# Part I

## Biology of Endonucleases (Zinc-Finger Nuclease, TALENs and CRISPRs) and Regulatory Networks

# 1 Targeted Genome Editing Techniques in *C. elegans* and Other Nematode Species

Behnom Farboud and Te-Wen Lo

## 1.1 Introduction

*C. elegans* is a genetically tractable model organism that has pioneered studies to understand cell fate decisions (Sulston *et al.*, 1983), apoptosis (Conradt and Xue, 2005; Ellis and Horvitz, 1986; Ellis *et al.*, 1991; Hengartner *et al.*, 1992), RNA interference (RNAi) (Fire *et al.*, 1998), micro RNAs (Lee *et al.*, 1993; Reinhart *et al.*, 2000) and cellular aging (Dorman *et al.*, 1995; Kenyon *et al.*, 1993; Kimura *et al.*, 1997; Lin *et al.*, 1997; Morris *et al.*, 1996; Ogg *et al.*, 1997), to mention a few. It is a 1 mm free-living, non-parasitic nematode that has two sexes: males and hermaphrodites. The ability to culture self-fertile hermaphrodites allows for easy maintenance of isogenic strains. Additionally, their transparent cuticle has allowed tracking of the fate of every cell, from fertilization to adulthood, to generate a complete cell lineage map that has been pivotal for cell fate studies (Sulston *et al.*, 1983). And in a seminal effort, *C. elegans* became the first multicellular eukaryotic organism to have its entire genome sequenced (*C. elegans* Sequencing Consortium, 1998), yielding a well-defined genetic map and opening the door for development of numerous molecular tools to dissect gene function.

To study gene function, *C. elegans* researchers have utilized both forward and reverse genetic strategies. A common forward genetic screen introduces genetic lesions using mutagens such as ethyl methane sulfonate (EMS) and screens for the desired phenotype (Brenner, 1974). In addition to forward genetic screens, two reverse genetic methods were developed. One powerful reverse genetic tool is RNA interference (RNAi) (Fire *et al.*, 1998). RNAi-mediated knockdown of known target genes results in loss-of- or reduction-of-function phenotypes. Due to the inability to guarantee complete reduction of gene function and the transient nature of RNAi, scientists developed additional methods to create and screen for gene deletions. Individuals could make their own frozen deletion library (Edgley *et al.*, 2002; Jansen *et al.*, 1997) or request pre-isolated mutant strains from two groups: the *C. elegans* Gene Knockout Consortium (1998), and the Million Mutation Project (Thompson *et al.*, 2013). Unfortunately, since these mutants are generated randomly, there are no guarantees that there is a mutation in your

gene of interest or that the deletion occurs in a functional domain of the gene. Therefore, more precise methods for targeted gene manipulation were desired.

Introduction of custom transgenes into *C. elegans* allowed more direct dissection of genetic components (Figure 1.1). Microinjection of DNA plasmids encoding genes of interest results in the formation of heritable, multi-copy extra-chromosomal arrays (Mello *et al.*, 1991). These arrays, however, are not stable and are lost at an unpredictable rate during cell division. In order to create stable transgenic lines, extrachromosomal arrays can be integrated into the genome by UV irradiation (Mitani, 1995) or microparticle bombardment (Praitis *et al.*, 2001). These methods generate endogenous, low-copy transgenes, which are expressed at levels closer to that of native genes; however, single-copy integrates are rare. Also, using these methodologies, it is impossible to control for the site of integration. Utilization of existing Mos techniques (Boulin and Bessereau, 2007) overcame many of these challenges. MosSCI and miniMos techniques target transgenes for single-copy insertion at defined chromosomal loci (Frøkjær-Jensen *et al.*, 2012; 2014). Notably, single-copy transgenes can be expressed in the germline where extrachromosomal arrays are often silenced. One limitation of these methods, however, is that transgenes can only be inserted in locations where Mos sites already reside, thus limiting transgene placement.

The rapid development of next generation sequencing (NGS) technology has elevated the power of *C. elegans* as a model organism, by allowing molecular comparisons between *C. elegans* and closely related species for evolutionary

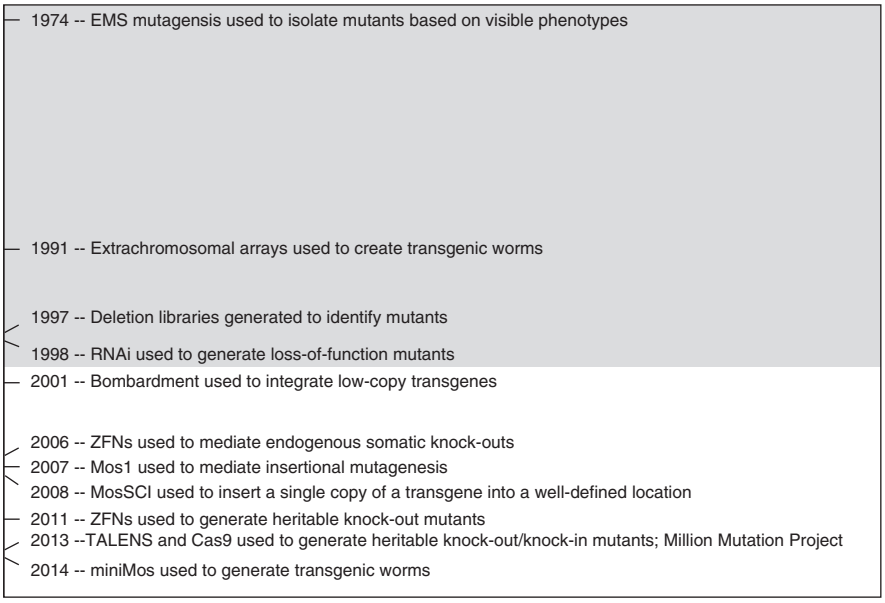


**Figure 1.1** Timeline of *C. elegans* mutant isolation strategies. Strategies shown in the gray area occurred prior to the sequencing of the *C. elegans* genome and strategies shown in the white area occurred after.

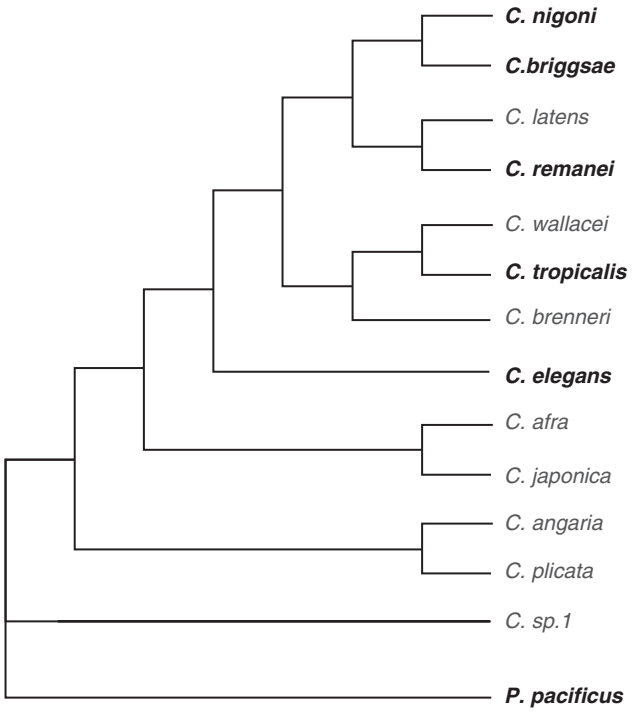


**Figure 1.2** Partial *Caenorhabditis* phylogeny (using data from Félix *et al.*, 2014; Kiontke *et al.*, 2011) of sequenced or partially sequenced (gray) species. Genome-edited species are shown in bold. Sequencing status is according to 959 Nematode Genomes.

studies (Figures 1.1, 1.2). Just in the past decade, dozens of related nematodes have been identified, significantly broadening the nematode phylum. Improvements in NGS technology have reduced the time and cost of sequencing new species and the ease of evolutionary comparisons between species that can further our understanding of all aspects of biology that would not be possible by studying one species alone.

Nematodes, and particularly *C. elegans*, have benefited from numerous molecular tools available to study gene function. However, conspicuously absent was the ability to precisely edit the genome to introduce custom mutations. With ZFNs, TALENs and now CRISPR/Cas9, programmable nucleases have revolutionized the study of nematode biology. Below, we discuss how these nucleases have been used in nematodes and how they have the potential to transform any nematode species into a model organism amenable for biological research.

## 1.2 ZFN-mediated Genome Editing in Nematodes

Zinc-finger nucleases (ZFNs) are customizable DNA binding molecules fused to FokI (Figure 1.3A; Urnov *et al.*, 2010). FokI is a restriction endonuclease that functions as an obligate dimer. Once a FokI dimer is bound to duplex DNA, the DNA cleavage domains are activated and both DNA strands are cleaved. The DNA

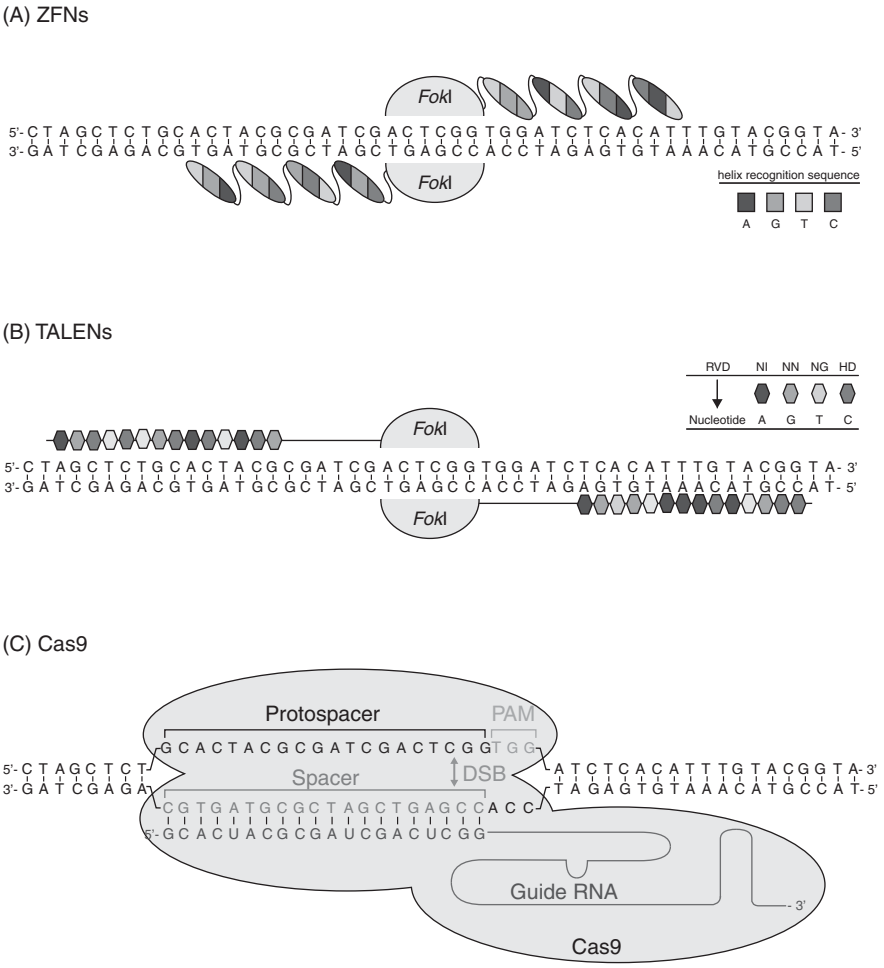


**Figure 1.3** ZFN, TALEN and Cas9 DNA recognition. (A) Shown is a ZFN pair bound to its DNA target. Each monomer is composed of a DNA binding domain fused to a FokI cleavage domain. The DNA binding domain is assembled from several zinc-finger DNA recognition domains (four are shown, depicted as ovals) and each domain can recognize three unique nucleotides, noted in different colors. Strung together, the multiple domains can dictate DNA binding to unique genomic sequences with high specificity. When two monomers are brought together by binding adjacent DNA sequences, the FokI domains dimerize and the nuclease is active. (B) A TALEN pair bound to its DNA target. As with ZFNs, TALENs are composed of a DNA binding domain fused to the FokI cleavage domain. The binding specificity is dictated by the repeat variable diresidues (RVD) within each TALE repeat unit (shown are the different RVDs that bind each nucleotide). (C) Cas9 is targeted to DNA sequences by a bound sgRNA. If a protospacer adjacent motif (PAM) is present, the first 20 nucleotides of the guide RNA can form Watson–Crick base pairs with the target spacer sequence, and activate Cas9 nuclease activity. (A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

binding domains of ZFNs can be designed to target specific sequences in an organism's genome. Each zinc-finger DNA binding domain makes contact primarily with a separate DNA triplet (Pabo *et al.*, 2001; Pavletich and Pabo, 1991) and can be assembled in unique combinations to target specific DNA sequences

within the genome. Increasing the number of zinc-finger domains in each zinc-finger protein increases specificity. Typically, 3–4 zinc-finger domains are assembled resulting in 9–12 bp specificity. Taken together with the need for FokI to form a dimer for cleavage, this results in 24 bp specificity for each ZFN pair.

In 2006, Morton *et al.* demonstrated that custom-engineered ZFNs can generate targeted double-strand breaks (DSBs) in somatic tissues of *C. elegans* by injecting plasmids encoding ZFNs (Morton *et al.*, 2006). Repair of these breaks by non-homologous end joining (NHEJ) results in small insertions and deletions at the desired locus. Unfortunately, these mutations are somatic and not heritable. The lack of heritable mutagenesis may result from the absence of germline nuclease expression. Morton *et al.* suggested that germline expression was undetectable because ZFN expression from transgene arrays was suppressed by RNAi. Building upon these early successes, Wood *et al.* demonstrated heritable germline mutagenesis with ZFNs in *C. elegans* and *C. briggsae* by injecting mRNA into the germline, in order to avoid silencing associated with transgene arrays (Wood *et al.*, 2011). Wood *et al.* chose to edit the endogenous *ben-1* locus due to its easily scored mutant phenotype; mutants are mobile on media containing benomyl. ZFN-encoding mRNAs with 5′ and 3′ untranslated regions (UTRs) permissive for germline expression were injected to obtain *ben-1*/+ F1 progeny. Methods were also established to screen F1 progeny for mutations without a visible phenotype using the endonuclease CEL-1, which recognizes and cleaves heteroduplex DNA at sites of mismatches. PCR amplicons that contain the ZFN-targeted site were generated from potential F1 mutant heterozygotes. The amplicons from F1s were then pooled, denatured and slowly reannealed. Mutagenized amplicons will randomly anneal to wild-type amplicons and CEL-1 can recognize and cleave the mismatched bases. Mutant CEL-1-positive clones can be easily detected by resolving the cleaved, mismatched PCR product from the wild-type full-length PCR fragment by agarose gel electrophoresis.

To demonstrate that ZFN-directed mutagenesis could be ported to other nematode species, Wood *et al.* targeted the *sdc-2* gene in *C. briggsae*. Without any modifications to the protocols established in *C. elegans*, ZFNs were successfully utilized to isolate custom mutations in *C. briggsae* (Wood *et al.*, 2011), thus demonstrating that these genome editing methods could be applied across species to achieve targeted genome editing.

## 1.3 TALEN-mediated Genome Editing in Nematodes

Similar to ZFNs, transcription activator-like effector nucleases (TALENs) consist of a customizable DNA binding domain (TALE domain) fused to the FokI endonuclease (Figure 1.3B). The TALE domain consists of modular repeating units that are 33 or 34 amino acids long. Within each repeat unit, the 12th and 13th residues, termed the repeat variable diresidues (RVDs), are unique, and dictate which DNA base a particular repeat unit will bind. By stringing together repeat units with unique RVDs, TALENs could be easily designed to predictably target DNA (Boch *et al.*, 2009; Bogdanove *et al.*, 2010; Moscou and Bogdanove, 2009). As with ZFNs,

TALENs function as dimers, so two TALENs must be targeted to adjacent DNA sequences to reconstitute an active nuclease. Typical TALE domains contain 15–20 RVDs where the first RVD targets a "T" with 15–30 base pairs between TALEs. Similarly to ZFNs, there is a requirement of dimerization for FokI cleavage, so two TALEs are required for function, thus increasing specificity up to 40 base pairs.

Using protocols established for ZFN mutant isolation, Wood *et al.* engineered TALE nucleases to target genes in *C. elegans* and *C. briggsae* (Figure 1.2). The proof-of-principle experiment targeted TALENs to the same *ben-1* locus that was successfully targeted using ZFNs. Rates of TALEN-mediated genome editing of *ben-1* were comparable to rates of ZFN-mediated genome editing. To further demonstrate the portability of custom nucleases, TALENs were used to generate mutations in another nematode *C. nigoni* (previously named *C. species 9*) and *P. pacificus* (Figure 1.2). Wei *et al.* further adapted protocols to also utilize TALENs in both *C. briggsae* and *C. tropicalis* (Figure 1.2) (Wei *et al.*, 2014).

The editing discussed thus far is imprecise; after the double-strand break, DNA is repaired through NHEJ, resulting in small insertions and deletions. To demonstrate that specific site-directed changes could also be introduced at the DSB, Lo *et al.* used TALENs to insert three different restriction sites into the *ben-1* locus. The authors co-injected mRNA encoding TALENs directed at the *ben-1* locus with three single-stranded oligos. Each oligo possessed a unique restriction site flanked by sequence that was homologous to the TALEN's target site. Along with small insertion and deletion alleles, precise engineered knockins possessing the unique restriction site sequences were recovered (Lo *et al.*, 2013).

## 1.4 Cas9-mediated Genome Editing in Nematodes

Studies of the bacterial and Archael adaptive immune systems that utilize clustered regularly interspersed short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins have led to the identification of a family of RNA-guided endonucleases that can readily be repurposed to target and cleave double-stranded DNA sequences with very few constraints (reviewed in Carroll, 2014; Doudna and Charpentier, 2014; Hsu *et al.*, 2014; Mali *et al.*, 2013b). The most well-studied family member is Cas9, a component of the type IIa CRISPR system. Cas9 is targeted to genomic loci by a guide RNA that encodes a 20-nucleotide region of homology to the DNA target of choice (Figure 1.3C; Garneau *et al.*, 2010; Jinek *et al.*, 2012; Mojica *et al.*, 2009). Cas9-mediated DNA target recognition and cleavage has only one sequence requirement, a trinucleotide motif adjacent to the 20-nucleotide DNA target sequence, called the protospacer adjacent motif (PAM) (Anders *et al.*, 2014; Garneau *et al.*, 2010; Jinek *et al.*, 2012; Sapranauskas *et al.*, 2011; Sternberg *et al.*, 2014). For Cas9, the PAM is NGG. The RNA guide naturally consists of two separate RNAs: a CRISPR RNA (crRNA) that can form Watson–Crick base pairs with the DNA target (spacer) backbone and a transactivating RNA (trRNA) that binds the crRNA and forms a ternary complex with Cas9 (Figure 2.3C). This can be further simplified and a chimeric single-guide RNA (sgRNA) made from fusing the crRNA and trRNA (Jinek *et al.*, 2012).

The minimal requirement for target design and the straightforward construction of guide RNAs have made Cas9 a very attractive alternative to other site-specific nucleases, which have more stringent requirements for target selection and require more effort to design and construct. Ease of implementation has led to widespread use of CRISPR/Cas9, from bacteria and yeast to plants and metazoans, including nematodes.

In the following sections, we will discuss general considerations for performing Cas9-mediated genome editing in nematodes. While the majority of studies in nematodes have involved *C. elegans*, most principles can be applied to other nematode species and some of these principles have already been adapted to accomplish genome editing in other nematodes, including *C. briggsae, C. tropicalis, C. remanei* (Yin and Haag, unpublished data), *C. nigoni* and *P. pacificus* (Figure 1.2; Lo *et al.*, 2013; Markov *et al.*, 2016; Witte *et al.*, 2015). Finally, we discuss newer applications of CRISPR systems in nematodes.

### 1.4.1 Delivery of Cas9 and RNA Guides into Nematodes

Cas9 and guide RNAs can be introduced into nematodes using three different strategies, all relying on microinjection. With the first method, mRNA encoding Cas9 and RNA guides are delivered by co-injection into the gonad of young adult hermaphrodites (Chiu *et al.*, 2013; Katic and Großhans, 2013; Lo *et al.I*, 2013). Initial studies using this approach blindly scored offspring of injected animals for the presence of the desired mutation, assuming the mutation did not have a phenotype. While successful, this approach often required screening in excess of 1000 worms to identify the desired mutation. The second alternative approach of injecting DNA encoding Cas9 and the sgRNA along with transformation markers reduced the burden of screening (Figure 1.4A) (Dickinson *et al.*, 2013; Friedland *et al.*, 2013; Waaijers *et al.*, 2013). Only worms expressing transformation makers, and presumably Cas9 and sgRNAs, were examined and the odds of isolating the desired mutant were improved.

There are several parameters to consider when performing DNA-based introduction of Cas9 and guide RNAs into *C. elegans*. To ensure Cas9 is present early enough in development to generate heritable genomic modifications, constructs are codon-optimized, and germline-permissive 5′ and 3′ UTRs are utilized to maximize transcriptional and translational efficiency (Dickinson *et al.*, 2013; Friedland *et al.*, 2013; Waaijers *et al.*, 2013). Several constructs employ the *eft-3* promoter and the *tbb-2* 3′ UTR, while an alternate vector uses a heat shock promoter to drive expression of Cas9. Heat shock induction produces short bursts of Cas9 expression, potentially reducing embryonic lethality observed using the *eft-3* promoter and reducing the chances for off-target cleavage by the continual expression of Cas9 (Waaijers *et al.*, 2013).

Two of the most widely used vectors, *Peft-3::cas9-SV40-NLS::tbb-2* 3′UTR and pDD162, drive Cas9 expression with the *eft-3* promoter and *tbb-2* 3′ UTR (Dickinson *et al.*, 2013; Friedland *et al.*, 2013), but minimal differences between the two vectors have a dramatic impact on mutagenesis efficiency. A recent study found that the sequence of the flexible linker between Cas9 and the nuclear

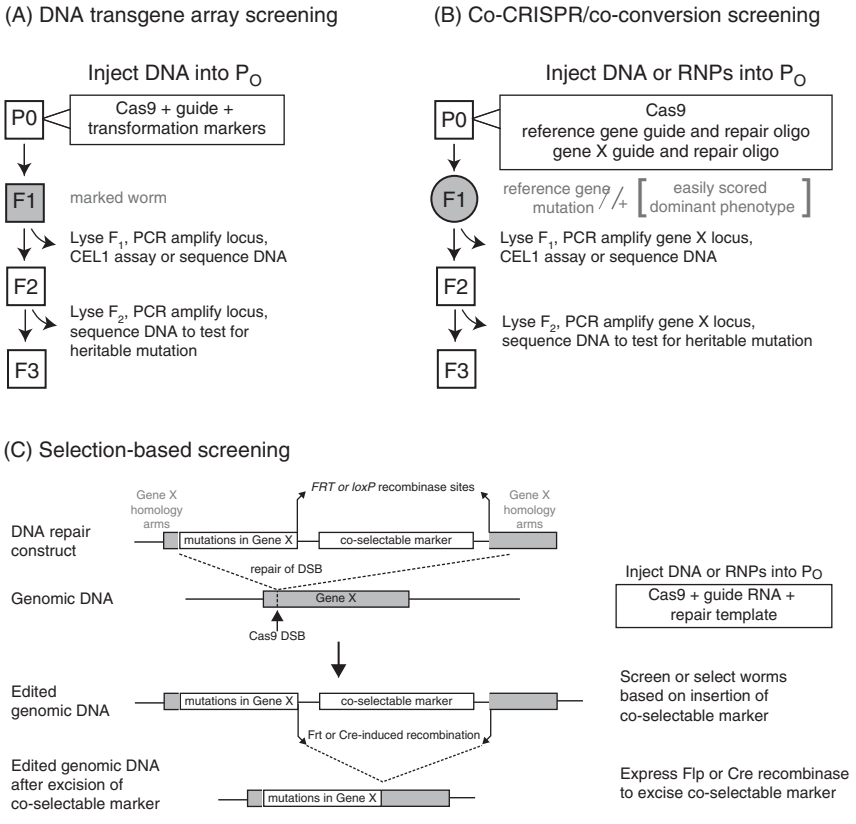


**Figure 1.4** Three different strategies for isolating mutants generated by Cas9. (A) DNA transgene array screening: P0 hermaphrodites are microinjected with DNA encoding Cas9, a sgRNA and co-transformation markers. In some F1 offspring, the injected DNA forms transgene arrays and worms are selected based on the expression of the transformation markers from the arrays (e.g. expression of fluorescent makers). After F2 progeny are laid, worms are examined for the presence of the desired mutation. (B) Co-CRISPR/co-conversion screening: P0 animals are microinjected with DNA encoding Cas9 and sgRNAs or RNPs targeting a gene of interest and a co-conversion reference gene (Arribere *et al.*, 2014; Kim *et al.*, 2014; Ward, 2015). Appropriate editing of a reference gene using a co-injected repair oligonucleotide yields F1s with an easily scored dominant phenotype. F1 worms mutant for the reference gene are selected and the gene of interest is examined. (C) Selection-based screening: P0s are injected with DNA encoding Cas9, a sgRNA and a DNA repair construct. Repair using the template results in insertion of the desired mutation and a co-selectable marker flanked by loxP or Frt recombination sites (Chen *et al.*, 2013b; Dickinson *et al.*, 2013; 2015). After selection or screening of worms based on insertion of the co-selectable marker, expression of Flp or Cre recombinase leads to excision of the selectable marker, leaving behind the mutation of interest.

localization signal accounts for the impaired efficiency of *Peft-3::cas9-SV40-NLS::tbb-2* 3′UTR (Zhao *et al.*, 2016).

Guide RNA expression vectors utilize U6 small RNA promoters that are transcribed by RNA polymerase III (polIII) (Dickinson *et al.*, 2013; Friedland *et al.*, 2013; Waaijers *et al.*, 2013). The first 20 nucleotides of the guide are DNA-target-specific, so successful expression of guides depends on predictable transcriptional initiation and polIII transcriptional initiation from the U6 promoter is well