertion sites and sgRNA sequences.

unc-119 selection

We generated target plasmids containing a codon-optimized synthetic *unc-119* rescue marker ("*syn-unc-119(*+)") flanked by 250 bp of homology to each side of the insertion site (pSEM203 Chr. II, pSEM224 Chr. I, pSEM226 Chr. IV). We injected these targeting plasmids (individually) into *unc-119*(*ed3*) animals at 25 ng/ul with 25 ng/ul pCFJ2474 (P*smu-2*::Cas9::*gpd-2*::*mCherry*), 15 ng/ul of sgRNA (specific to each site) and 10 ng/ul of fluorescent co-injection markers pSEM233 (P*mlc-1*::*tagRFP*) or pSEM231 (P*mlc-1*::*gfp*) (El Mouridi *et al.* 2020) for a final DNA concentration of 100 ng/ul. After injection, P0 plates were placed at 25°C to increase Cas9 expression. The screen for Unc rescue was done after starvation. Strains that were homozygous for the landing sites were subsequently injected with a plasmid encoding Cre recombinase (pMDJ39) to remove the synthetic *unc-119* rescue marker (flanked by LoxP sites) and identified by screening for Unc animals. We validated all MosTI insertions by PCR and Sanger sequencing. Additionally, we whole-genome sequenced two MosTI insertion strains, CFJ77 (chr. I) and CFJ42 (chr. II) to identify background mutations (see **Supplementary Table S1**). All other insertions strains (for example, strains with split hygroR and P*mlc-2*::*gfp* selection) were derived from the same genetic background (originally an *unc-119*(*ed3*) mutant strain) and are expected to share the same background SNPs.

Converting MosTI landing site to a different selection marker

We converted MosTI strains with split *unc-119* selection to split P*mlc-2*::*gfp* and hygromycin selections. MosTI strains for a given chromosomal location with *unc-119* selection were injected with a conversion plasmid (pSEM258 for hygroR and pSEM260 for P*mlc-2*::*gfp*) and standard plasmids used for MosTI single-copy insertions (Cas9, *hygroR*, sgRNA, *tagRFP* or *gfp* fluorescent co-injection markers). After injection, P0 animals were placed at 25°C. After 48 hours, 500 ul of a 4 mg/ml hygromycin solution was added to the NGM plates to select for transgenic animals. Single-copy insertion of the conversion transgenes was identified by *unc-119* rescue after the food was exhausted. All converted MosTI sites were validated by Sanger sequencing. The conversion strategy should, in principle, work for any selection marker that can be split into two non-rescuing fragments.

Single-copy transgene insertion using MosTI

Constitutive Cas9 protocol: In this protocol, a codon-optimized Cas9 containing PATCs in introns to prevent silencing and a fluorophore to monitor expression (pCFJ2474, P*smu-2*::*cas9*::*gpd-2 tagRFP-T*) (Aljohani *et al.* 2020) was used for constitutive expression from extra-chromosomal arrays. The injection mix consisted of 25 ng/ul linearized Cas9 (pCFJ2474), 25 ng/ul target vector (various), 15 ng/ul pCFJ782 (HygroR), 15 ng/ul sgRNA (pSEM318), and 10 ng/ul of fluorescent co-injection markers (P*mlc-1*::*tagRFP*, pSEM233 or P*mlc-1*::*gfp*, pSEM231) for a total DNA concentration of 100 ng/ul. The plasmids were injected into young adult Unc animals grown at 15°C or 20°C from strains with MosTI landing sites. Single, injected animals were placed on NGM plates seeded with OP50 and incubated at 25°C. 48 hours after injection, 500 ul of a 4 mg/ml hygromycin solution was added to P0 plates to select for transgenic animals. Single-copy insertions were identified by screening plates for any moving animals (*unc-119* rescue) once the food was exhausted (approximately seven to ten after injection).

Heat-shock protocol: In this protocol, codon-optimized Cas9 was expressed under a heat-shock promoter for inducible Cas9 expression. The injection mix contained 25 ng/ul of P*hsp-16.41*::Cas9 (pMDJ231) linearized with ApaLI, 25 ng/ul target vector, 15 ng/ul pCFJ782 (HygroR), 15 ng/ul sgRNA (pSEM318), and 10 ng/ul of muscle co-injection markers (P*mlc-1*::*tagRFP*, pSEM233 or P*mlc-1*::*gfp*, pSEM231 for a total DNA concentration of 100 ng/ul. The plasmids were injected into young adult Unc animals grown at 15°C or 20°C from strains with MosTI landing sites. Single, injected animals were placed on NGM plates seeded with OP50 and incubated at 25°C. 48 hours after injection, 500 ul of a 4 mg/ml hygromycin solution was added to P0 plates to select for transgenic animals. Two different heat-shock protocols were used to generate single-copy insertions: (a) 37°C for 1 hour or (b) 30°C for 18 hours (Nonet 2020). Animals with single-copy insertions were identified based on *unc-119* rescue after the food was exhausted. Note, we frequently observed transgene insertions prior to heat-shock, suggesting that the heat-shock promoter is leaky or that transgenic animals are under stress (*e.g.*, due to temperature, antibiotic selection, or starvation). The longer heat-shock protocol (30°C for 18 hours) was slightly more efficient and was more convenient as all plates placed in an incubator will be equilibrated to the surrounding temperature without the need for wrapping plates in parafilm and using a water bath (Boulin & Bessereau, 2007) or splitting plates into a single layer (Frøkjær-Jensen et al., 2014).

Heat-shock induced MosTI multiplex insertions over generations

A MosTI strain containing a landing site on Chr. II (CFJ42) was injected with a mix containing 10 ng/ul linearized transgene(s), 10 ng/ul pCFJ782 (HygroR), 10 ng/ul pSEM318 (sgRNA), 10 ng/ul pMDJ231 (P*hsp-16.41*::Cas9) linearized with ApaLI, and DNA ladder (1 kb ladder, Thermo Fisher Scientific, cat. no. SM1331) added to a final concentration of 100 ng/ul. Injected animals were singled to NGM plates seeded with OP50 and placed at 25°C. One day after injection, 500 ul of 4 mg/ml hygromycin was added to injection plates to select for transgenic strains. The experiment was performed in parallel with seven independent transgenic strains. For each transgenic strain, a single non-rescued transgenic Unc animal was picked to an NGM plate seeded with OP50 seven days after injection and placed at 25°C. For every generation, a single non-rescued transgenic animal was transferred to a new NGM plate with OP50 before the food was exhausted. Subsequently, the transgenic animals were heat-shocked to induce Cas9 expression seven days after being singled to NGM plates. The plates were screened for single-copy transgene insertions (based on moving *unc-119*-rescued animals) after the food was exhausted. In some experiments, the injection mix contained several target transgenes (P*rpl-7A*::*gfp*, P*eft-3*::*mMaple*, and P*eft-3*::*mScarlet*). We determined which transgene had been inserted based on fluorescence, PCR, and Sanger sequencing.

Extrachromosomal array insertions

Extra-chromosomal arrays can be inserted into MosTI sites or into the endogenous *unc-119* locus (in *ed3* mutants) using *unc-119* selection. We have also generated

targeting fragments to insert arrays into MosTI sites that use hygromycin and P*mlc-2*::*gfp* but have not tested these fragments (although we deposited the reagents with Addgene). The protocols are identical except for using targeting fragments and sgRNAs that are specific to each insertion site (listed in **Supplementary Table 1**). Here we describe validated array insertion protocols based on *unc-119* rescue.

One-step direct plasmid-based array integration

With this strategy, all injected plasmids are "directly" integrated at the target site, including plasmids expressing Cas9 and the sgRNA. We injected adult *unc-119* hermaphrodites with a mix composed of 10 ng/ul of pCFJ2474 (P*smu-2*::*cas9*::*gpd-2 tagRFP-T*), 5 ng/ul of array target fragment (*e.g.*, pSEM371 for integration at the endogenous *ce-unc-119* locus), 15 ng/ul of sgRNA targeting the integration site and the array target fragment (*e.g.*, pSEM376 for *ce-unc-119*), 10 ng/ul of transgenes for integration, and DNA ladder (1 kb plus, Invitrogen) to a final concentration of 100 ng/ul. Injected animals were placed on NGM plates and grown at 25°C degrees. Seven to ten days after injection, we screened plates for array integration based on *unc-119* rescue and expression of injected transgenes (typically, fluorescent markers). We treated any plate with rescued animals as a single independent integration event and picked only a single clonal strain derived from any injected animal.

One-step plasmid-based array integration in pre-injected strain

With this strategy, only the transgenes, the array target fragment, and the sgRNA plasmid were integrated. First, a MosTI strain containing a landing site on Chr. II (CFJ42) was injected with a mix containing: 25 ng/ul pMDJ231 (P*hsp-16.41*::Cas9) linearized with ApaLI, 15 ng/ul pCFJ594 (NeoR), 10 ng/ul pSEM235 (P*mlc-1*::*mCherry*), and DNA ladder (1 kb ladder, Thermo Fisher Scientific) added to a final concentration of 100 ng/ul. Injected animals were singled to NGM plates seeded with OP50 and placed at 25°C. One day after injection, 500 ul of 25 mg/ml neomycin was added to injection plates to select for transgenic strains (not phenotypically rescued for *unc-119*(*ed3*)). A single strain was selected as an injection strain for a second step, where adult animals were injected with a mix containing: 15 ng/ul pSEM376 (sgRNA), 15 ng/ul pCFJ782 (hygroR), 15 ng/ul pSEM234 (P*mlc-1::mCherry::NSL*), 15 ng/ul pSEM233 (P*mlc-1::gfp*), 5 ng/ul pSEM371 (integration fragment), and 1 kb DNA ladder (Invitrogen) to a total concentration of 100 ng/ul. Single, injected animals were placed on NGM plates seeded with OP50 and incubated at 25°C. 48 hours after injection, 500 ul of a 4 mg/ml hygromycin solution was added to P0 plates to select for transgenic animals. Seven to ten days after injection (before the bacterial lawn was exhausted), strains with the two extra-chromosomal arrays were heat-shocked (30°C for 18 hours) and screened for *unc-119* rescue one generation later. We subsequently allowed the initial array (containing NeoR and P*mlc-1*::*mCherry*) to be lost on standard NGM plates with no antibiotic selection and the integrants were rendered homozygous. Cas9 should only cut the array formed in the second injection and as expected, we only observed antibiotic resistance and markers from the second injection mix in the integrated strains.

Two-step plasmid-based array integration

With this strategy, only the transgenes and the array target fragment are integrated; plasmids encoding Cas9 and the sgRNA are injected in a second round and are not part of the integrated array. As a first step, we generated transgenic lines by injecting a mix composed of 2.5 ng/ul of the array target fragment (*e.g.*, pSEM371 for *ce-unc-119* integration), 15 ng/ul of pCFJ782 (HygroR), 10 ng/ul of transgenes, and DNA ladder (1 kb plus, Invitrogen) to a final concentration of 100 ng/ul. Injected animals were singled to NGM plates seeded with OP50 and placed at 25°C. One day after injection, 500 ul of 4 mg/ml hygromycin was added to injection plates to select for stable transgenic strains. In the F2 or F3 generation we picked a single clonal animal to a seeded NGM plate with hygromycin and expanded the population (note that these animals are not rescued for the *unc-119* phenotype but are hygromycin resistant). For the second step, we injected Unc transgenic animals carrying extra-chromosomal arrays with a mix containing 20 ng/ul pCFJ2474 (P*smu-2*::*cas9*::*gpd-2 tagRFP-T*), 15 ng/ul of plasmid expressing an sgRNA to cut the integration site and the array fragment (*e.g.*, pSEM376 for *ce-unc-119* integration), 10 ng/ul of fluorescent co-injection marker (pSEM233 or pSEM231), and DNA ladder (1 kb plus, Invitrogen) to a final concentration of 100 ng/ul. After injection, we placed single individual injected animals on seeded NGM plates with hygromycin and grew the animals at 25°C. We identified strains with array integrations seven to ten days after injection based on unc-119 rescue. We treated any plate with rescued animals as a single independent integration event and picked only a single clonal strain derived from any injected animal.

Two-step protein-based array integration

With this strategy, only the transgenes and the array target fragment are integrated; Cas9 protein (IDT) and crRNA/tracrRNA (IDT) are injected in a second round. Integrations are generated faster with this protocol compared to plasmid-based protocols as array integrants are identified in the progeny of the injected animals. As for the two-step plasmid protocol, we first generated transgenic lines by injecting a mix composed of 2.5 ng/ul of the array target fragment (*e.g.*, pSEM371 for *ce-unc-119* integration), 15 ng/ul of pCFJ782 (HygroR), 10 ng/ul of transgenes, and DNA ladder (1 kb plus, Invitrogen) to a final concentration of 100 ng/ul. Injected animals were singled to NGM plates seeded with OP50 and placed at 25°C. One day after injection, 500 ul of 4 mg/ml hygromycin was added to injection plates to select for stable transgenic strains. In the F2 or F3 generation we picked a single clonal animal to a seeded NGM plate with hygromycin and expanded the population (note that these animals are not rescued for the *unc-119* phenotype but are hygromycin resistant). For the second step, we prepared guide RNA duplexes *in vitro* by incubating 3 ul of 100 uM crRNA (*e.g.*, targeting *ce-unc-119*), 3 ul 100 uM tracrRNA, and 4 ul nuclease-free duplex buffer (IDT) at 95°C for 5 min, followed by a 5 min incubation at room temperature. We mixed 3 ul of the crRNA::tracrRNA duplex with 10 ug recombinant Cas9 protein (Alt-R S.p. Cas9 Nuclease V3, IDT), 10 ng/ul fluorescent co-injection marker (pSEM233 or pSEM231), and nuclease-free water to a final volume of 10 ul. We injected transgenic Unc animals carrying extra-chromosomal arrays with this Cas9 protein mix and placed individual injected animals on NGM plates seeded with OP50 at 25°C. Rescued animals were identified in the progeny of the injected animals (the F1 generation)

and isolated before the food was exhausted. We treated any plate with rescued animals as a single independent integration event and picked only a single clonal strain derived from any injected animal.

Characterization of arrays integrated by MosTI

Quantification of fluorescence expression using COPAS large-particle flow cytometer

We obtained similar mixed populations by placing four L4 worms on small NGM plates seeded with OP50 and used four plates per strains. The animals were grown at 25°C for seven days and then washed off with a M9 buffer and placed on ice. After 10 min, the pellet was washed with water to remove bacteria. We quantified the fluorescence of each strain using a COPAS flow cytometer (Union Biometrica). Between each strain, we washed all the tubing by running water for at least one minute. Settings for one-step insertion: Gain = 2, green: 400 volts; red: 450 volts. Settings for two-step insertion: Gain = 2, green: 300 volts; red: 400 volts. Both green and red fluorescence were expressed along the entire length of the worm, so we based our quantification on the integral of the fluorescence signal. We selected animals with a TOF between 600 and 900 (mainly young adults) and with an extinction coefficient below 35000 to select for animals that came straight through the flow cell. We manipulated the data in Microsoft Excel (version 16.59) and graphed the data with GraphPad Prism (version 9.3).

Background SNP analysis in injection strains

Raw sequencing reads were trimmed to remove adapters and filtered for low quality bases (Fastqc) (Andrews 2010). The filtered reads were aligned to the *C. elegans* reference genome (WBcel235) with BWA V0.7.17 (Li 2013). Single-nucleotide variants and indels were called with GATK4 Haplotypecaller (Poplin *et al.* 2018) using filters: QD < 2.0, FS > 60, MQ < 40, SOR > 4, MQRankSum < -8, ReadPosRankSum < -8. Next, we used snpEff to filter out SNPs and indels outside protein-coding genes and those with “modifier” impact (Cingolani *et al.* 2012).

Plasmid copy-number estimation

Genomic DNA from eight strains with integrated arrays was isolated using a DNeasy blood and tissue kit according to the manufacturer's instructions (Qiagen, cat. no. 69504). Whole-genome sequencing (150 bp paired end) was performed by Novogene. Short Illumina reads were mapped to the unique regions of integrated transgenes with BWA V0.7.17 (Li 2013). We calculated copy-number estimates with the formula: (mapped transgene reads * genome size) / (total genomic mapped reads / mapped transgene length). All copy-number estimates are per haploid genome.

Integrated array assembly from Oxford Nanopore sequencing

High molecular weight genomic DNA was extracted from CFJ157 (*kstIs6* [pSEM371; pSEM233; pSEM231, 1kb ladder] III) using a MagAttract HMW DNA kit following the manufacturer's instructions (Qiagen, cat #67563). Oxford Nanopore sequencing was performed by Novogene (China) using a PromethION instrument. We removed *C. elegans* reads by mapping base-called and quality filtered reads to the *C. elegans* reference genome (WBcel235) with transgene sequence homology and 40 kb of the reference sequence flanking the insertion site masked. Unmapped reads (corresponding to the integrated array, flanking genomic sequences and *E. coli* genome) were assembled with the Canu long-read assembler using default settings except for: 'corOutCoverage=9999' 'minOverlapLength=2500' 'minReadLength=2500' 'corMinCoverage=0' 'corMaxEvidenceErate=0.15' (Koren *et al.* 2017). Array contigs were identified by aligning integrated transgene sequences to contigs with MUMmer V4 (Marçais *et al.* 2018). A single contig (~5Mb) corresponding to the *E. coli* genome was excluded from further analysis.

Supplementary Materials

Figure S1 - Two-step protocol to generate initial MosTI landing sites

Figure S2 - Generating new MosTI landing site.

Figure S3 - MosTI overview.

Figure S4 - Position effect variegation following array integration.

Figure S5 - Indels in integration fragments within integrated arrays.

Figure S6 - Fluorophore expression is unchanged after ten generations for integrated array.

Figure S7 - Sequencing read coverage to detect large-scale duplications and deletions .

Supplementary Table S1 - Strains, plasmids, insertion sites, SNPs CFJ42 and CFJ77

Supplementary File 1 - GenBank annotated plasmid sequence files.

References

Aljohani M. D., S. El Mouridi, M. Priyadarshini, A. M. Vargas-Velazquez, and C. Frøkjær-Jensen, 2020 Engineering rules that minimize germline silencing of transgenes in simple extrachromosomal arrays in C. elegans. Nat Commun 11: 6300. https://doi.org/10.1038/s41467-020-19898-0

Andrews S., 2010 FastQC: A Quality Control Tool for High Throughput Sequence Data

Aram R., K. MacGillivray, C. Li, and A. Saltzman, 2019 Tools for Mos1-mediated single copy insertion (MosSCI) with excisable unc-119(+) or NeoR (G418) selection cassettes. MicroPubl Biol 2019. https://doi.org/10.17912/micropub.biology.000146

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 genetics.114.169730. https://doi.org/10.1534/genetics.114.169730

Au V., E. Li-Leger, G. Raymant, S. Flibotte, G. Chen, *et al.*, 2019 CRISPR/Cas9 Methodology for the Generation of Knockout Deletions in Caenorhabditis elegans. G3 (Bethesda) 9: 135–144. https://doi.org/10.1534/g3.118.200778

Bessereau J.-L., A. Wright, D. C. Williams, K. Schuske, M. W. Davis, *et al.*, 2001 Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. Nature 413: 70–74. https://doi.org/10.1038/35092567

Boer C. G. de, E. D. Vaishnav, R. Sadeh, E. L. Abeyta, N. Friedman, *et al.*, 2020 Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nature Biotechnology 38: 56–65. https://doi.org/10.1038/s41587-019-0315-8

Brenner S., 1974 The genetics of Caenorhabditis elegans. Genetics 77: 71–94.

Chen C., L. A. Fenk, and M. de Bono, 2013 Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. Nucleic Acids Res. https://doi.org/10.1093/nar/gkt805

Chiu H., H. T. Schwartz, I. Antoshechkin, and P. W. Sternberg, 2013 Transgene-Free Genome Editing in Caenorhabditis elegans Using CRISPR-Cas. Genetics. https://doi.org/10.1534/genetics.113.155879

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. https://doi.org/10.1534/genetics.113.155853

Cingolani P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. https://doi.org/10.4161/fly.19695

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. https://doi.org/10.1126/science.1231143

Davis M. W., J. J. Morton, D. Carroll, and E. M. Jorgensen, 2008 Gene Activation Using FLP Recombinase in C. elegans. PLOS Genetics 4: e1000028. https://doi.org/10.1371/journal.pgen.1000028

Davis M. W., and E. M. Jorgensen, 2022 ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program. Frontiers in Bioinformatics 2.

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. https://doi.org/10.1038/nmeth.2641

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. https://doi.org/10.1534/genetics.115.178335

Dokshin G. A., K. S. Ghanta, K. M. Piscopo, and C. C. Mello, 2018 Robust Genome Editing With Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in Caenorhabditis elegans. Genetics genetics.301532.2018. https://doi.org/10.1534/genetics.118.301532

El Mouridi S., C. Lecroisey, P. Tardy, M. Mercier, A. Leclercq-Blondel, *et al.*, 2017 Reliable CRISPR/Cas9 Genome Engineering in Caenorhabditis elegans Using a Single Efficient sgRNA and an Easily Recognizable Phenotype. G3 (Bethesda) 7: 1429–1437. https://doi.org/10.1534/g3.117.040824

El Mouridi S., Y. Peng, and C. Frøkjær-Jensen, 2020 Characterizing a strong pan-muscular promoter (Pmlc-1) as a fluorescent co-injection marker to select for single-copy insertions. MicroPubl Biol 2020. https://doi.org/10.17912/micropub.biology.000302

El Mouridi S., S. AlHarbi, and C. Frøkjær-Jensen, 2021 A histamine-gated channel is an efficient negative selection marker for C. elegans transgenesis. MicroPubl Biol 2021. https://doi.org/10.17912/micropub.biology.000349

Engler C., R. Kandzia, and S. Marillonnet, 2008 A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3: e3647. https://doi.org/10.1371/journal.pone.0003647

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. https://doi.org/10.1371/journal.pone.0005553

Etchberger J. F., and O. Hobert, 2008 Vector-free DNA constructs improve transgene expression in C. elegans. Nature Methods 5: 3–3. https://doi.org/10.1038/nmeth0108-3

Evans T., 2006 Transformation and microinjection. WormBook. https://doi.org/10.1895/wormbook.1.108.1

Farboud B., and B. J. Meyer, 2015 Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. Genetics 199: 959–971. https://doi.org/10.1534/genetics.115.175166

Farboud B., A. F. Severson, and B. J. Meyer, 2019 Strategies for Efficient Genome Editing Using CRISPR-Cas9. Genetics 211: 431–457. https://doi.org/10.1534/genetics.118.301775

Flytzanis N. C., C. N. Bedbrook, H. Chiu, M. K. M. Engqvist, C. Xiao, *et al.*, 2014 Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. Nature Communications 5: 4894. https://doi.org/10.1038/ncomms5894

Friedland A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaiácovo, G. M. Church, *et al.*, 2013 Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods 10: 741–743. https://doi.org/10.1038/nmeth.2532

Frøkjær-Jensen C., M. Wayne 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. https://doi.org/10.1038/ng.248

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. https://doi.org/10.1038/nmeth.1865

Frøkjær-Jensen C., M. W. Davis, M. Sarov, J. Taylor, S. Flibotte, *et al.*, 2014 Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat Meth 11: 529–534. https://doi.org/10.1038/nmeth.2889

Frøkjær-Jensen C., N. Jain, L. Hansen, M. W. Davis, Y. Li, *et al.*, 2016 An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the C. elegans Germline. Cell 166: 343–357. https://doi.org/10.1016/j.cell.2016.05.072

Gibson D. G., L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison 3rd, *et al.*, 2009 Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6: 343–345. https://doi.org/10.1038/nmeth.1318

Gilbert L. A., M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, *et al.*, 2014 Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 159: 647–661. https://doi.org/10.1016/j.cell.2014.09.029

Giordano-Santini R., S. Milstein, N. Svrzikapa, D. Tu, R. Johnsen, *et al.*, 2010 An antibiotic selection marker for nematode transgenesis. Nat Meth 7: 721–723. https://doi.org/10.1038/nmeth.1494

Hashemi N. A., A. C. F. Bergs, C. Schüler, A. R. Scheiwe, W. S. Costa, *et al.*, 2019 Rhodopsin-based voltage imaging tools for use in muscles and neurons of Caenorhabditis elegans. PNAS 116: 17051–17060. https://doi.org/10.1073/pnas.1902443116

Heger P., B. Marin, and E. Schierenberg, 2009 Loss of the insulator protein CTCF during nematode evolution. BMC Molecular Biology 10: 84. https://doi.org/10.1186/1471-2199-10-84

Ho J. W. K., Y. L. Jung, T. Liu, B. H. Alver, S. Lee, *et al.*, 2014 Comparative analysis of metazoan chromatin organization. Nature 512: 449–452. https://doi.org/10.1038/nature13415

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. https://doi.org/10.1126/science.1225829

Katic I., and H. Großhans, 2013 Targeted Heritable Mutation and Gene Conversion by Cas9-CRISPR in Caenorhabditis elegans. Genetics. https://doi.org/10.1534/genetics.113.155754

Kaymak E., B. M. Farley, S. A. Hay, C. Li, S. Ho, *et al.*, 2016 Efficient generation of transgenic reporter strains and analysis of expression patterns in Caenorhabditis elegans using library MosSCI. Developmental Dynamics 245: 925–936. https://doi.org/10.1002/dvdy.24426

Kelly W. G., S. Xu, M. K. Montgomery, and A. Fire, 1997 Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics 146: 227–238.

Kerr R., V. Lev-Ram, G. Baird, P. Vincent, R. Y. Tsien, *et al.*, 2000 Optical imaging of calcium transients in neurons and pharyngeal muscle of C. elegans. Neuron 26: 583–594.

Kim H., T. Ishidate, K. S. Ghanta, M. Seth, D. Conte, *et al.*, 2014 A Co-CRISPR Strategy for Efficient Genome Editing in Caenorhabditis elegans. Genetics 197: 1069–1080. https://doi.org/10.1534/genetics.114.166389

Koren S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, *et al.*, 2017 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27: 722–736. https://doi.org/10.1101/gr.215087.116

Leyva-Díaz E., N. Stefanakis, I. Carrera, L. Glenwinkel, G. Wang, *et al.*, 2017 Silencing of Repetitive DNA Is Controlled by a Member of an Unusual Caenorhabditis elegans Gene Family. Genetics 207: 529–545. https://doi.org/10.1534/genetics.117.300134

Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Lin Z., Y. Xie, W. Nong, X. Ren, R. Li, *et al.*, 2021 Formation of artificial chromosomes in Caenorhabditis elegans and analyses of their segregation in mitosis, DNA sequence composition and holocentromere organization. Nucleic Acids Research 49: 9174–9193. https://doi.org/10.1093/nar/gkab690

Liu T., A. Rechtsteiner, T. A. Egelhofer, A. Vielle, I. Latorre, *et al.*, 2010 Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res. https://doi.org/10.1101/gr.115519.110

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. https://doi.org/10.1534/genetics.113.155382

Logsdon G. A., M. R. Vollger, and E. E. Eichler, 2020 Long-read human genome sequencing and its applications. Nat Rev Genet 21: 597–614. https://doi.org/10.1038/s41576-020-0236-x

Long L., H. Guo, D. Yao, K. Xiong, Y. Li, *et al.*, 2015 Regulation of transcriptionally active genes via the catalytically inactive Cas9 in *C. elegans and D. rerio*. Cell Research 25: 638–641. https://doi.org/10.1038/cr.2015.35

Maduro M., and D. Pilgrim, 1995 Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. Genetics 141: 977–988.

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. https://doi.org/10.1126/science.1232033

Marçais G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, *et al.*, 2018 MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol 14: e1005944. https://doi.org/10.1371/journal.pcbi.1005944

McDiarmid T. A., V. Au, D. G. Moerman, and C. H. Rankin, 2020 Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome editing in Caenorhabditis elegans. PLoS One 15: e0238950. https://doi.org/10.1371/journal.pone.0238950

Mello C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros, 1991 Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. EMBO J 10: 3959–3970.

Mello C., and A. Fire, 1995 Chapter 19 DNA Transformation, pp. 451–482 in *Methods in Cell Biology*, Caenorhabditis elegans: Modern Biological Analysis of an Organism. edited by Epstein H. F., Shakes D. C. Academic Press.

Merritt C., D. Rasoloson, D. Ko, and G. Seydoux, 2008 3' UTRs are the primary regulators of gene expression in the C. elegans germline. Curr. Biol 18: 1476–1482. https://doi.org/10.1016/j.cub.2008.08.013

Moreno-Mateos M. A., C. E. Vejnar, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, *et al.*, 2015 CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nature Methods 12: 982–988. https://doi.org/10.1038/nmeth.3543

Noma K., and Y. Jin, 2018 Rapid Integration of Multi-copy Transgenes Using Optogenetic Mutagenesis in Caenorhabditis elegans. G3: Genes, Genomes, Genetics 8: 2091–2097. https://doi.org/10.1534/g3.118.200158

Nonet M. L., 2020 Efficient Transgenesis in Caenorhabditis elegans Using Flp Recombinase-Mediated Cassette Exchange. Genetics. https://doi.org/10.1534/genetics.120.303388

Nonet M., 2021 Additional Landing Sites for Recombination-Mediated Cassette Exchange in C. elegans. MicroPubl Biol 2021. https://doi.org/10.17912/micropub.biology.000503

Okkema P. G., S. W. Harrison, V. Plunger, A. Aryana, and A. Fire, 1993 Sequence requirements for myosin gene expression and regulation in Caenorhabditis elegans. Genetics 135: 385–404.

Paix A., Y. Wang, H. E. Smith, C.-Y. S. Lee, D. Calidas, *et al.*, 2014 Scalable and Versatile Genome Editing Using Linear DNAs with Micro-Homology to Cas9 Sites in Caenorhabditis elegans. Genetics genetics.114.170423. https://doi.org/10.1534/genetics.114.170423

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. https://doi.org/10.1534/genetics.115.179382

Paix A., H. Schmidt, and G. Seydoux, 2016 Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs. Nucl. Acids Res. gkw502. https://doi.org/10.1093/nar/gkw502

Poplin R., V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, *et al.*, 2018 Scaling accurate genetic variant discovery to tens of thousands of samples. 201178.

Praitis V., E. Casey, D. Collar, and J. Austin, 2001 Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics 157: 1217–1226.

Priyadarshini M., J. Z. Ni, A. M. Vargas-Velazquez, S. G. Gu, and C. Frøkjær-Jensen, 2022 Reprogramming the piRNA pathway for multiplexed and transgenerational gene silencing in C. elegans. Nat Methods 19: 187–194. https://doi.org/10.1038/s41592-021-01369-z

Radman I., S. Greiss, and J. W. Chin, 2013 Efficient and rapid C. elegans transgenesis by bombardment and hygromycin B selection. PLoS ONE 8: e76019. https://doi.org/10.1371/journal.pone.0076019

Robert V., and J.-L. Bessereau, 2007 Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. EMBO J. 26: 170–183. https://doi.org/10.1038/sj.emboj.7601463

Schwartz M. L., and E. M. Jorgensen, 2016 SapTrap, a Toolkit for High-Throughput CRISPR/Cas9 Gene Modification in Caenorhabditis elegans. Genetics 202: 1277–1288. https://doi.org/10.1534/genetics.115.184275

Semple J. I., R. Garcia-Verdugo, and B. Lehner, 2010 Rapid selection of transgenic C. elegans using antibiotic resistance. Nat Meth 7: 725–727. https://doi.org/10.1038/nmeth.1495

Silva-García C. G., A. Lanjuin, C. Heintz, S. Dutta, N. M. Clark, *et al.*, 2019 Single-Copy Knock-In Loci for Defined Gene Expression in Caenorhabditis elegans. G3: Genes, Genomes, Genetics 9: 2195–2198. https://doi.org/10.1534/g3.119.400314

Stevenson Z. C., M. J. Moerdyk-Schauwecker, B. Jamison, and P. C. Phillips, 2020 Rapid self-selecting and clone-free integration of transgenes into engineered CRISPR safe harbor locations in Caenorhabditis elegans. bioRxiv 2020.05.26.117390. https://doi.org/10.1101/2020.05.26.117390

Stinchcomb D. T., J. E. Shaw, S. H. Carr, and D. Hirsh, 1985 Extrachromosomal DNA transformation of Caenorhabditis elegans. Mol. Cell. Biol. 5: 3484–3496.

Tyson J. R., N. J. O'Neil, M. Jain, H. E. Olsen, P. Hieter, *et al.*, 2018 MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res. 28: 266–274. https://doi.org/10.1101/gr.221184.117

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. https://doi.org/10.1534/genetics.113.156075

Vallin E., J. Gallagher, L. Granger, E. Martin, J. Belougne, *et al.*, 2012 A genome-wide collection of Mos1 transposon insertion mutants for the C. elegans research community. PLoS ONE 7: e30482. https://doi.org/10.1371/journal.pone.0030482

Voutev R., and E. J. A. Hubbard, 2008 A "FLP-Out" System for Controlled Gene Expression in Caenorhabditis elegans. Genetics 180: 103–119. https://doi.org/10.1534/genetics.108.090274

Waaijers S., V. Portegijs, J. Kerver, B. B. C. G. Lemmens, M. Tijsterman, *et al.*, 2013 CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis elegans. Genetics. https://doi.org/10.1534/genetics.113.156299

Wang H., J. Liu, S. Gharib, C. M. Chai, E. M. Schwarz, *et al.*, 2017 cGAL, a temperature-robust GAL4–UAS system for *Caenorhabditis elegans*. Nature Methods 14: 145–148. https://doi.org/10.1038/nmeth.4109

Wang H., J. Liu, K. P. Yuet, A. J. Hill, and P. W. Sternberg, 2018 Split cGAL, an intersectional strategy using a split intein for refined spatiotemporal transgene control in Caenorhabditis elegans. PNAS 115: 3900–3905. https://doi.org/10.1073/pnas.1720063115

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. https://doi.org/10.1534/genetics.114.172361

Wei X., C. J. Potter, L. Luo, and K. Shen, 2012 Controlling gene expression with the Q repressible binary expression system in Caenorhabditis elegans. Nat. Methods 9: 391–395. https://doi.org/10.1038/nmeth.1929

Wheeler B. S., E. Anderson, C. Frøkjær-Jensen, Q. Bian, E. Jorgensen, *et al.*, 2016 Chromosome-wide mechanisms to decouple gene expression from gene dose during sex-chromosome evolution. Elife 5. https://doi.org/10.7554/eLife.17365

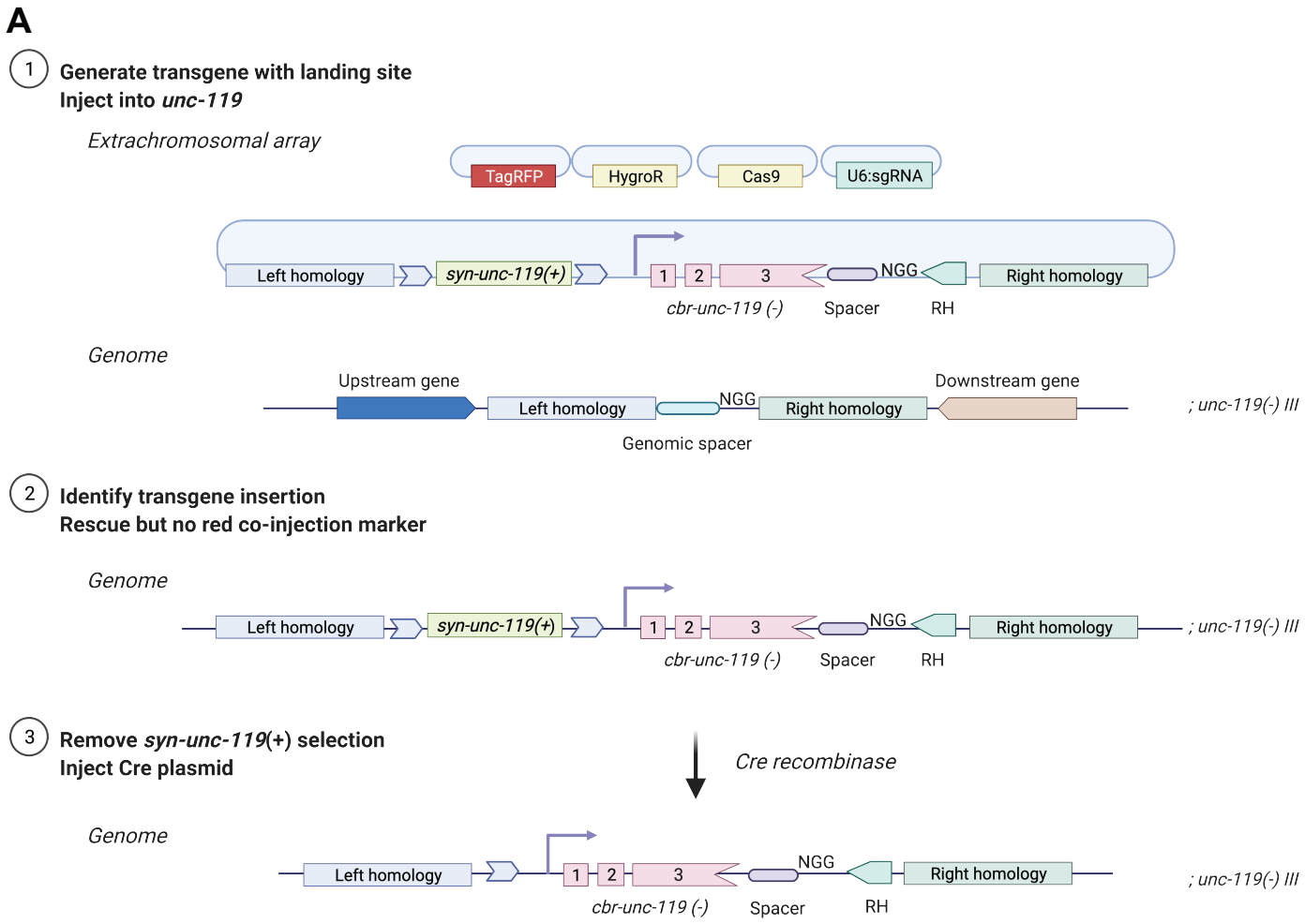
Woglar A., K. Yamaya, B. Roelens, A. Boettiger, S. Köhler, *et al.*, 2020 Quantitative cytogenetics reveals molecular stoichiometry and longitudinal organization of meiotic chromosome axes and loops. PLOS Biology 18: e3000817. https://doi.org/10.1371/journal.pbio.3000817

Yemini E., A. Lin, A. Nejatbakhsh, E. Varol, R. Sun, *et al.*, 2021 NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell 184: 272-288.e11. https://doi.org/10.1016/j.cell.2020.12.012

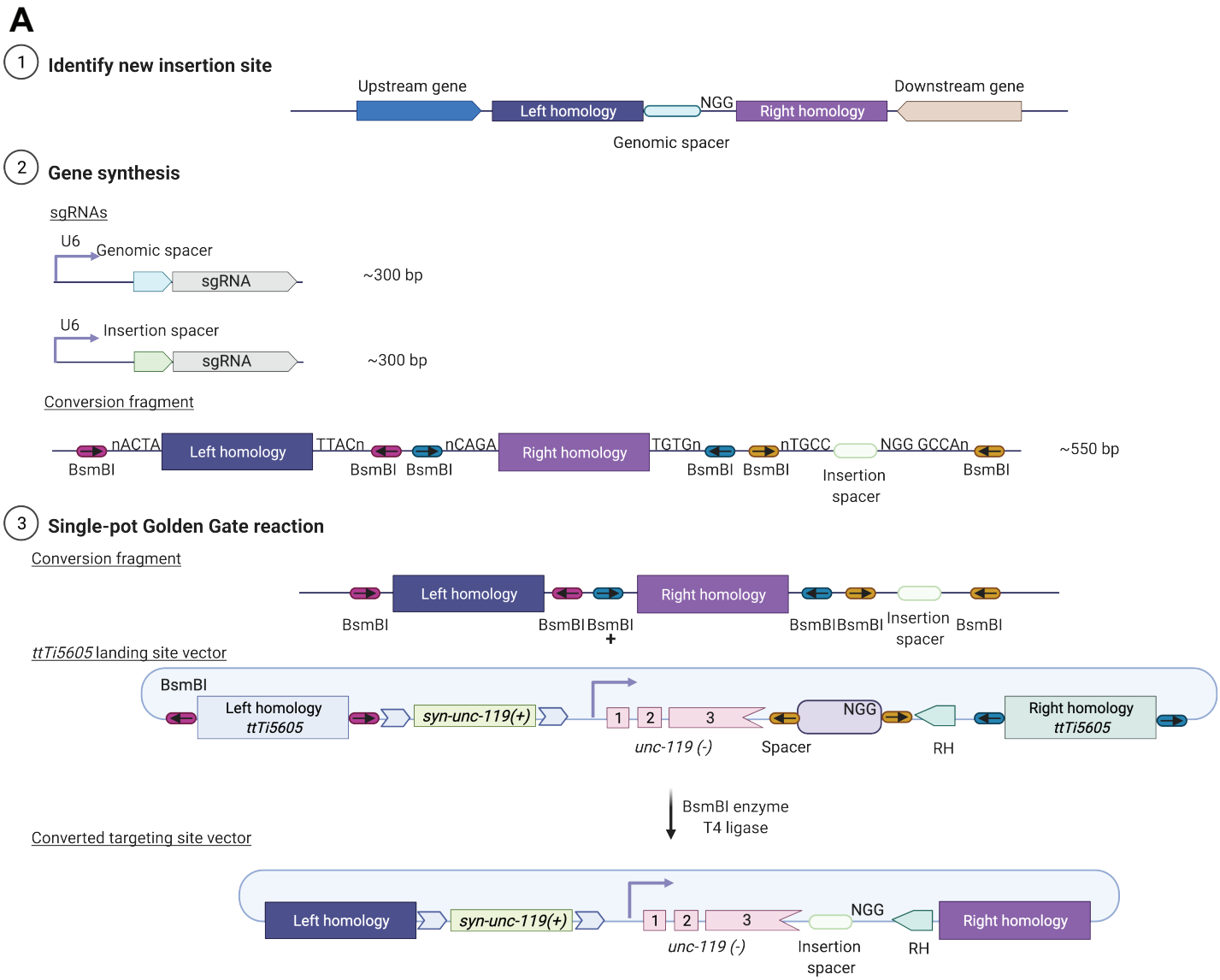
Yoshina S., Y. Suehiro, E. Kage-Nakadai, and S. Mitani, 2016 Locus-specific integration of extrachromosomal transgenes in C. elegans with the CRISPR/Cas9 system. Biochemistry and Biophysics Reports 5: 70–76. https://doi.org/10.1016/j.bbrep.2015.11.017

Zeiser E., C. Frøkjær-Jensen, E. Jorgensen, and J. Ahringer, 2011 MosSCI and gateway compatible plasmid toolkit for constitutive and inducible expression of transgenes in the C. elegans germline. PLoS ONE 6: e20082. https://doi.org/10.1371/journal.pone.0020082

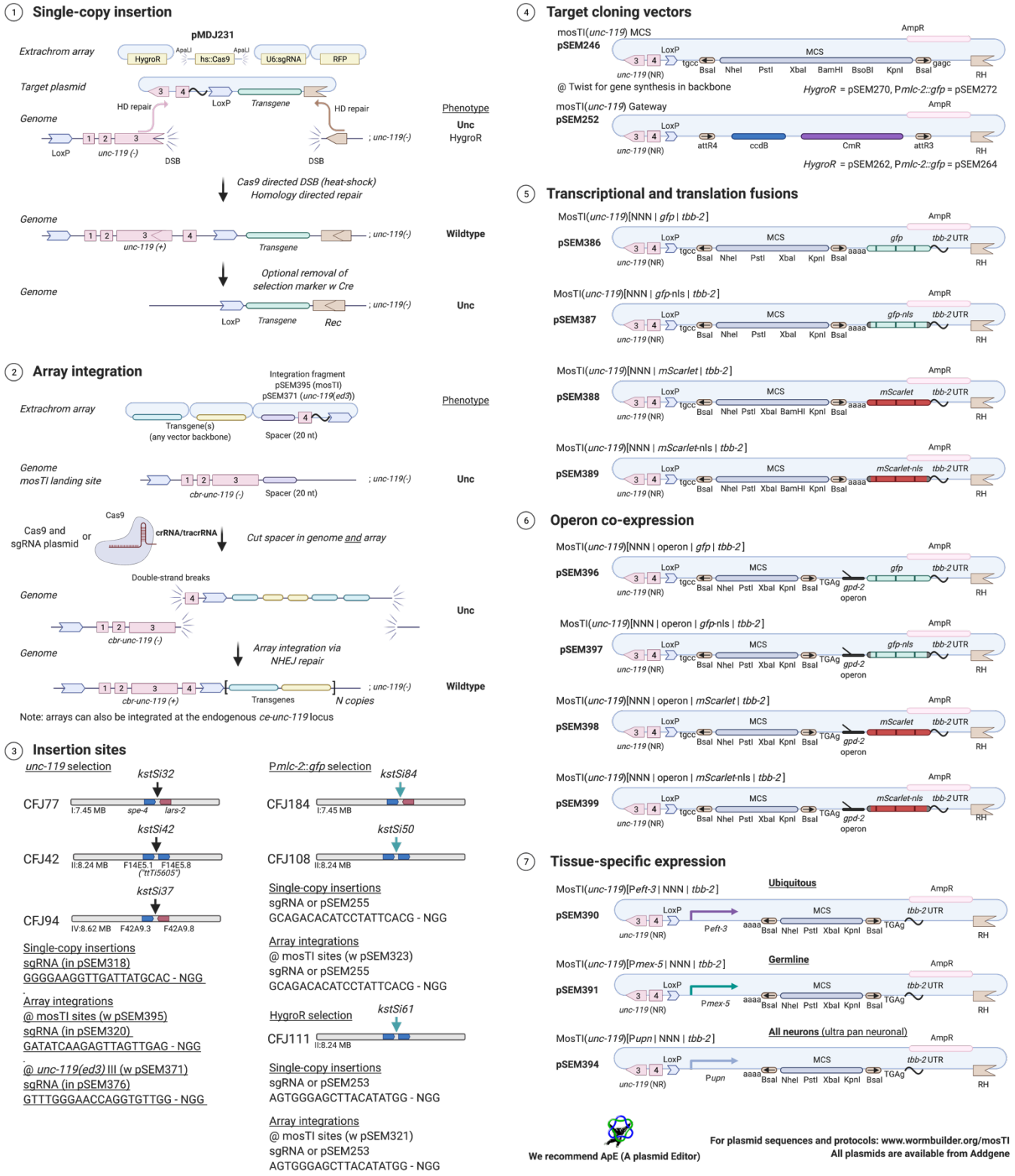
Zhang L., J. D. Ward, Z. Cheng, and A. F. Dernburg, 2015 The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development 142: 4374–4384. https://doi.org/10.1242/dev.129635



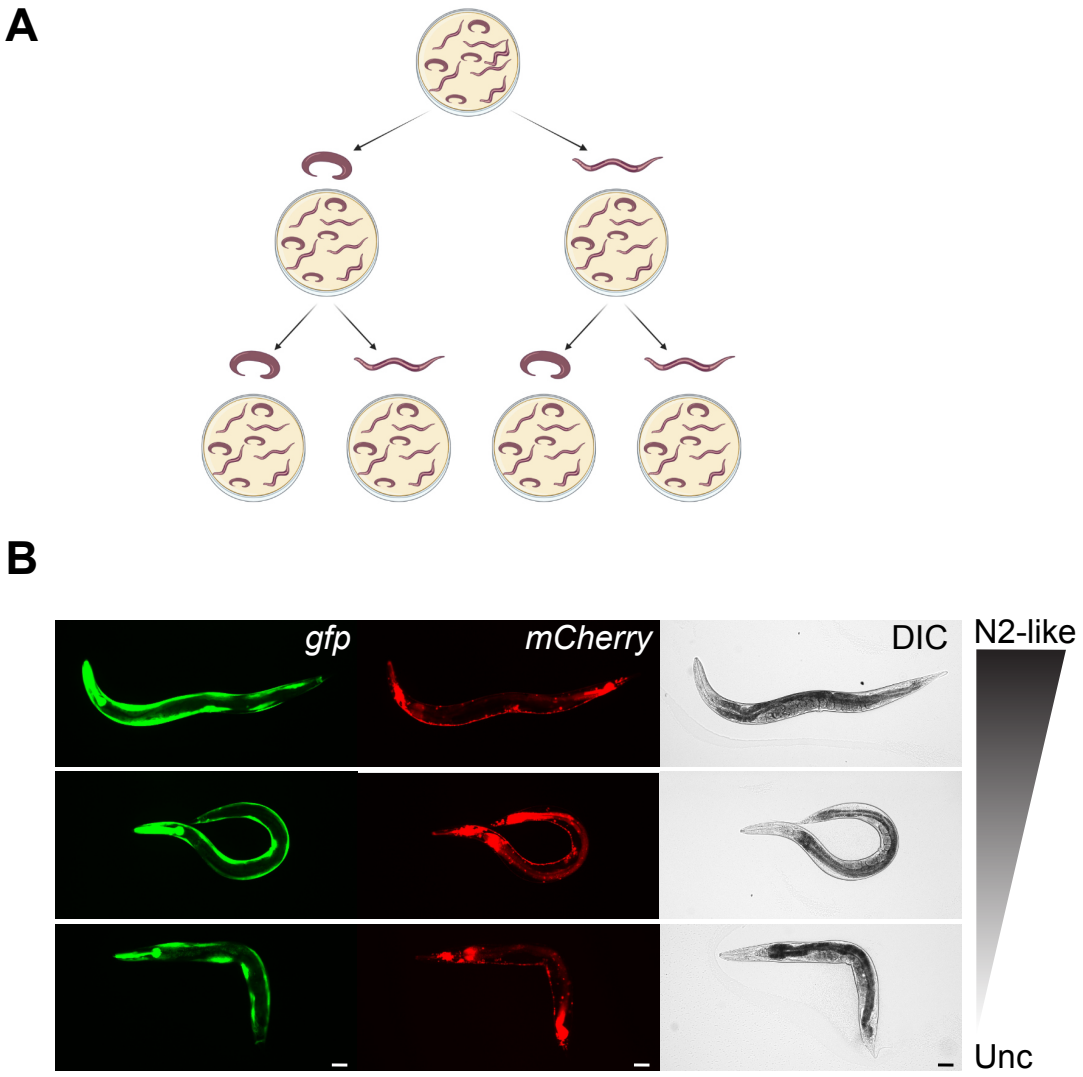
Supplementary Figure 1 | Two-step protocol to generate initial MosTI landing sites
A. 1. As a first step, we generated a targeting construct that contains homology regions (left and right homology) to the insertion site, a coding-optimized *unc-119* rescue fragment ("*syn-unc-119(+)*"), the split *cbr-unc-119* marker ("*cbr-unc-119(-)*"), a spacer sequence, and a right homology region for transgene insertions ("RH"). We selected sites in permissive chromatin domains (Ho *et al.* 2014) and between two convergently transcribed genes to minimize the effect of nearby promoters and enhancers. The targeting construct, fluorescent co-injection markers ("tag-RFP"), an antibiotic selection marker ("HygroR"), plasmids expressing Cas9, and sgRNA are injected into a strain with the *unc-119*(*ed3*) mutation. 2. Animals with the landing site inserted at the target genomic location were identified based on Unc rescue and loss of the extra-chromosomal array (based on loss of fluorescent co-injection markers). 3. Finally, *syn-unc-119(+)* selection was removed from the landing site by injecting a plasmid that expresses Cre recombinase and screening for Unc animals.



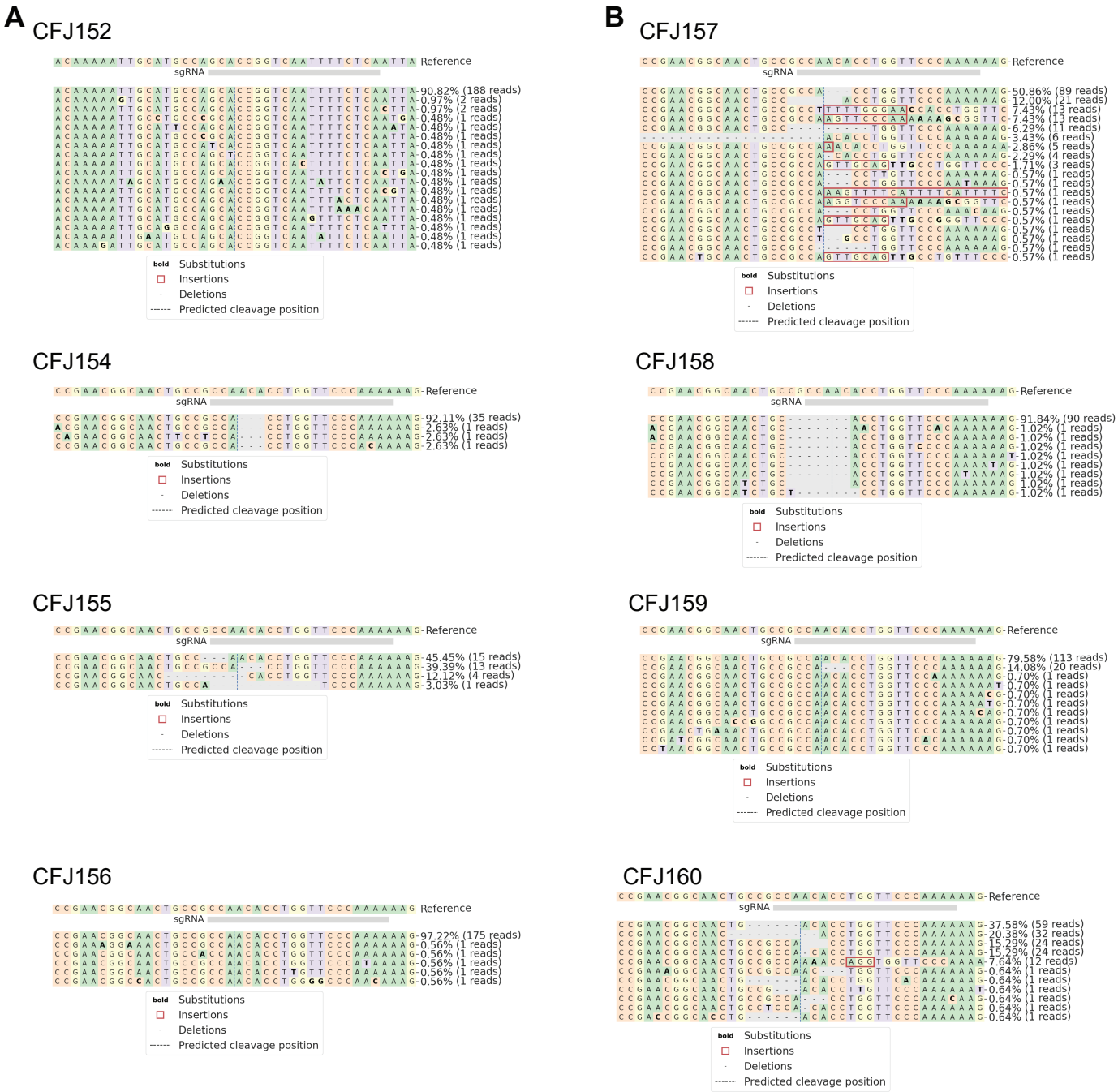
Supplementary Figure 2 | Generating new MosTI landing site.
A. 1. As a first step, we identified appropriate insertion sites. For "safe-harbor" landing sites, we selected insertion sites located in permissive chromatin (Ho *et al.* 2014) and between the 3'UTRs of two convergently transcribed endogenous genes. 2. In a second step, we synthesized two sgRNAs: one for inserting the landing site and one for inserting transgenes at the landing site (optional - not required if using the same sgRNA as used for the standard set of MosTI landing sites). We synthesized a conversion fragment containing left and right homology regions (250 - 500 bp) and a novel insertion spacer flanked by the appropriate Golden-Gate cloning overhangs (indicated in figure). Note: any BsmBI sites in the homology regions and spacer must be removed. 3. We performed a Golden-Gate reaction using BsmBI between the conversion fragment and the *ttTi5605* MosTI landing site vector (pSEM203). This reaction exchanged the homology arms and the insertion site spacer but maintained the codon-optimized *syn-unc-119(+)* and split *cbr-unc-119* selection marker.



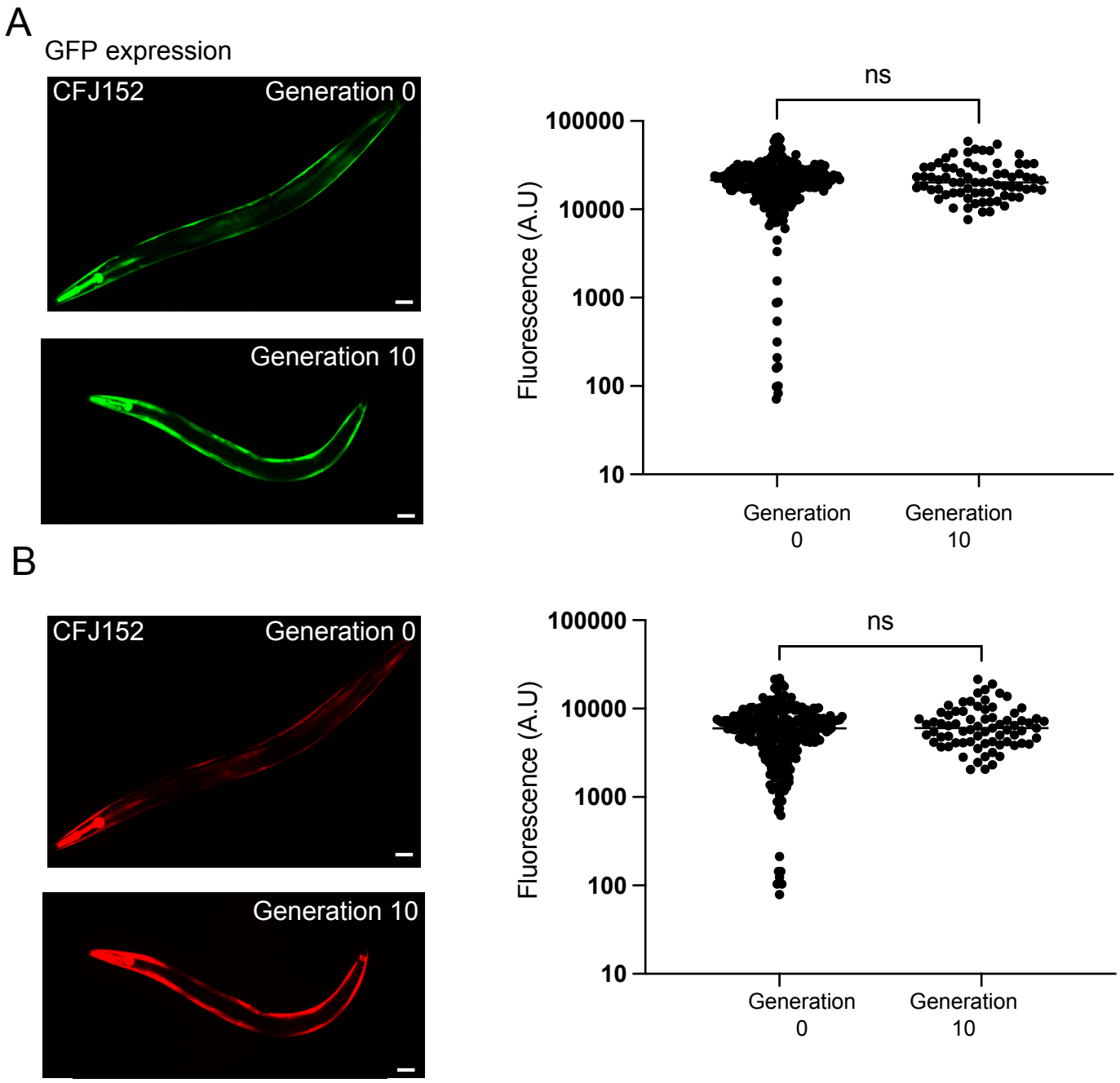
Supplementary Figure 3 | MosTI overview.
This figure contains a single overview of MosTI methods and reagents.
1. Schematic overview for generating MosTI single-copy insertions.
2. Schematic overview for generating MosTI array integrations.
3. Overview of single-copy MosTi sites.
4. Cloning vectors for MosTi single-copy insertions. All vectors are compatible with restriction enzyme cloning into the multiple cloning site (MCS) or Golden-gate based cloning using BsaI. Note that pSEM246 has been deposited with Twist Bioscience for direct gene synthesis.
5. MosTI vectors for transcriptional and translation fusions (*unc-119*-based selection).
6. MosTI vectors for fluorophore co-expression using the *gpd-2* operon. Transgene expression can be monitored by fluorophore co-expression.
7. Tissue-specific MosTI expression vectors. P*eft-3* is ubiquitously expressed (including the germline), P*mex-5* is germline specific (Zeiser et al. 2011), and Pupn is pan-neuronal ("ultra pan neuronal")(Yemini et al. 2021).



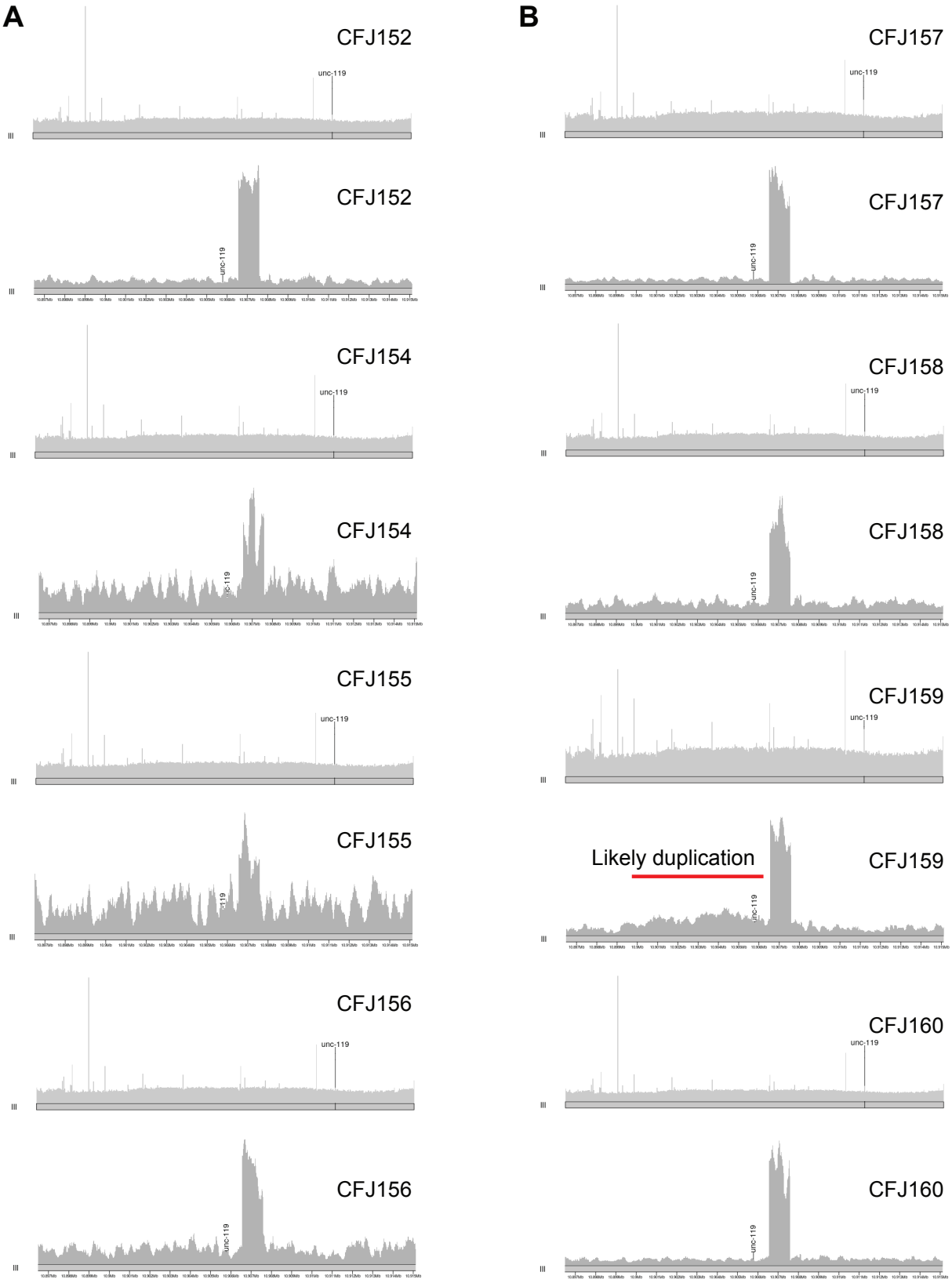
Supplementary Figure 4 | Position effect variegation following array integration.
A. Schematic overview of position effect variegation after targeted integration of an extra-chromosomal array into a MosTI site by the "two-step heat-shock plasmid" protocol. We observed *unc-119* rescue indicating that an NHEJ event had happened between the genomic location and the array, but we were unable to generate a homozygous rescued strain. We observed a continuum of phenotypes from fully rescued to non-rescued animals suggesting phenotypic variegation. Non-rescued animals would segregate a mix of rescued and non-rescued animals, like the segregation pattern of rescued animals. **B.** Fluorescence images of rescued and non-rescued transgenic animals. All animals contained the injected fluorescence markers regardless of the degree of rescue. We were also unable to detect any obvious mosaicism suggesting that the strains did not contain an extra-chromosomal array with unusual segregation. Scale bar = 20 microns, 10x magnification.



Supplementary Figure 5 | Indels in integration fragments within integrated arrays.
We analyzed Illumina short-read whole-genome sequence data (150 bp paired-end reads) using CRISPResso2 (Clement *et al.* 2019) to determine the integrity and copy-number number of the integration fragment within MosTI integrants. The integration fragments (and the genomic location) are cut by the sgRNA, and parts of the extra-chromosomal array are integrated by non-homologous end-joining. The number of different alleles with reasonable read coverage (>1 read) are a reasonable lower bound for the copy-number of the integration plasmid. See **Figure 6A** for plasmid-copy number estimates based on sequence coverage. **A.** Array integrations generated by the "one-step plasmid" protocol. **B.** Array integrations generated by the "two-step protein" protocol.



Supplementary Figure 6 | Fluorophore expression is unchanged after ten generations for integrated array.
We propagated the strain CFJ152 (*kstIs1*[pCFJ2474 (Cas9); pSEM371(integration fragment); pSEM376 (sgRNA1), pSEM377(sgRNA2) ;pSEM233 (P*mlc-1*::*tagRFP*); pSEM231 (P*mlc-1*::*gfp*), 1kb ladder] III) for ten generations without selecting for any fluorescence to determine if the integrated array changed over time. **A.** GFP expression at generation 0 and generation 10. **B.** TagRFP expression at generation 0 and generation 10.
Left, fluorescence images of adult animals. Scale bar = 20 microns. Right. Quantification of fluorescence levels using COPAS large-particle flow sorter (animals selected with time-of-flight between 700-900 to select mainly adult animals). Statistics: Mann-Whitney t-test.



Supplementary Figure 7 | Sequencing read coverage to detect large-scale duplications and deletions.
Read coverage for eight strains carrying integrated arrays at the *unc-119*(*ed3*) locus. **A.** Array integrations generated by the "one-step plasmid" protocol. **B.** Array integrations generated by the "two-step protein" protocol. Note, the likely short duplication (approximately 6-7 kb) at the *unc-119* locus in CFJ159.