bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Title

Split-wrmScarlet and split-sfGFP: tools for faster, easier fluorescent labeling of endogenous proteins in *Caenorhabditis elegans*

Authors

Jérôme Goudeau[1,3], Jonathan Paw[1], Laura Savy[2], Manuel D. Leonetti[2], Andrew G. York[1], Cynthia Kenyon[1,3], and Maria Ingaramo[1,3]

[1]*Calico Life Sciences LLC, South San Francisco, California 94080, United States*
[2]*Chan Zuckerberg Biohub, San Francisco, California 94158, United States*
[3]Corresponding authors: Calico Life Sciences LLC, 1170 Veterans Blvd., South San Francisco, CA 94080. E-mail: jerome@calicolabs.com, cynthia@calicolabs.com and ingaramo@calicolabs.com

Abstract

We create and share a new red fluorophore, along with a set of strains, reagents and protocols, to make it faster and easier to label endogenous *C. elegans* proteins with fluorescent tags. CRISPR-mediated fluorescent labeling of *C. elegans* proteins is an invaluable tool, but it is much more difficult to insert fluorophore-size DNA segments than it is to make small gene edits. In principle, high-affinity asymmetrically split fluorescent proteins solve this problem in *C. elegans*: the small fragment can quickly and easily be fused to almost any protein of interest and can be detected wherever the large fragment is expressed and complemented. There is currently only one available strain stably expressing the large fragment of a split fluorescent protein, restricting this solution to a single tissue (the germline) in the highly autofluorescent green channel. No available *C. elegans* lines express unbound large fragments of split red fluorescent proteins, and even state-of-the-art split red fluorescent proteins are dim compared to the canonical split-sfGFP protein. In this study, we engineer a bright, high-affinity new split red fluorophore, split-wrmScarlet, and generate transgenic *C. elegans* lines to allow easy single-color labeling in muscles and dual-color labeling in somatic cells. We validate these strains by targeting split-wrmScarlet to several genes whose products label distinct organelles, and we provide a protocol for an easy, cloning-free method for CRISPR/Cas9 editing.

Keywords

C. elegans, Cas9, CRISPR, genome engineering, GFP, protein localization, mScarlet

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Genetically-expressed fluorophores are essential tools for visualizing and quantifying cellular proteins. In *C. elegans*, fluorescent proteins have traditionally been introduced on extrachromosomal arrays (Kimble *et al.* 1982; Mello 1991) or via MosSCI-based integration (Frøkjær-Jensen *et al.* 2008; Frøkjær-Jensen *et al.* 2012). These methods have enabled important discoveries but can also lead to artifacts due to supraphysiological gene-expression levels and lack of endogenous regulatory control. In recent years, the repertoire of *C. elegans* transgenic tools has expanded [see (Nance and Frøkjær-Jensen 2019) for review], particularly due to advances in CRISPR/Cas9 genome-editing technologies (Paix *et al.* 2014; Dickinson and Goldstein 2016). CRISPR/Cas9 allows precise transgene insertion by homology-directed repair (HDR) and can be used to label an endogenous gene at its native locus with a fluorescent protein (Dokshin *et al.* 2018; Farboud *et al.* 2019; Vicencio *et al.* 2019).

However, relative to CRISPR/Cas9-mediated integration of smaller transgenes, genomic insertion of large DNA fragments like those encoding fluorescent proteins remains a challenge, both because repair with double-stranded templates is less efficient than repair with single-stranded oligodeoxynucleotide donors (ssODN) (Farboud *et al.* 2019), and because of the requirement for cloning to prepare the HDR donor template. Recent methods such as 'hybrid' (Dokshin *et al.* 2018) and 'nested' (Vicencio *et al.* 2019) CRISPR remove the need for cloning but still require preparation of the DNA template or several rounds of injections and selection of transgenic progeny. As a result, using CRISPR with small ssODN templates is currently faster, easier, cheaper and more efficient than with large templates. In our lab, we routinely make *C. elegans* genome edits with short ssODN with almost guaranteed success. In contrast, in our experience, large edits using double-stranded DNA templates are more time-consuming and have higher failure rates.

Our preferred approach is to combine the utility of full-length fluorescent proteins with the convenience of short genomic edits, by using high-affinity asymmetrically-split fluorescent proteins (Cabantous *et al.* 2004). These fluorophores typically separate a GFP-like protein between the 10th and 11th strands of the beta barrel, splitting it asymmetrically into a large (FP_{1-10}) and a small (FP_{11}) fragment. The fragments are not individually fluorescent, but upon binding one another, recapitulate the fluorescent properties of an intact fluorophore (Figure 1A). Unlike the low-affinity split fluorescent proteins used in BiFC assays (Hu *et al.* 2002), high-affinity binding between the fragments is critical here. Our preferred approach for tagging a new cellular protein begins with a *C. elegans* strain expressing the large FP_{1-10} fragment in cells of interest, unattached to any cellular protein. This way, only the small FP_{11} fragment (<60 nt) needs to be inserted to tag the target protein, which will only fluoresce in compartments where it can bind the large fragment. These short insertions tend to be faster, easier, and more reliable to make than inserting a >600 nt full-length fluorescent protein (Paix

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

et al. 2015; Prior *et al.* 2017, Dokshin *et al.* 2018; Richardson *et al.* 2018). Therefore, collections of *C. elegans* lines stably expressing the large FP_{1-10} in different tissues are an invaluable resource allowing rapid fluorescent tagging in a cell type of choice. Stable lines with red FP_{1-10} fragments would be especially useful, given *C. elegans'* substantial autofluorescence in the GFP channel.

Green and red asymmetrically-split fluorescent proteins have been used to combine cell and protein specificity in *C. elegans* neurons and synapses (Noma *et al.* 2017; He *et al.* 2019; Feng *et al.* 2019); however, these strains used extrachromosomal arrays, not stable lines, which are more time-consuming to maintain and can have variable expression levels. To the best of our knowledge, there is only one available unbound FP_{1-10} stable *C. elegans* line, which expresses $sfGFP_{1-10}$ in the germline (Hefel and Smolikove 2019), and there are no available lines with red FP_{1-10} fragments. Existing red split fluorophores are also much dimmer in *C. elegans* than green ones, despite recent improvements like split-sfCherry3 (Feng *et al.* 2019).

Here, we describe tools that reduce these obstacles for convenient fluorescent labeling of endogenous *C. elegans* proteins. We engineered a new split red fluorescent protein based on mScarlet (Bindels *et al.* 2016; El Mouridi *et al*. 2017), which is three times brighter in worms than split-sfCherry3 (https://www.addgene.org/138966/). We generated *C. elegans* lines carrying single-copy insertions of $wrmScarlet_{1-10}$ expressed in somatic cells (https://cgc.umn.edu/strain/CF4582) and in muscle (https://cgc.umn.edu/strain/CF4610), and are making them available to the *C. elegans* community. We provide a protocol for an easy, cloning-free method to label endogenous genes with FP_{11}s using CRISPR/Cas9, commercially available synthetic single-stranded oligodeoxynucleotide (ssODN) donors, and microinjection (https://www.protocols.io/view/step-by-step-guide-to-tag-endogenous-genes-with-sp-bamkic4w). We validate this protocol by targeting $wrmScarlet_{11}$ to six different genes whose products have distinct cellular locations. We also show that labeling with tandem $wrmScarlet_{11}$-repeats increases fluorescence *in vivo*, and we provide the plasmid necessary to generate the dsDNA template through Addgene (link pending, expected in early August 2020). We also generated a strain expressing an integrated copy of $sfGFP_{1-10}$ (Pédelacq *et al.* 2005) in somatic cells (https://cgc.umn.edu/strain/CF4587). Finally, to further expand the toolkit, we generated a dual-color strain expressing both $sfGFP_{1-10}$ and $wrmScarlet_{1-10}$ in somatic cells (https://cgc.umn.edu/strain/CF4588), for two-color applications such as colocalization studies or organelle interaction. We hope that these resources will facilitate the study of *C. elegans* biology.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MATERIALS AND METHODS

Mutagenesis and screening. For the initial screenings in *E. coli,* we introduced a 32 amino-acid spacer between the tenth and eleventh β-strands of full-length wrmScarlet in a pRSET vector (Feng *et al.* 2017). This starting construct was non-fluorescent, but we restored low fluorescence levels by introducing the superfolder mutation G220A. Semi-random mutagenesis was carried out using rolling-circle amplification with NNK primers at positions I8, K10, F15, G32, Q43, A45, K46, L47, G52, G53, D60, S63, P64, Q65, F66, S70, R71, T74, K75, D79, Y84, W94, R96, T107, V108, Q110, E115, L125, R126, T128, K139, K140, W144, E145, S147, T148, E149, R150, I162, K163, M164, L175, F178, K179, K183, K185, K186, N195, R198, I202, T203, S204, D208, Y209, T210, V211, V212, E213, Q214, Y215, E216, R217, S218, E219, A220, H222, S223, T224, G225, G226, M227, D228, and E229 with Phusion polymerase (NEB) in GC buffer, followed by pooling of the PCR products, *DpnI* digestion and transformation into BL21(DE3) *E. coli*. These positions covered areas deemed important for brightness or stability, and the interface between FP_{11} and FP_{1-10}. Primers were resynthesized if a mutation interfered with neighboring mutagenic primer binding. The brightest three to five colonies were identified using a Leica M165 FC fluorescent stereomicroscope, and their plasmid DNA subjected to a new mutagenesis round. After five rounds, we separated the two fragments of a version of split wrmScarlet that had fluorescence comparable to the parent protein into two *S. cerevisiae* plasmids to test for complementation. Since we did not detect fluorescence, we continued selection using two plasmids in yeast. For screening on two plasmids, a pRSET vector expressing $mScarlet_{1-10}$ and a pD881-MR vector (ATUM) expressing mTagBFP-$mScarletS_{11}$ (without the MDELYK tail from the C-terminus) were used to perform the semi-random mutagenesis. The libraries were co-electroporated into *E. coli* and expression induced with 1% rhamnose and 1mM IPTG. The library was enriched for fluorescent clones using FACS, and then subcloned to make pRS-GPD-$wrmScarlet_{1-10}$ and p416-TEF-membrane-mTagBFP-$wrmScarlet_{11}$. The yeast plasmids were co-transformed into a URA^-, HIS^-, LEU^-, MET^- *S. cerevisiae* strain and selected for in SC media without uracil and histidine, and FACS was used again for enrichment of clones with the highest red to blue ratio. After three rounds of semi-random mutagenesis with the two plasmids strategy, a final round of random mutagenesis was performed using the GeneMorph II kit (Agilent). Yeast plasmids will be available through Addgene in early August 2020.

***C. elegans* strains and maintenance.** Animals were cultured under standard growth conditions with *E. coli* OP50 at 20°C (Brenner 1974). Strains generated in this work are listed in the Supplementary Material, Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nucleic acid reagents. Synthetic nucleic acids were purchased from Integrated DNA Technologies (IDT), GenScript or Genewiz. For knock-in of a single wrmScarlet$_{11}$ or sfGFP$_{11}$ sequence, 200-mer HDR templates were ordered in ssODN form (synthetic single-stranded oligodeoxynucleotide donors) from IDT. For knock-in of wrmScarlet$_{11}$ repeats, HDR template were ordered in dsDNA form (plasmids) from GenScript or Genewiz. For plasmids injected as extrachromosomal arrays, sequences were synthesized and cloned into the pUC57 vector (Genewiz). The complete set of crRNAs and DNA sequences used for the experiments described here can be found in Supplementary Material, Table S1.

Strain generation: CRISPR/Cas9-triggered homologous recombination. CRISPR insertions were performed using published protocols (Paix *et al.* 2015, 2016). Ribonucleoprotein complexes (protein Cas9, tracrRNA, crRNA) and DNA templates were microinjected into the gonad of young adults using standard methods (Evans 2006). Injected worms were singled and placed at 25ºC overnight. All crRNA and DNA template sequences used to generate the strains described in this work are listed in the Supplementary Material, Table S1. WrmScarlet$_{11}$ and sfGFP$_{11}$ integrants were identified by screening for fluorescence in the F1 or F2 progeny of injected worms. The co-CRISPR *dpy-10(cn64)* mutation was used as a marker when generating non-fluorescent strains. The CF4582 strain (*muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*) was generated by replacing the *tir-1::mRuby* sequence from the strain CA1200 (*ieSi57 II; unc-119(ed3) III* (Zhang *et al.* 2015) with the *wrmScarlet$_{1-10}$* sequence. The CF4587 strain (*muIs253[(Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*) was generated by replacing the *let-858* promoter from the strain COP1795 (*knuSi785 [pNU1687(Plet-858::sfGFP$_{1-10}$::unc-54 3'UTR, unc-119(+)] II; unc-119(ed3) III*) with the *eft-3* promoter. Both CF4582 and CF4587 strains were generated using long, partially single-stranded DNA donors (Dokshin *et al.* 2018). The CF4610 strain (*muIs257[Pmyo-3::wrmScarlet$_{1-10}$::unc-54 3'UTR] I*) was generated by inserting the *wrmScarlet$_{1-10}$* sequence in the WBM1126 strain following the SKI LODGE protocol (Silva-García *et al.* 2019). The strain PHX731 (*vha-13(syb731[wrmScarlet::vha-13]) V*) was generated by SunyBiotech using CRISPR services. Strains generated were genotyped by Sanger sequencing of purified PCR products (Genewiz).

Strain generation: Mos1-mediated single-copy insertion. The COP1795 strain was generated by NemaMetrix using MosSCI services. The PHX1797 strain was generated by SunyBiotech using MosSCI services.

Strain generation: genetic crosses. The following *C. elegans* strains were created by standard genetic crosses: CF4588 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)], muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-*

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

119(ed3) III) and CF4602 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)], muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III; fib-1(muIs254[wrmScarlet$_{11}$::fib-1]), his-3(muIs255[his-3::sfGFP$_{11}$]) V*). Non-fluorescent parental lines CF4582, CF4587 and CF4610 generated using *dpy-10(cn64)* co-CRISPR were backcrossed at least once.

Strain generation: plasmid microinjection. *Peft-3::3NLS::mTagBFP2::wrmScarlet$_{11}$::T2A::mNeonGreen::wrmScarlet$_{1-10}$::fib-1 3'UTR*, *Peft-3::3NLS::mTagBFP2::sfCherry3$_{11}$::T2A::mNeonGreen::sfCherry3$_{1-10}$::fib-1 3'UTR*, *Pmyo-3::mTagBFP2::wrmScarlet$_{11}$::T2A::mNeonGreen::wrmScarlet$_{1-10}$::fib-1 3'UTR*, or *Pmyo-3::mTagBFP2::sfCherry3$_{11}$::T2A::mNeonGreen::sfCherry3$_{1-10}$::fib-1 3'UTR* constructs were microinjected at (20 ng.μL$^{-1)}$ using a standard microinjection procedure (Mello 1991). Germline gene expression was achieved using a microinjection-based protocol with diluted transgenic DNA (Kelly *et al.* 1997), *Psun-1::mNeonGreen::linker::wrmScarlet$_{11}$::tbb-2 3'UTR* construct (5 ng.μL^{-1}) was co-injected with PvuII-digested genomic DNA fragments from E. coli (100 ng.μL^{-1}).

Microscopy. Confocal fluorescence imaging was performed using the NIS Elements imaging software on a Nikon confocal spinning disk system equipped with an Andor EMCCD camera, a CSU-X1 confocal scanner (Yokogawa), 405, 488, and 561 nm solid-state lasers, and 455/50, 525/26 and 605/70 nm emission filters. Transgenic animals expressing sfGFP$_{11}$ or wrmScarlet$_{11}$ were screened using a Leica M165 FC fluorescent stereomicroscope equipped with a Sola SE-V with GFP and mCherry filters.

Image analysis. Images were analyzed using Fiji. Image manipulations consisted of maximum intensity projections along the axial dimension, rolling ball radius background subtraction, smoothing, and LUT minimum and maximum adjustments. Masks were created by thresholding and setting the pixels under the threshold cutoff to NaN. Plotting of values per pixel was carried out in python 3, using numpy and matplotlib. When performing normalizations for split-sfCherry3 vs split-wrmScarlet, the red channel was divided by the green channel (mNeonGreen-FP$_{1-10}$) because the localization of both fragments is expected to be cytosolic. For normalization of signals where mTagBFP-FP$_{11}$ is targeted to the membrane, the blue channel was used instead.

Mounting worms for microscopy. Pads made of 3% agarose (GeneMate) were dried briefly on Kimwipes (Kimtech) and transferred to microscope slides. Around 10 μL of 2 mM levamisole (Sigma) was pipetted onto the center of the agarose pads. Animals were transferred to the levamisole drop, and a cover slip was placed on top before imaging.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Brood size analysis. Eight single synchronized adults grown at 20°C were transferred to fresh plates every 24 hours until cessation of reproduction, and the number of viable progeny produced by each worm was scored.

Developmental toxicity assay. Ten N2E wild-type animals were microinjected with either *Peft-3::3NLS::mTagBFP2::wrmScarlet*$_{11}$*::T2A::mNeonGreen::wrmScarlet*$_{1-10}$*::fib-1 3'UTR* or *Peft-3::3NLS::mTagBFP2::sfCherry3*$_{11}$*::T2A::mNeonGreen::sfCherry3*$_{1-10}$*::fib-1 3'UTR* construct at (20 ng.μL^{-1}) and were singled. mNeonGreen-positive F1 animals were scored and their development was monitored for up to five days from egg-laying. The number of fluorescent dead eggs, arrested larvae (i.e. animals never reaching adulthood) or adults were scored for each group.

Lifespan assays. NGM plates were supplemented with 5-Fluorouracil (5-FU, Sigma, 15 μM) (Goudeau *et al.* 2011) in order to prevent progeny from hatching and with Kanamycin sulfate to prevent bacterial contamination (Sigma, 25 μg.mL^{-1}). Animals fed with kanamycin-resistant OP50 were scored manually as dead or alive, from their L4 larval stage defined as day 0. A worm was considered alive if it moved spontaneously or, in cases where it wasn't moving, if it responded to a light touch stimulus with a platinum wire. Animals that crawled off the plates, had eggs that accumulated internally, burrowed or ruptured were censored and included in the analysis until the time of censorship.

Structure prediction and rendering of split-wrmScarlet. Phyre2 was used to predict the three-dimensional modelling in intensive mode with default parameters (Kelley *et al.* 2015). The 3D model obtained was visualized using PyMOL (v 2.2.0).

Statistical analysis. Differences in fluorescence intensity between groups were compared using unpaired *t*-test with Welch's correction. Data are presented as means ± SD. Kaplan-Meier estimates of survival curves were calculated using *survival* (v 2.38–3) and *rms* (v 4.5–0) R packages and differences were tested using log-rank test. The number of animals used in each experiment is indicated in the figure legends.

Data availability. Strains expressing a single-copy of wrmScarlet$_{1-10}$ and/or sfGFP$_{1-10}$ (CF4582, CF4587, CF4588 and CF4610) will be made available via the *Caenorhabditis* Genetics Center (CGC). The vector pJG100 carrying *Peft-3::wrmScarlet*$_{1-10}$*::unc-54 3'UTR*, is deposited, along with sequence and map at Addgene. Other strains and plasmids are available upon request. The authors state that all data necessary for confirming the conclusions presented here are represented fully within the article. A detailed protocol to generate *C. elegans* with sfGFP$_{11}$ and/or wrmScarlet$_{11}$ integrants is

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

available at dx.doi.org/10.17504/protocols.io.bamkic4w. Supplemental material will be made available at Figshare.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RESULTS

Split-wrmScarlet

To engineer split-wrmScarlet, first we introduced a 32 amino acid spacer between the 10th and 11th β-strands of full-length wrmScarlet, following a strategy described previously (Feng *et al.* 2017). We subjected the spacer-inserted wrmScarlet sequence to several rounds of semi-random mutagenesis in *E. coli*, generating a split version of wrmScarlet with fluorescence comparable to the full-length wrmScarlet when expressed in bacteria. However, upon separating the two fragments into two *S. cerevisiae* plasmids to test for complementation, we observed no detectable fluorescence in yeast. We decided to continue with several rounds of selection of new mutant libraries in yeast using FACS, by fusing the wrmScarlet$_{11}$ sequence (without the MDELYK C-terminus residues) from our brightest *E. coli* clone to a plasma-membrane targeted blue FP (mTagBP2), and expressing soluble mScarlet$_{1-10}$ from a high-copy number vector containing a strong promoter. The brightest resulting protein, which we named split-wrmScarlet, contained 10 amino acid substitutions relative to the C-terminal truncated wrmScarlet (Figure S1, A and B). Fluorescent microscopy of yeast containing both plasmids corroborated that split-wrmScarlet showed the expected membrane localization and can reach brightness comparable to that of intact wrmScarlet in yeast (Figure S2, A and B).

Split-wrmScarlet is three-fold brighter than split-sfCherry3 in *C. elegans* muscles

In order to compare split-wrmScarlet to split-sfCherry3, the brightest published red split-FP at the time of the experiment, we combined the FP$_{1-10}$ and FP$_{11}$ fragments into a single plasmid for each fluorophore. Specifically, we generated plasmids encoding three nuclear localization signals (NLS), mTagBFP2, FP$_{11}$, a T2A peptide bond skipping sequence, mNeonGreen and the corresponding FP$_{1-10}$, driven by the ubiquitous somatic *Peft-3* promoter (Figure S3A). Each FP$_{11}$ was linked to mTagBFP2 in order to reduce the risk of proteolysis of the short peptide, and mNeonGreen was linked to FP$_{1-10}$ to monitor its expression, and for normalization purposes. Each construct was injected into wild-type animals and fluorescent progeny were analyzed. Unexpectedly, split-sfCherry3 turned out to be toxic when expressed ubiquitously, whereas 99% of split-wrmScarlet-overexpressing worms became viable adults (Figure S3B).
In an attempt to reduce split-sfCherry3-associated toxicity, we modified our construct by using the muscle-specific *myo-3* promoter and removing the NLS sequence (Figure 1B). We did not detect toxicity associated with the expression of these constructs and were able to compare the fluorescence of split-sfCherry3 and split-wrmScarlet in young adults. Red fluorescence emitted from split-wrmScarlet was 2.9-fold higher than that of split-sfCherry3 when normalized to the mNeonGreen signal (Figure 1, B and C).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

wrmScarlet$_{11}$-mediated tagging

Our protein-tagging approach was analogous to existing split-FP methods developed for human cells (Kamiyama *et al.* 2016, Leonetti *et al.* 2016) and *C. elegans* (Hefel and Smolikove 2019). It requires wrmScarlet$_{1-10}$ (i.e. wrmScarlet without the 11th β-strand) to be expressed in the cell or tissue of interest, and the short wrmScarlet$_{11}$ fragment to be inserted at an endogenous locus to tag a protein of interest (Figure 1A).
To build strains expressing single-copy insertions of wrmScarlet$_{1-10}$, we first optimized its sequence for *C. elegans* codon usage (Redemann *et al.* 2011) and included three introns (Table S1). The strain expressing wrmScarlet$_{1-10}$ in all somatic cells (driven by the *eft-3* promoter and *unc-54* 3'UTR) was generated by editing the genome of the existing MosSCI line CA1200 (Zhang *et al.* 2015) and replacing the sequence encoding *tir-1::mRuby* with *wrmScarlet$_{1-10}$* using CRISPR/Cas9 and hybrid DNA templates (Paix *et al.* 2015; Dokshin *et al.* 2018) (Table S1). In order to perform tissue-specific labeling, we generated a strain expressing muscle-specific wrmScarlet$_{1-10}$ using the SKI-LODGE system in the strain WBM1126 (Silva-García *et al.* 2019) (Table S1). The expression of wrmScarlet$_{1-10}$ in these two lines did not affect the number of viable progeny (Figure S4A) nor lifespan (Figure S4B), suggesting that the expression of wrmScarlet$_{1-10}$ has no deleterious effect. To tag a gene of interest with the wrmScarlet$_{11}$ fragment, we used microinjection of preassembled Cas9 RNPs because this method enables high-efficiency genome editing in worms (Paix *et al.* 2015). The most efficient insertion of short sequences in *C. elegans* was previously shown to be achieved using ssODN donors (Paix *et al.* 2015; Prior *et al.* 2017; Dokshin *et al.* 2018}. A great advantage of this strategy is that all of the components required for editing are commercially available or can be synthesized rapidly in the lab (Leonetti *et al.* 2016). Synthetic ssODNs have a typical size limit of 200 nt. The small size of wrmScarlet$_{11}$ (18 a.a.) is key: 200 nt can encompass wrmScarlet$_{11}$ (66 nt, including a 4 a.a. linker) flanked by two 67 nt homology arms for HDR. In principle, a few days after the somatic and/or muscle-specific wrmScarlet$_{1-10}$ strain(s) are microinjected, progeny can be screened for red fluorescence, genotyped and sequenced to check the accuracy of editing (Figure 2; a detailed protocol is available dx.doi.org/10.17504/protocols.io.bamkic4w). If desired, co-CRISPR strategies such as *dpy-10(cn64)* (Paix *et al.* 2015) or co-injection with pRF4 (Dokshin *et al.* 2018) can be used to help screening for correct candidates and to control for microinjection efficacy and payload toxicity.

To test our approach, we used it to tag six proteins with distinct subcellular localizations. Starting with the somatic wrmScarlet$_{1-10}$ parental strain CF4582, we introduced wrmScarlet$_{11}$ at the N-terminus of TBB-2, FIB-1 or VHA-13 or at the C-terminus of EAT-6, HIS-3 and TOMM-20 (Table S1). These proteins mark the cytoskeleton, nucleoli, lysosomes, plasma membrane, nuclei and mitochondria, respectively. Importantly, for transmembrane targets, the wrmScarlet$_{11}$ tag was

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

introduced at the terminus exposed to the cytosol. wrmScarlet fluorescence from all six proteins matched their expected subcellular localization in somatic cells (Figure 3, A-F). To test the muscle-specific wrmScarlet$_{1-10}$ line CF4610, we tagged the N-terminus of the endogenous FIB-1 with wrmScarlet$_{11}$ and confirmed the fluorescence from nucleoli in muscle cells (Figure 3, G and H). Together, our results show that split-wrmScarlet enables rapid fluorescent tagging of proteins with disparate cytoplasmic or nuclear locations expressed from their endogenous loci.

wrmScarlet$_{11}$ tandem repeats increase fluorescence

To benchmark the fluorescence intensity of split-wrmScarlet to its full-length counterpart, we first generated *wrmScarlet::vha-13* transgenic animals and compared their fluorescence to *wrmScarlet$_{11}$::vha-13* in worms expressing wrmScarlet$_{1-10}$ somatically (Figure 4, A and B). At the *vha-13* locus, split-wrmScarlet was about half as bright as a full-length fluorophore (48%), a ratio comparable to that of split-mNeonGreen2 and its full-length counterpart in human cells (Feng *et al.* 2017). Since visualizing endogenous proteins of low abundance can be challenging, it is key to address this limitation. Increasing the number of FP$_{11}$ domains tagged to an endogenous protein multiplies the number of the corresponding FP$_{1-10}$ recruited, increasing the overall fluorescent signal in human cells (Leonetti *et al.* 2016) and in *C. elegans* (He *et al.* 2019; Hefel and Smolikove 2019). To test whether split-wrmScarlet fluorescence would be enhanced by wrmScarlet$_{11}$ tandem repeats, we introduced two wrmScarlet$_{11}$ domains at the N-terminus of *vha-13* and three wrmScarlet$_{11}$ domains at the C-terminus of *his-3*, using CRISPR/Cas9 and dsDNA as donor template (Table S1), in animals expressing somatic wrmScarlet$_{1-10}$. Compared to animals carrying a single wrmScarlet$_{11}$ at the identical locus, carrying a tandem of wrmScarlet$_{11}$ increased overall fluorescence by 1.5-fold, while carrying three increased it by 2.3-fold (Figure 4, C and D). Therefore, increasing the number of wrmScarlet$_{11}$ repeats improves visualization of low-abundance proteins.

sfGFP$_{11}$-mediated tagging in somatic cells

Split-sfGFP has been used successfully in worms before (Noma *et al.* 2017; He *et al.* 2019; Hefel and Smolikove 2019). However, there was still a need for a strain that ubiquitously expressed sfGFP$_{1-10}$ in the soma from an integrated single-copy insertion in order to avoid heterogeneity of expression, and time-consuming manual maintenance. To build this strain, we first optimized the original sfGFP$_{1-10}$ sequence for *C. elegans* codon usage and included one intron (Cabantous *et al.* 2004; Redemann *et al.* 2011) (Table S1). We initially generated a strain expressing sfGFP$_{1-10}$ driven by the *let-858* promoter and *unc-54* 3'UTR using MosSCI (Table S1), but later replaced the *let-858* promoter with the *eft-3* promoter using CRISPR/Cas9 and hybrid DNA donor template because we observed that *Peft-3* resulted in significantly higher levels of gene

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

expression (Paix *et al.* 2015; Dokshin *et al.* 2018) (Table S1). To validate this strain, we inserted sfGFP$_{11}$ at the N-terminus of lysosomal VHA-13 or at the C-terminus of nuclear-localized HIS-3 (Figure 5, A and B). Both strains yielded relatively bright signals in accordance with their predicted subcellular localization.

Dual color protein labeling with split-wrmScarlet and split-sfGFP

Finally, to test the compatibility of split-wrmScarlet and split-sfGFP *in vivo,* we crossed the strains *Peft-3::sfGFP$_{1-10}$; his-3::sfGFP$_{11}$* (CF4592) and *Peft-3::wrmScarlet$_{1-10}$; wrmScarlet$_{11}$::fib-1* (CF4601). This cross resulted in the generation of the line *Peft-3::sfGFP$_{1-10}$, Peft-3::wrmScarlet$_{1-10}$* (CF4588) as well as the dually labeled strain *Peft-3::sfGFP$_{1-10}$, Peft-3::wrmScarlet$_{1-10}$; wrmScarlet$_{11}$::fib-1, his-3::sfGFP$_{11}$* (CF4602, Figure 5C). The fluorescent signals from both split-FPs appeared in their respective subcellular compartments, strongly suggesting the absence of interference between the two systems. We would like to note an additional advantage of the strain CF4588. The loci of *wrmScarlet$_{1-10}$* and *sfGFP$_{1-10}$* are genetically linked (only 0.96 cM apart), which facilitates outcrossing when needed. In addition, our *C. elegans* lines expressing *wrmScarlet$_{1-10}$* and *sfGFP$_{1-10}$* are viable homozygotes, so the strains do not require special maintenance.

Split-wrmScarlet$_{1-10}$ was not functional in the *C. elegans* germline or mammalian cells

In an attempt to generate a strain with tissue-specific expression of wrmScarlet$_{1-10}$ in the germline, we optimized the sequence of wrmScarlet$_{1-10}$ with three introns and engineered it to avoid piRNA recognition in order to prevent transgene silencing (Wu *et al.* 2018; Zhang *et al.* 2018) (Table S1). For reasons we do not understand, we were unable to detect wrmScarlet fluorescence in the germline wrmScarlet$_{1-10}$ strain we generated, when injecting a plasmid encoding mNeonGreen::linker::wrmScarlet$_{11}$ (Figure S5, A and B), despite detecting mNeonGreen fluorescence, potentially due to compromised expression, folding or maturation of the protein wrmScarlet$_{1-10}$. A similar negative result was obtained when attempting to express split-mScarlet in mammalian cells, in spite of efforts to rescue its fluorescence by screening an mScarlet$_{11}$ single/double mutant library in HEK293T cells (Figure S6, A and B, and supplementary text).

Discussion

Several considerations should be taken into account when using this method. First, as with all existing split-FP systems, detection of a given protein labeled with a FP$_{11}$ can only occur in a cellular compartment where the corresponding FP$_{1-10}$ is present. Proteins tagged with wrmScarlet$_{11}$ or sfGFP$_{11}$ generated in this work were either exposed to the cytosol or nucleoplasm (nuclei or nucleoli), where both

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

wrmScarlet$_{1-10}$ and/or sfGFP$_{1-10}$ were present. For proteins or epitopes located within the lumen of organelles, such as the endoplasmic reticulum or mitochondria, one might need to generate and validate *C. elegans* lines expressing wrmScarlet$_{1-10}$ or sfGFP$_{1-10}$ containing a mitochondrial localization sequence or ER signal peptide and retention signals, respectively. These approaches have been used successfully in mammalian cells with split-sfGFP when tagging ER-resident polypeptides (Kamiyama *et al.* 2016) and with split-sfCherry2 to detect protein present in the mitochondrial matrix (Ramadani-Muja *et al.* 2019).

Second, as for any other protein tag, it is important to select, when possible, a site that is unlikely to interfere with protein folding, function or localization (Snapp 2005; Nance and Frøkjær-Jensen 2019). For example, N-termini of membrane- and organelle-resident proteins often contain signal peptides or localization signals, and C-termini may contain sequences that regulate protein turnover (degrons). Interestingly, there are examples of proteins that become toxic when tagged with a full-length GFP, but tolerate labeling with a split protein. For example, SYP-4 was reported to be mostly functional when endogenously-tagged with sfGFP$_{11}$ in a strain expressing sfGFP$_{1-10}$ specifically in the germline, but not functional when labeled with full-length GFP (Hefel and Smolikove 2019). For proteins of interest present at low levels, we provided an alternative protocol to insert an additional two or three wrmScarlet$_{11}$ fragments, which increases the overall fluorescence substantially. However, the number of wrmScarlet$_{11}$ fragments could likely be increased further, to at least seven tandem repeats, based on approaches used successfully with split-sfGFP in human cells (Feng *et al.* 2017) and *C. elegans* (Noma *et al.* 2017; He *et al.* 2019; Hefel and Smolikove 2019).

Third, we would like to emphasize differences between our technique and the bimolecular fluorescence-complementation (BiFC) assay. When used together, the green and red split fluorescent proteins used here can provide information on co-localization, but unlike BiFC split proteins (Hu *et al.* 2002), they are not intended to assess protein-protein interactions directly. This is because BiFC split proteins require finely tuned weak affinities that do not disrupt the underlying interaction being studied. In our approach, only the wrmScarlet$_{11}$ fragment is attached to a protein of interest, the wrmScarlet$_{1-10}$ one is expressed in excess and untagged. In principle, it could be possible to use secondary assays that assess target protein interactions using split-FP$_{1-10}$/FP$_{11}$ proteins, like FRET or anisotropy, if one so desired and one can account for the possibility of incomplete complementation.

Lastly, we would like to note that in spite of its being three times brighter than the latest split-sfCherry3 in worms, our current split-wrmScarlet was not visible in the mammalian cell line we examined (Figure S6). Its ability to fluoresce is not restricted to worms, because it can reach wild-type levels of brightness in yeast. We do not know the basis for this discrepancy, nor why the protein is not visible in the *C. elegans* germline. It is possible that the concentration of the mScarlet$_{1-10}$ fragment in mammalian cells is

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

too low to drive complementation with $mScarlet_{11}$. This could potentially be overcome by further mutagenizing split-mScarlet and screening for fluorescence at low expression levels in mammalian cells.

In conclusion, we believe our system can substantially increase the speed, efficiency, and easiness of *in vivo* microscopy studies in *C. elegans*. We expect it to facilitate two-color and co-localization experiments and to find wide use in the worm community. We believe that these strains could facilitate novel or large-scale experiments, such as efforts to tag the entire genome of *C. elegans*.

Author contributions

M.I. developed the split-wrmScarlet in A.G.Y. laboratory. J.P. performed the cell sorting. J.G. performed *C. elegans* experiments in C.K. laboratory. L.S. conducted the mammalian cell experiments in M.D.L. laboratory. J.G. wrote the initial draft. All authors provided intellectual contributions to the collaboration.

Acknowledgements

We thank Katie Podshivalova, Rex Kerr, Calvin Jan and David Botstein for comments on the manuscript, Peichuan Zhang for experimental suggestions, and members of the Kenyon lab for fruitful discussions. We also thank Behnom Farboud from Barbara Meyer's laboratory for discussing CRISPR/Cas9 protocols and Liangyu Zhang from Abby Dernburg's laboratory for sharing sequences of the CA1200 strain. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This project was supported by Calico Life Sciences LLC.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Literature cited

Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature Publishing Group* **14,** 53–56 (2016).

Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77,** 71–94 (1974).

Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **23,** 102–107 (2004).

Dickinson, D. J. & Goldstein, B. CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering. *Genetics* **202,** 885–901 (2016).

Dokshin, G. A., Ghanta, K. S., Piscopo, K. M. & Mello, C. C. Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in Caenorhabditis elegans. *Genetics* **210,** 781–787 (2018).

El Mouridi, S. *et al.* Reliable CRISPR/Cas9 Genome Engineering in Caenorhabditis elegansUsing a Single Efficient sgRNA and an Easily Recognizable Phenotype. *G3* **7,** 1429–1437 (2017).

Evans, T. C., ed. Transformation and microinjection (April 6, 2006), WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/ wormbook.1.108.1, http://www.wormbook.org.

Farboud, B., Severson, A. F. & Meyer, B. J. Strategies for Efficient Genome Editing Using CRISPR-Cas9. *Genetics* 211, 431–457 (2019).

Feng, S. *et al.* Improved split fluorescent proteins for endogenous protein labeling. *Nature Communications* **8,** 370 (2017).

Feng, S. *et al.* Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. *Communications Biology* **2,** 344–12 (2019).

Frøkjær-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated transgenesis in C. elegans. *Nature Publishing Group* **9,** 117–118 (2012).

Frøkjær-Jensen, C. *et al.* Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet* **40,** 1375–1383 (2008).

Goudeau, J. *et al.* Fatty acid desaturation links germ cell loss to longevity through NHR-80/HNF4 in C. elegans. *PLoS Biol* **9,** e1000599 (2011).

He, S., Cuentas-Condori, A. & Miller, D. M. NATF (Native and Tissue-Specific Fluorescence): A Strategy for Bright, Tissue-Specific GFP Labeling of Native Proteins in Caenorhabditis elegans. *Genetics* **212,** 387–395 (2019).

Hefel, A. & Smolikove, S. Tissue-Specific Split sfGFP System for Streamlined Expression of GFP Tagged Proteins in the Caenorhabditis elegans Germline. *G3* **9,** 1933–1943 (2019).

Hu, C. D., Chinenov, Y. & Kerppola TK. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell* **9**, 789-98 (2002)

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ramadani-Muja, J. *et al.* Visualization of Sirtuin 4 Distribution between Mitochondria and the Nucleus, Based on Bimolecular Fluorescence Self-Complementation. *Cells* **8,** (2019).

Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nature Communications* **7,** 11046–9 (2016).

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10,** 845–858 (2015).

Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct Requirements for Somatic and Germline Expression of a Generally Expressed Caenorhabditis elegans Gene. *Genetics* **146,** 227–238 (1997).

Kimble, J., Hodgkin, J., Smith, T. & Smith, J. Suppression of an amber mutation by microinjection of suppressor tRNA in C. elegans. *Nature* **299,** 456–458 (1982).

Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E3501–8 (2016).

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

Nance, J. & Frøkjær-Jensen, C. The Caenorhabditis elegans Transgenic Toolbox. *Genetics* **212,** 959–990 (2019).

Noma, K., Goncharov, A., Ellisman, M. H. & Jin, Y. Microtubule-dependent ribosome localization in C. elegans neurons. *eLife* **6,** 218 (2017).

Paix, A. *et al.* Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in Caenorhabditis elegans. *Genetics* **198,** 1347–1356 (2014).

Paix, A. High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. 1–26 (2015). doi:10.1534/genetics.115.179382/-/DC1

Paix, A., Schmidt, H. & Seydoux, G. Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs. *Nucleic Acids Res* **44,** e128 (2016).

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger & T. C., Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24**(1), 79-88 (2005)

Prior, H., Jawad, A. K., MacConnachie, L. & Beg, A. A. Highly Efficient, Rapid and Co-CRISPR-Independent Genome Editing in Caenorhabditis elegans. *G3* **7,** 3693–3698 (2017).

Redemann, S. *et al.* Codon adaptation-based control of protein expression in C. elegans. *Nature Publishing Group* **8,** 250–252 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Richardson, C. D. *et al.* CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat Genet* **50,** 1132–1139 (2018).

Silva-García, C. G. *et al.* Single-Copy Knock-In Loci for Defined Gene Expression in Caenorhabditis elegans. *G3* **9,** 2195–2198 (2019).

Snapp, E. Design and use of fluorescent fusion proteins in cell biology. *Current Protocols in Cell Biology* **Chapter 21,** 21.4.1–21.4.13 (2005).

Vicencio, J., Martínez-Fernández, C., Serrat, X. & Cerón, J. Efficient Generation of Endogenous Fluorescent Reporters by Nested CRISPR in Caenorhabditis elegans. *Genetics* **211,** 1143–1154 (2019).

Wu, W.-S. et al. pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans. Nucleic Acids Res 46, W43–W48 (2018).

Zhang, L., Ward, J. D., Cheng, Z. & Dernburg, A. F. The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Development* **142,** 4374–4384 (2015).

Zhang, D. et al. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science 359, 587–592 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FIGURE LEGENDS

Figure 1. Engineering and evaluating split-wrmScarlet
(A) Principle of endogenous protein labeling with split-wrmScarlet. (B) Schematic of the plasmids encoding split-wrmScarlet and split-sfCherry3. Each plasmid consists of the large FP_{1-10} sequence fused to mNeonGreen, and the corresponding small FP_{11} sequence fused to the mTagBFP2. The T2A sequence ensures that mTagBFP2::FP_{11} and the corresponding mNeonGreen::FP_{1-10} are separated. The images are representative displays of the ratio of red to green fluorescence intensity from images acquired under identical conditions after background subtraction and masking with the same threshold. Scale bar, 50 µm. (C) Emission intensities from split-sfCherry3 and split-wrmScarlet normalized to mNeonGreen. Mean ± s.d. Circles are individuals (n=6 for each split protein). ****$P < 0.0001$.

Figure 2. wrmScarlet$_{11}$-mediated tagging.
Schematic representation of the split-wrmScarlet workflow to visualize endogenous proteins in muscles or somatic tissues. Some illustrations were created with BioRender.com.

Figure 3. Split-wrmScarlet labeling of proteins with distinct subcellular locations.
Endogenous proteins tagged with wrmScarlet$_{11}$ in animals expressing wrmScarlet$_{1-10}$ in somatic tissues or in muscles. (A-F) Confocal images of worms expressing somatic wrmScarlet$_{1-10}$ and (A) EAT-6::wrmScarlet$_{11}$ (membrane), (B) wrmScarlet$_{11}$::TBB-2 (cytoskeleton), (C) wrmScarlet$_{11}$::FIB-1 (nucleoli), (D) HIS-3::wrmScarlet$_{11}$ (nuclei), (E) wrmScarlet$_{11}$::VHA-13 (lysosomes), or (F) TOMM-20::wrmScarlet$_{11}$ (mitochondria). (G,H) Transgenic worms expressing wrmScarlet$_{1-10}$ in muscle and wrmScarlet$_{11}$::FIB-1 (A-G) Maximum intensity projections of 3D stacks shown. Scale bars, 50 µm.

Figure 4. wrmScarlet$_{11}$ tandem repeats increase fluorescence.
(A) Images of animals carrying either full-length wrmScarlet, wrmScarlet$_{11}$ or two tandem repeats of wrmScarlet$_{11}$ inserted at the endogenous VHA-13 N-terminus. (B) wrmScarlet emission intensities of animals carrying full-length wrmScarlet, wrmScarlet$_{11}$ or dual wrmScarlet$_{11}$ inserted at the VHA-13 N-terminus. Mean ± s.d. Circles are individuals. ***$P < 0.001$. (C) Images of animals carrying either a single wrmScarlet$_{11}$ or three tandem repeats of wrmScarlet$_{11}$ inserted at the HIS-3 C-terminus. (D) wrmScarlet emission intensities from animals carrying a single wrmScarlet$_{11}$ or three tandem repeats of wrmScarlet$_{11}$ knockin at the HIS-3 C-terminus. Mean ± s.d. Circles are individuals. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.005$. Images from each comparison were taken under identical instrument conditions using confocal microscopy and are

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

shown using identical brightness and contrast settings. Images shown are from a single confocal plane. Scale bars, 50 μm.

Figure 5. Split-sfGFP and split-wrmScarlet dual color protein labeling.
Images of animals stably expressing sfGFP$_{1-10}$ in somatic tissues (A) CF4592 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III; his-3(muIs255[his-3::sfGFP$_{11}$] V*) or (B) CF4589 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III; vha-13(muIs268[sfGFP$_{11}$::vha-13]) V*
). (C) Dual color protein labeling with split-wrmScarlet and split-sfGFP in somatic cells. Composite display of red and green channels of animals expressing wrmScarlet$_{1-10}$ and sfGFP$_{1-10}$ in somatic tissues, HIS-3::sfGFP$_{11}$ and wrmScarlet$_{11}$::FIB-1; CF4602 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)], muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III; fib-1(muIs254[wrmScarlet$_{11}$::fib-1]), his-3(muIs255[his-3::sfGFP$_{11}$]) V*). Maximum intensity projections of 3D stacks shown. Scale bars, 50 μm.

Figure S1. Split-wrmScarlet sequence comparison to mScarlet.
(A) Protein sequence alignment of the full-length mScarlet and split-wrmScarlet. The ten amino acid substitutions are highlighted in red, the sequence corresponding to wrmScarlet$_{1-10}$ in light gray and the sequence of wrmScarlet$_{11}$ in dark gray.
(B) Protein structure from split-wrmScarlet generated with Phyre2 and PyMOL. Mutations of split-wrmScarlet relative to mScarlet are highlighted with surface rendering, and the wrmScarlet$_{11}$ strand is colored in black.

Figure S2. **Split-wrmScarlet brightness in *S. cerevisiae.***
(A) Composite display of red and blue channels for membrane-localized mTagBFP-mScarlet (wild-type) fusion or wrmScarlet$_{1-10}$ plus membrane localized mTagBFP-wrmScarlet$_{11}$ in yeast. Images were acquired and are displayed under identical conditions. Note that the heterogeneity inherent to expression from plasmids is large, but split-wrmScarlet is capable of brightness levels similar to the parent protein. A schematic of the plasmids transformed is presented above each image.
(B) Histograms displaying the pixel-per-pixel ratio of red to blue fluorescence for background corrected, masked images. mTagBFP/full-length Scarlet ratios are displayed in blue, and split-wrmScarlet in orange. The inset displays the average red/blue ratio.

Figure S3. Developmental toxicity in worms expressing split-sfCherry3 in somatic nuclei.
(A) Schematic of the plasmids encoding split-wrmScarlet and split-sfCherry3 used for comparison. Each plasmid consists of a large FP$_{1-10}$ fused to mNeonGreen, and the

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

corresponding small FP_{11} fused to the blue fluorescent protein 2 (mTagBFP2), preceded with three SV40 nuclear localization sequences (NLS). The T2A sequence ensures the separation of NLS::mTagBFP2::FP_{11} and the corresponding mNeonGreen::FP_{1-10}.
(B) Quantification of mNeonGreen-positive animals into one of three classes, dead eggs, larvae or adults.

Figure S4. Brood size and lifespan of wrmScarlet$_{1-10}$ and sfGFP$_{1-10}$ lines.
Split-wrmScarlet$_{1-10}$ and split-sfGFP$_{1-10}$ lines produced wild-type numbers of progeny (A) and a wild-type lifespan (B). Genotypes: N2E (wild-type), CF4582 (*muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*), CF4587 (*muIs253[(Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*), CF4588 (*muIs253[Peft-3::sfGFP$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)], muIs252[Peft-3::wrmScarlet$_{1-10}$::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*), CF4610 (*muIs257[Pmyo-3::wrmScarlet1-10::unc-54 3'UTR] I*) and WBM1126 (*wbmIs61[myo-3p::3XFLAG::dpy-10 crRNA::unc-54 3'UTR] I*). Supplementary Material, Table S1 shows survival statistics for all lifespan experiments.

Figure S5. Tissue-specific split-wrmScarlet fluorescence in the germline is undetectable
(A) Schematic of the plasmid encoding *Psun-1::mNeonGreen::linker::wrmScarlet$_{11}$::tbb-2* 3'UTR (left), which was injected into the (MosSCI) strain PHX1797 carrying a single, integrated copy of *Psun-1::wrmScarlet$_{1-10}$::sun-1* 3'UTR (right).
(B) Images of animal expressing mNeonGreen::linker::wrmScarlet$_{11}$ and wrmScarlet$_{1-10}$ in the germline. Scale bar, 20 µm.

Figure S6. Screen for split-mScarlet fluorescence in mammalian cells.
(A) FACS histograms of human codon-optimized mScarlet$_{1-10}$ expressed as a C-terminal GFP fusion. GFP expression verifies successful expression of the fusion protein in HEK293T cells by lentiviral transduction.
(B) Schematic of the CRISPR-based knock-in design for screening single and double mutants of mScarlet$_{11}$. Left panel shows that neither our original mScarlet$_{11}$ sequence nor its mutant library enabled detectable complementation as detected by FACS. Right panel shows that the control experiment using the sfGFP$_{1-10}$/sfGFP$_{11}$ system displays high levels of knock-in and complementation in HEK293T cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Text

Supplementary Materials and Methods

Mammalian cell culture. HEK293T cells (ATCC # CRL-3216) were cultured in high-glucose DMEM supplemented with 10% FBS, 1 mM glutamine and 100 µg/mL penicillin/streptomycin (Gibco). An mScarlet$_{1-10}$ cDNA codon-optimized for mammalian expression was fused to the C-terminus of eGFP and cloned into a pCDH lentiviral expression vector (SFFV GFP-mScarlet$_{1-10}$). Lentivirus was prepared using standard protocols (Kamiyama *et al.* 2016) and used to infect HEK293T cells. A polyclonal population of GFP-mScarlet$_{1-10}$ positive cells was isolated by FACS (using GFP fluorescence) and served as parental cell line for further experiments. For CLTA-N CRISPR engineering, *S. pyogenes* Cas9/sgRNA ribonucleoprotein complexes were prepared as in (Leonetti *et al.* 2016), mixed with HDR donor templates and electroporated into of GFP-mScarlet$_{1-10}$ cells by nucleofection.

CLTA-N mScarlet11 donor library. A cDNA pool of degenerate mScarlet$_{11}$ sequences was generated by oligonucleotide synthesis (GeneScript) and homology arms for HDR-mediated insertion at CLTA N-terminus were appended by PCR (Supplementary Material – Table-S1 for full sequences). Library diversity was verified by Illumina MiSeq deep-sequencing.

Supplementary Results

Split mScarlet screening in mammalian cells

We tested the applicability of the wrmScarlet$_{1-10}$ system for mammalian cell engineering but were surprisingly unsuccessful at detecting fluorescence. We designed a human codon-optimized mScarlet$_{1-10}$ cDNA and expressed it as a C-terminal GFP fusion in HEK293T cells by lentiviral transduction. Expression of GFP verified the successful expression of the fusion protein (Figure S6A). However, subsequent expression of mScarlet$_{11}$ fragments did not give rise to detectable red fluorescence despite numerous attempts. We reasoned that the mScarlet$_{11}$ amino-acid sequence might be sub-optimal for complementation in human cells and synthesized a library of degenerate mScarlet$_{11}$ sequences covering any possible single and double amino-acid mutants. Using an established assay for CRISPR-based knock-in of sequences at the CLTA N-terminus (a highly expressed gene in HEK293T cells (Leonetti *et al.* 2016), neither our original wrmScarlet$_{11}$ sequence nor its mutant library enabled detectable complementation (Figure S6B, left panels). By contrast, a control experiment using the GFP$_{1-10}$/GFP$_{11}$ system showed a high level of knock-in and complementation in HEK293T (Figure S6B, right panels). It is possible that mScarlet$_{1-10}$ is expressed in a non-functional form in

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

human cells, or that its binding to $mScarlet_{11}$ is occluded by competing interactions (with cellular chaperones, for example). In addition, we did not attempt complementation on primary non-transformed cell lines, like WI-38 cells, whose different proteostasis network and chaperones could aid split mScarlet folding. At this point, more experiments will be required to fully test the portability of split wrmScarlet to mammalian systems.

Supplementary Materials – Literature cited

Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nature Communications* **7,** 11046–9 (2016).

Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E3501–8 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185249; this version posted July 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Engineering and evaluating split-wrmScarlet

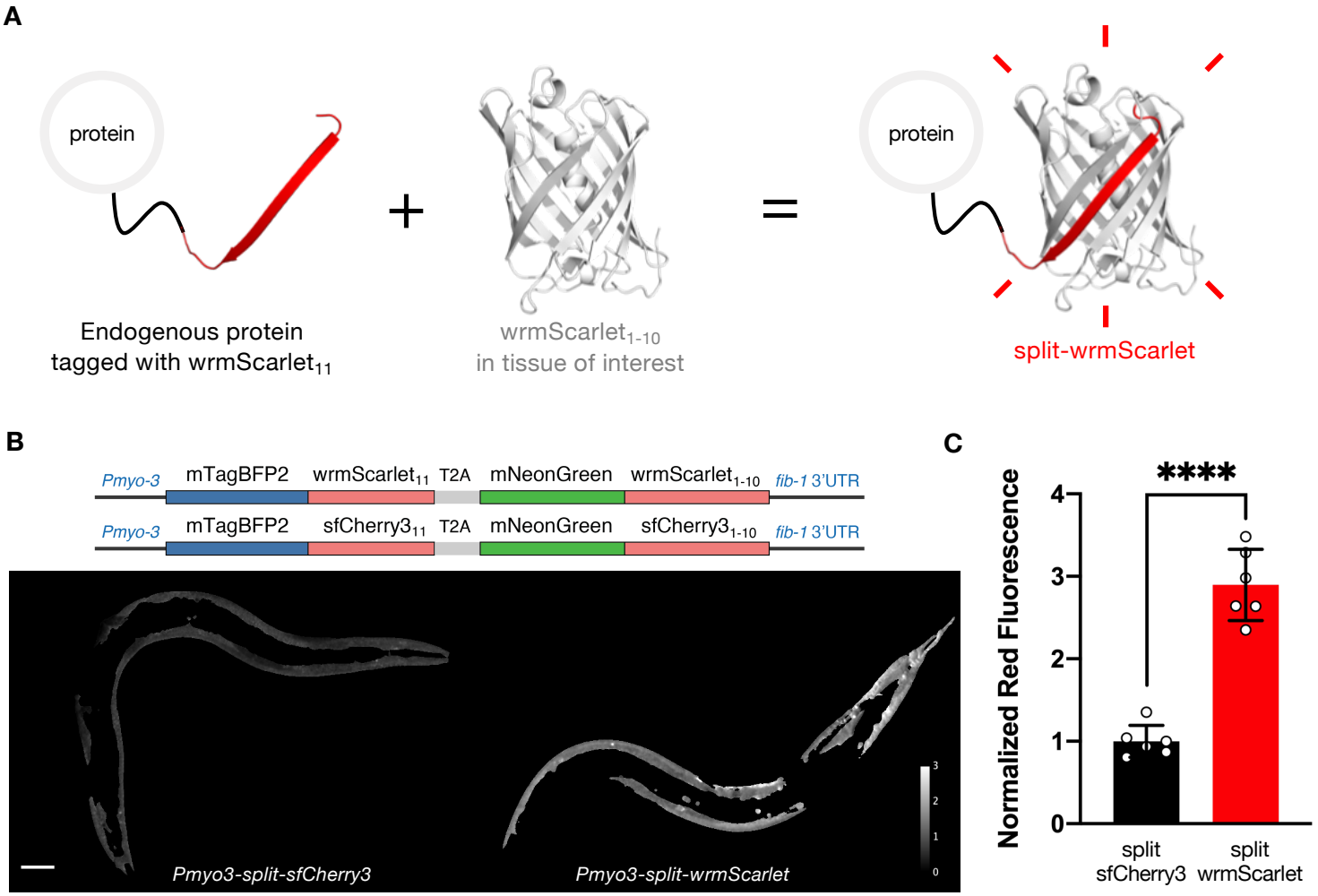
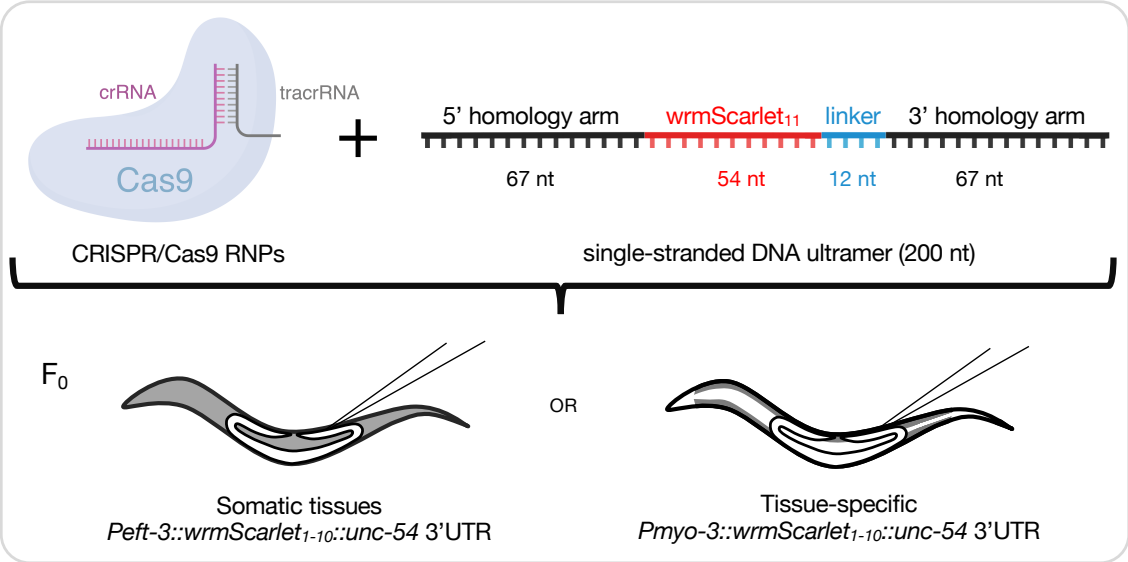
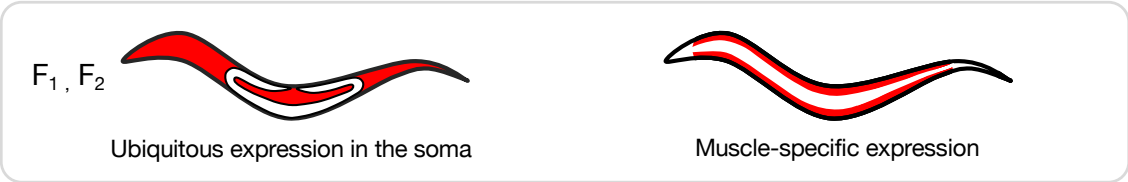


Figure 2. wrmScarlet$_{11}$-mediated tagging.

1. Microinjection into wrmScarlet$_{1\text{-}10}$ strain(s) of choice



2. Screen for fluorescent wrmScarlet$_{11}$-positive progeny



3. Lysis, PCR genotyping and sequencing

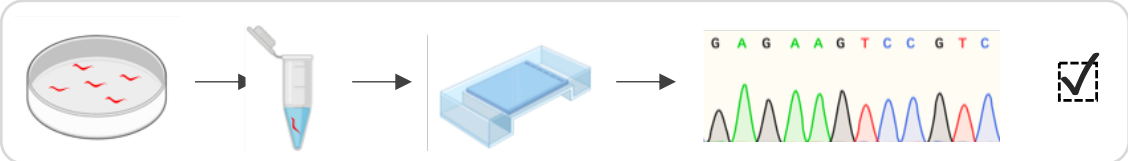


Figure 3. Split-wrmScarlet labeling of different structures

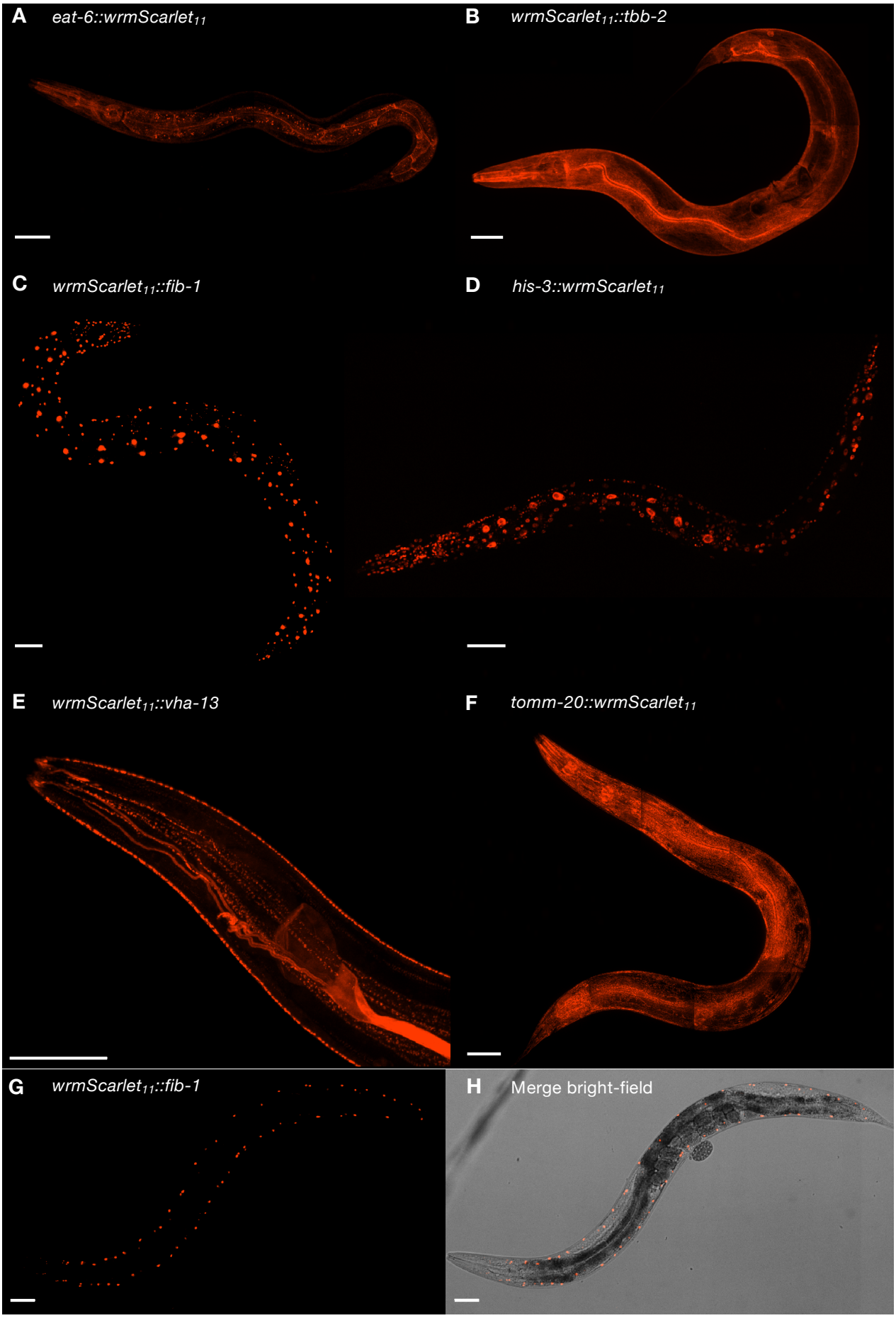
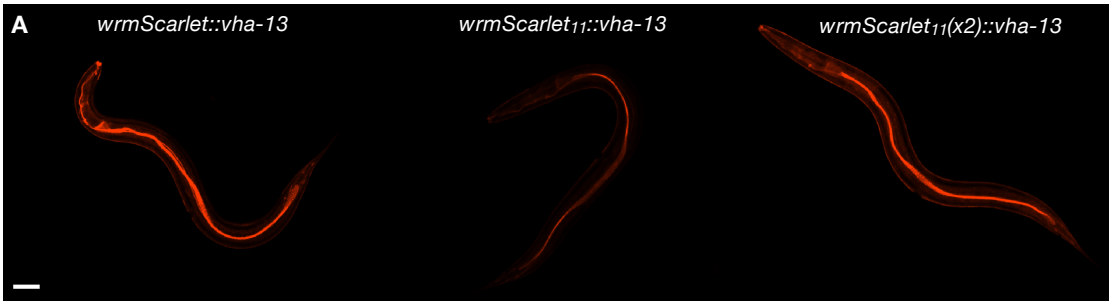
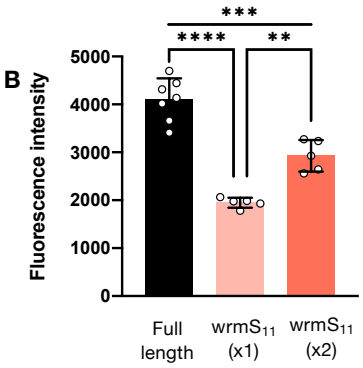
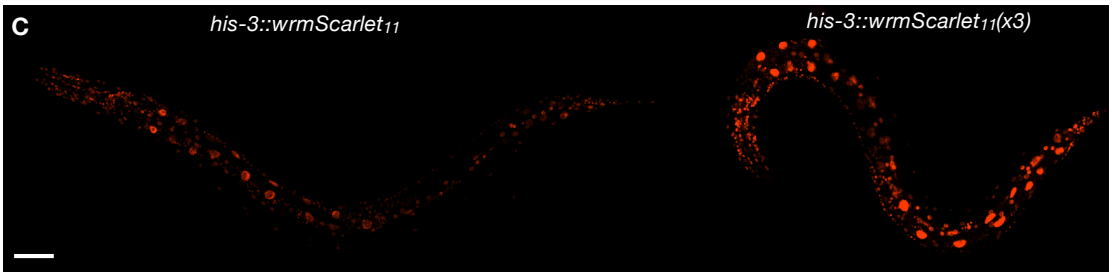


Figure 4. wrmScarlet$_{11}$ tandem repeats increase fluorescence







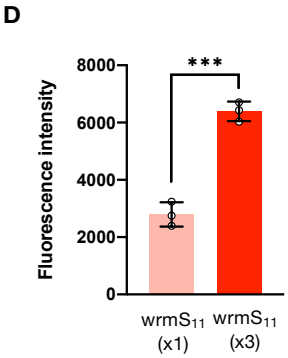


Figure 5. Split-sfGFP and split-wrmScarlet dual color protein labeling

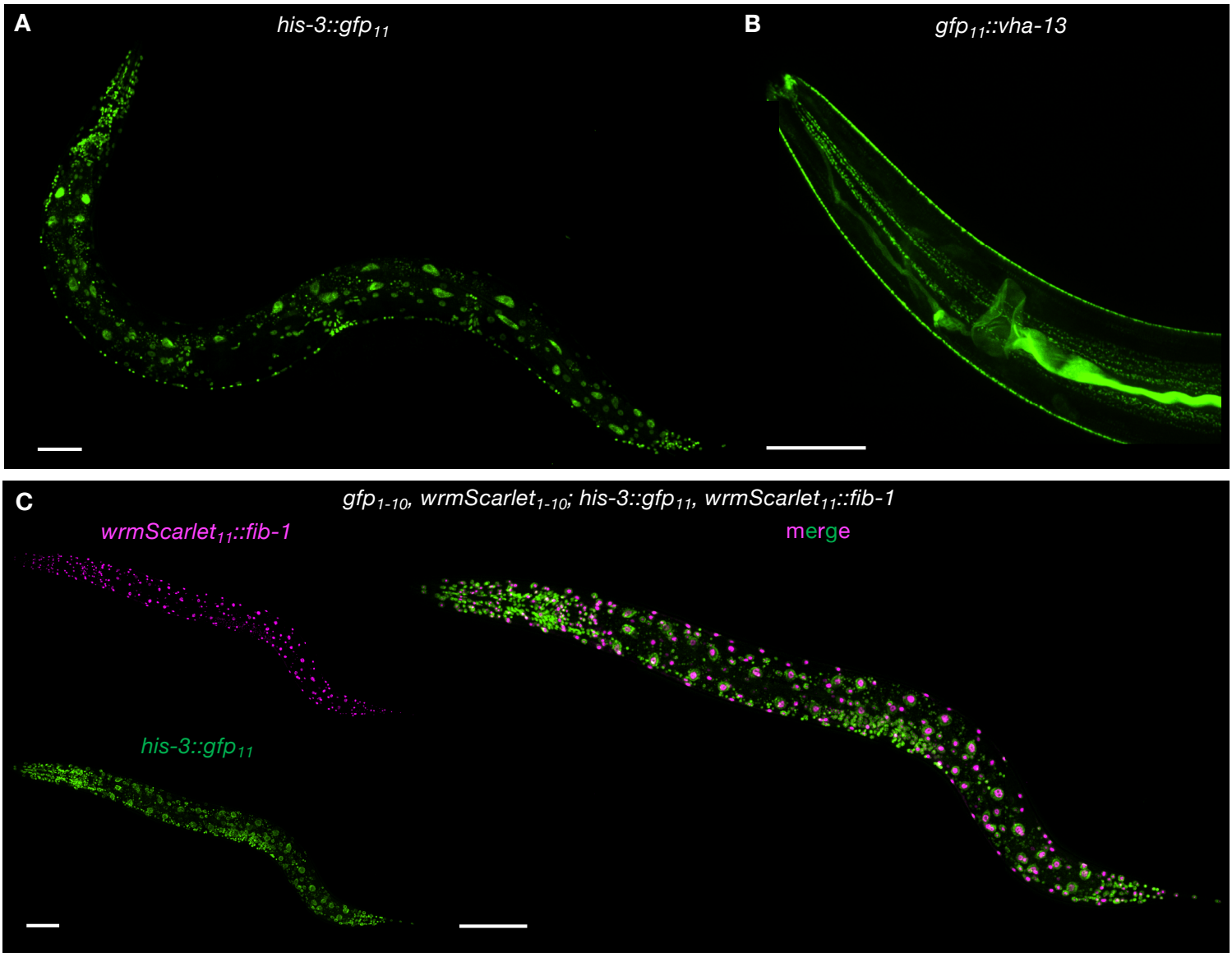


Table S1

Sequences and strains

Supplementary Material - Table S1	
Sheet 1	wrmScarlet1-10, wrmScarlet11, sfGFP1-10 and sfGFP11 DNA sequences
Sheet 2	*C. elegans* toolkit lines expressing single-copy of wrmScarlet1-10 and/or sfGFP1-10
Sheet 3	*C. elegans* strains, genotypes and sources
Sheet 4	crRNAs, HDR templates and oligonucleotide sequences
Sheet 5	Plasmids
Sheet 6	Adult lifespans of strains generated in this study

Sheet 1 - DNA Sequences of wrmScarlet1-10, wrmScarlet11, sfGFP1-10 and sfGFP11

Gene	DNA sequence
Codon-optimized wrmScarlet1-10 with 3 introns	ATGGTATCGAAGGGAGAAGCAGTAATCAAGGAGTTCATGCGTTTCAAGGTCCACATGGAGGGATCCATGAACGGACACGAGTTCGAGATCGAGGGGAGAGGGAGAGGGACGTCCATACGAGGGAACCCAAACCGCCAAGCTCAAGGTCACCAAGgtaagtttaaacatatatatactaactaaccctgatt atttaaattttcagGGAGGACCACTCCCATTCTCCTGGGACATCCTCTCCCCCACAATTCATGTACGGATCCCGTGCCTTCATCAAGCACCCAGCCGACATCCCAGACTACTACAAGCAATCCTTCCCAGAGGGATTCAAGTGGGAGCGTGTCATGAACTTCGGAGACGGGAGGCGTCTACGAACTTGA CCCAAAGACACCTCCCTCGAGGACGGAACCCTCATCTACAAGgtaagtttaaacagttcggtactaactaaccatacatatttaaattttcagGTCAAGCTCCGTGGAACCAACTTCCCACCAGACGGACCAGTCATGCAAAAGAAGACCATGGGATGGGAGGCCTCCACCGAGCGTCTCTACCCAGAGGACGGAGT CCTCAAGgtaagtttaaacatgatttttactaactaactaatctgatttaaattttcaggGAGACATCACCATGGCCCTCCGTCTCAAGGACGGAGGACGTTACCTCGCCGACACCTCCACCACCTACAAGGCCAAGAAGCCAGTCCAAATGCCAGGAGCCTACCTCGTCGACCGTAAGCTCGACATCACCTCCCACAA CGAGTAC
Codon-optimized wrmScarlet11	TACACCGTCGTCGAGCAATACGAGAAGTCCGTCGCCCGTCACTGCACCGGAGGA
Codon-optimized sfGFP1-10 with 1 intron	ATGTCTAAGGGAGAAGAGTTATTTACTGGAGTTGTGCCGATCCTCGTCGAGCTCGACGGAGACGTCAACGGACACAAGTTCTCCGTCCGTGGAGAGGGAGAGGGAGACGCCACCATCGGAAAGCTCACCCTCAAGTTCATCTGCACCACCGGAAAGCTCCCAGTCCCATGGCCAACCCTC GTCACCACCCTCACCTACGGAGTCCAATGCTTCTCCCGTTACCCAGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCAGAGGGATACGTCCAAGAGCGTACCATCTCCTTCAAGGACGACGGAAGTACAAGgtaagtttaaacatatatactaactaacccctgattattttaaattttcagACCCGT GCCGTCGTCAAGTTCGAGGGAGACACCCTCGTCAACCGTATCGAGCTCAAGGGAACCGACTTCAAGGAGGACGGAAACATCCTCGGACACAAGCTCGAGTACAACTTCAACTCCCACAACGTCTACATCACCGCCGACAAGCAAAAGAACGGAATCAAGGCCAACTTCACCGTCCGTCAC AACGTCGAGGACGGATCCGTCCAACTCGCCGACCACTACCAACAAAACACCCCAATCGGAGACGGACCAGTCCTCCTCCCAGACAACCACTACCTCTCCACCCAAACCGTCCTCTCCAAGGACCCAAACGAGAAG
Codon-optimized sfGFP11	CGTGACCACATGGTCCTTCATGAGTATGTAAATGCTGCTGGGATTACA

Sheet 2 - *C. elegans* lines expressing single-copy of wrmScarlet1-10 and/or sfGFP1-10

Strain	Genotype	Description	Genomic position	Genetic Position	Sequence
CF4582	*muls252[Peft-3::wrmScarlet1-10::unc-54 3'UTR, Cbr-unc-119(+)] II; unc-119(ed3) III*	Somatic wrmScarlet1-10 (eft-3 promoter)	II: 9.83 MB	II:1.73	gcacctttggctttttatgtcaactttcaattgctttctttcatttgttctgttaaattaatgaattttttcataaaataaagacattatacaataaaaatgaagaatttttatgaaaataaactgccagagagaaaaagtatgcaacactccgccgagagtgtttgaagtgtgtgacgtgtgt

Sheet 4 - crRNAs, HDR templates and oligonucleotide sequences

A. Sequences of crRNA and HDR template used for wrmScarlet11 and sfGFP11 knock-in experiments

Gene name	Gene ID	Tagged term	gene-specific crRNA sequence	1x wrmScarlet11 HDR donor sequence - Ultramer ssDNA (lower case: homology arms; red: wrmScarlet11; blue: linker)	Sequencing Primer (Forward)	Sequencing Primer (Reverse)
eat-6	B0365.3	C	ACAAGCUGUUCUUUAGUAGU	cgacgagatccgtcgtttctgattcgcagatatccagcggatgggcgcgtggacgctdacGGAGGAGGATCTACACCGTCGTCGAGCAATACGAGAAGTCGTTGCGCCGTCACTGCACCGGAGGAGGAGGATTCCGGAGGAGGAGATCCGGAGGAGGAGGATTCCGGAGGAGGAGGATTCCgagagatcgTcagtgaagcgggacaatgcgggaacaaatggatcaaat	CCTGGTTCATGTGCTATTGCC	CGACGACAGAAAGTAGCATCAC
fib-1	T01C3.7	N	AUAAUCGAUUUUUGUAGUAU	aatacgaaaaaatcctgaaattcagttatccctccgcccgccccacgattgaactcbgacgtccGGATCCTCCTCCTCCTCCGGTGCAGTGACGGGCGACGGACTTTCTCGTATTGCTCGACGACGGTGTAcatacacaaaaatcgattattttaacaaaagcgaaagcgaaattacgaaaaatacatccctcagcc	TTCTAGTCGATTCAGATCGACTGG	CGAAACTGCCACGATCACC
his-3	T10C6.12	C	GAAGAAAACCGGAGGAGACA	gatttttaatttgttttttttcgatttccatatagcttttcaaatcattcattccaggaaaagatgTACACCGTCGTCGAGCAATACGAGAAGTCGCGCCGTCACTGCACCGGAGGAGGGTGGCTCTGGAGGTCGCTCGTCGACGCGGTGTAGGATCCTCCTCCTCCT

Sheet 6 - Adult lifespans of strains in this study

Strain	Events / n initial	Mean lifespan +/- SEM (Days)	Median lifespan	% mean lifespan change vs. N2	*P*-value (log-rank) vs. N2
N2E	105 / 134	20.20 +/- 0.74	17		
CF4582	118 / 136	19.27 +/- 0.57	17	-4.60	0.19
CF4587	109 / 129	19.56 +/- 0.64	17	-3.17	0.29
WBM1126	112 / 127	19.31 +/- 0.61	17	-4.41	0.28
CF4610	108 / 128	20.18 +/- 0.69	17	-0.10	0.92

Strain	Events / n initial	Mean lifespan +/- SEM (Days)	Median lifespan	% mean lifespan change vs. WBM1126	*P*-value (log-rank) vs. WBM1126
WBM1126	112 / 127	19.31 +/- 0.61	17		
CF4610	108 / 128	20.18 +/- 0.69	17	4.51	0.33

Figure S1. Split-wrmScarlet sequence comparison to mScarlet

A

```
mScarlet        1 MVSKGEAVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWD  60
Split-wrmScarlet 1 MVSKGEAVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWD  60

mScarlet       61 ILSPQFMYGSRAFTKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLI 120
Split-wrmScarlet 61 ILSPQFMYGSRAFIKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLI 120

mScarlet      121 YKVKLRGTNFPPDGPVMQKKTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKT 180
Split-wrmScarlet 121 YKVKLRGTNFPPDGPVMQKKTMGWEASTERLYPEDGVLKGDITMALRLKDGGRYLADTST 180

mScarlet      181 TYKAKKPVQMPGAYNVDRKLDITSHNEDYTVVEQYERSEGRHSTGGMDELYK          232
Split-wrmScarlet 181 TYKAKKPVQMPGAYLVDRKLDITSHNEYYTVVEQYEKSVARHCTGG------ 226
                                                 split
```

B

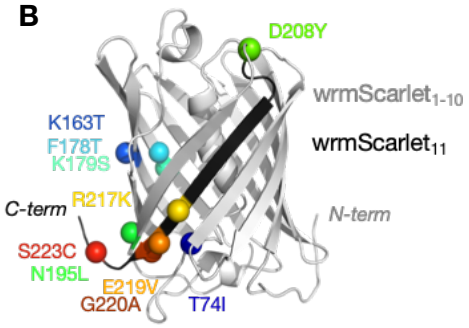


Figure S2. Split-wrmScarlet brightness in *S. cerevisiae*

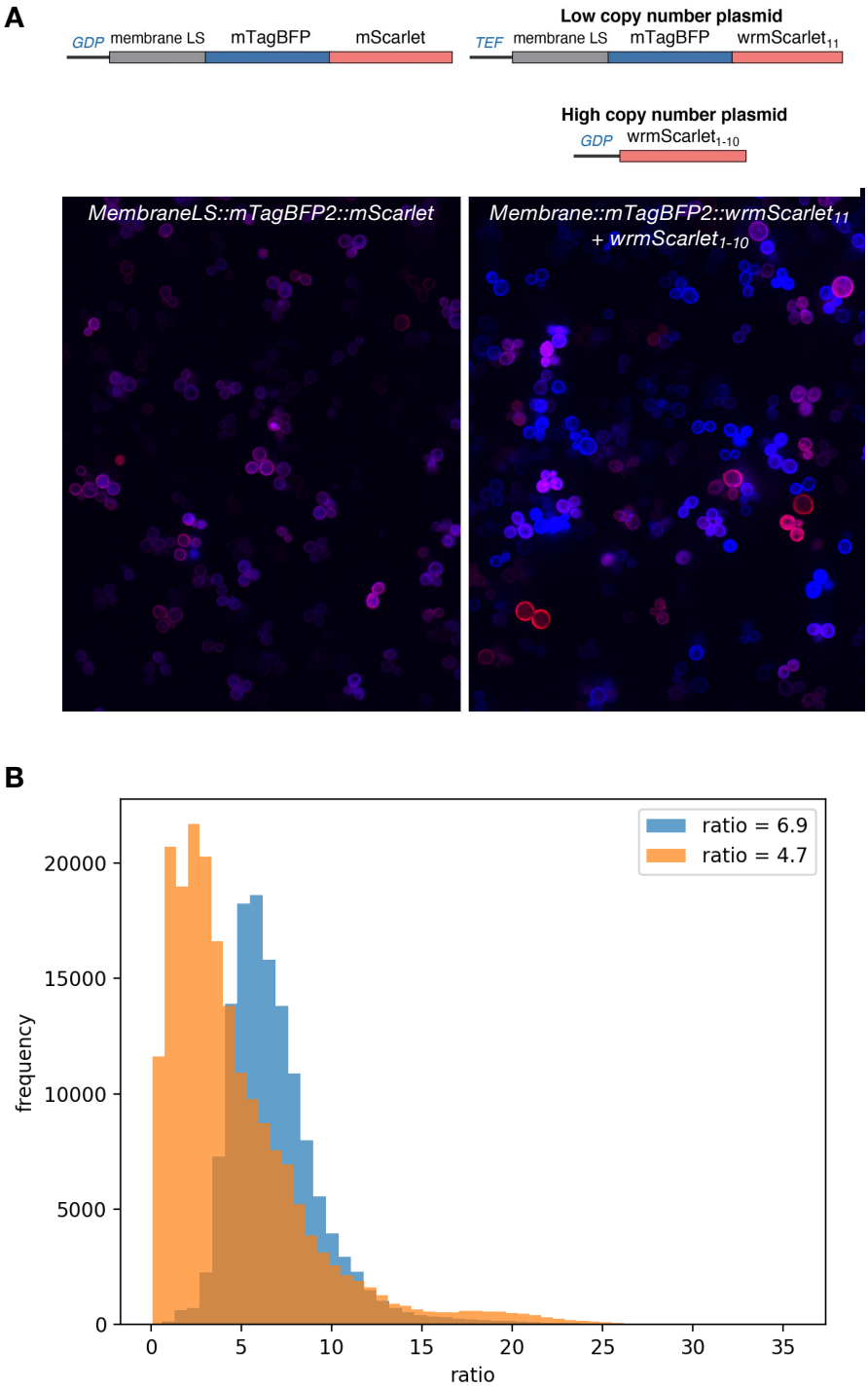


Figure S3. Developmental toxicity in worms expressing split sfCherry3 in somatic nuclei

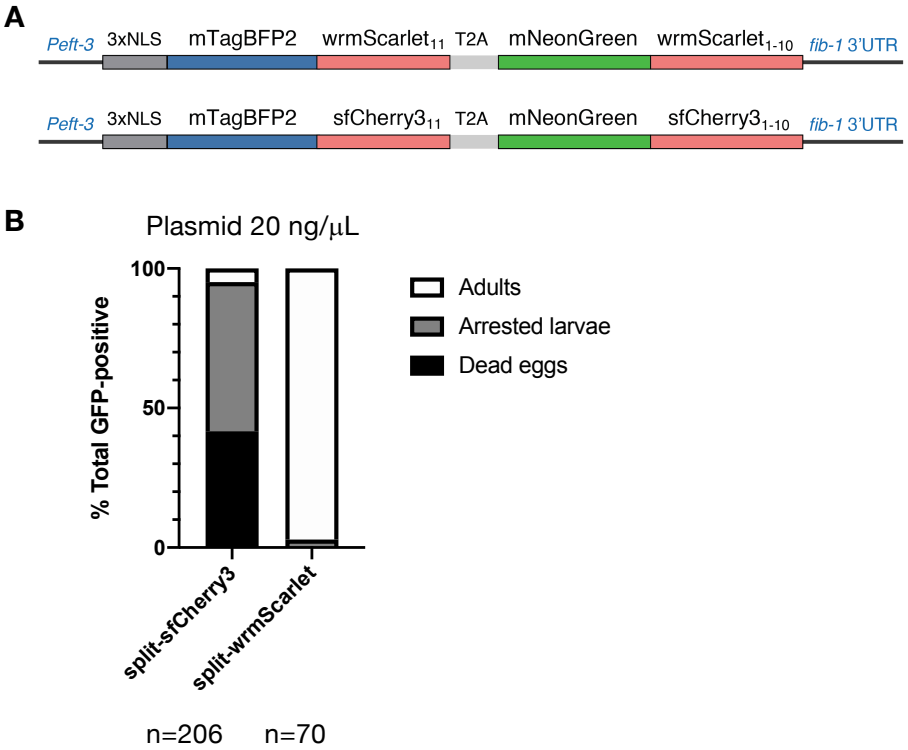
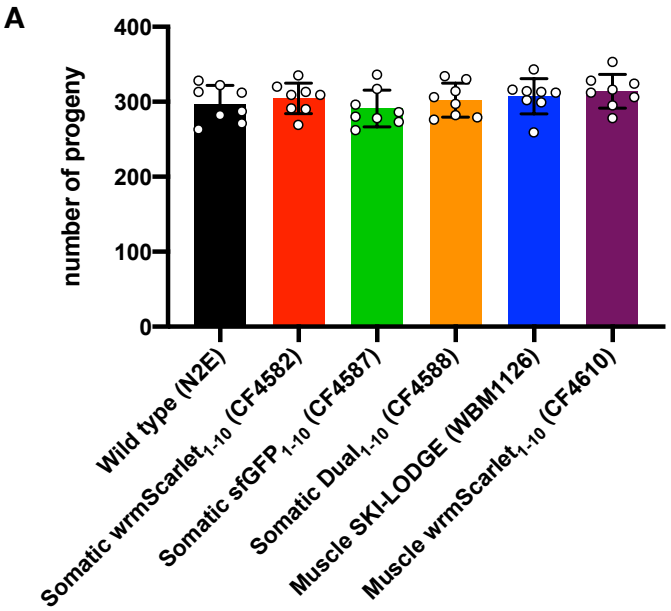
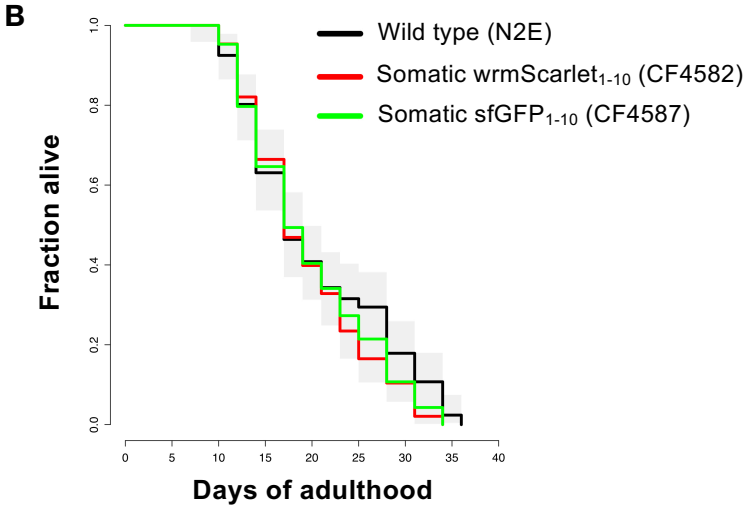


Figure S4 – Brood size and lifespan of wrmScarlet$_{1-10}$ and sfGFP$_{1-10}$ lines





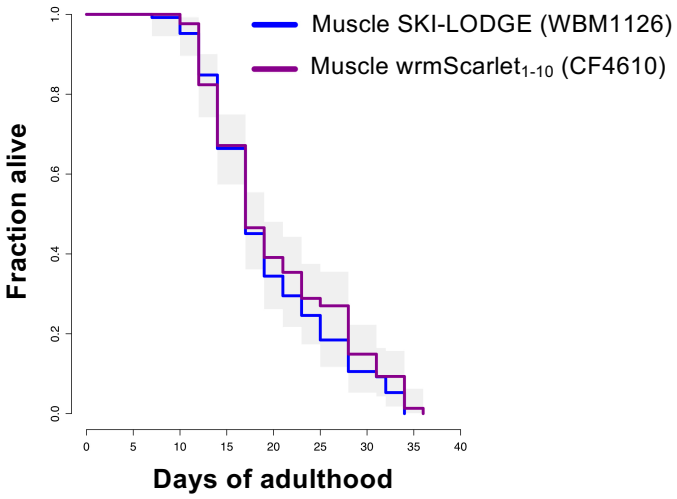


Figure S5 – Tissue-specific split-wrmScarlet fluorescence in the germline is undetectable

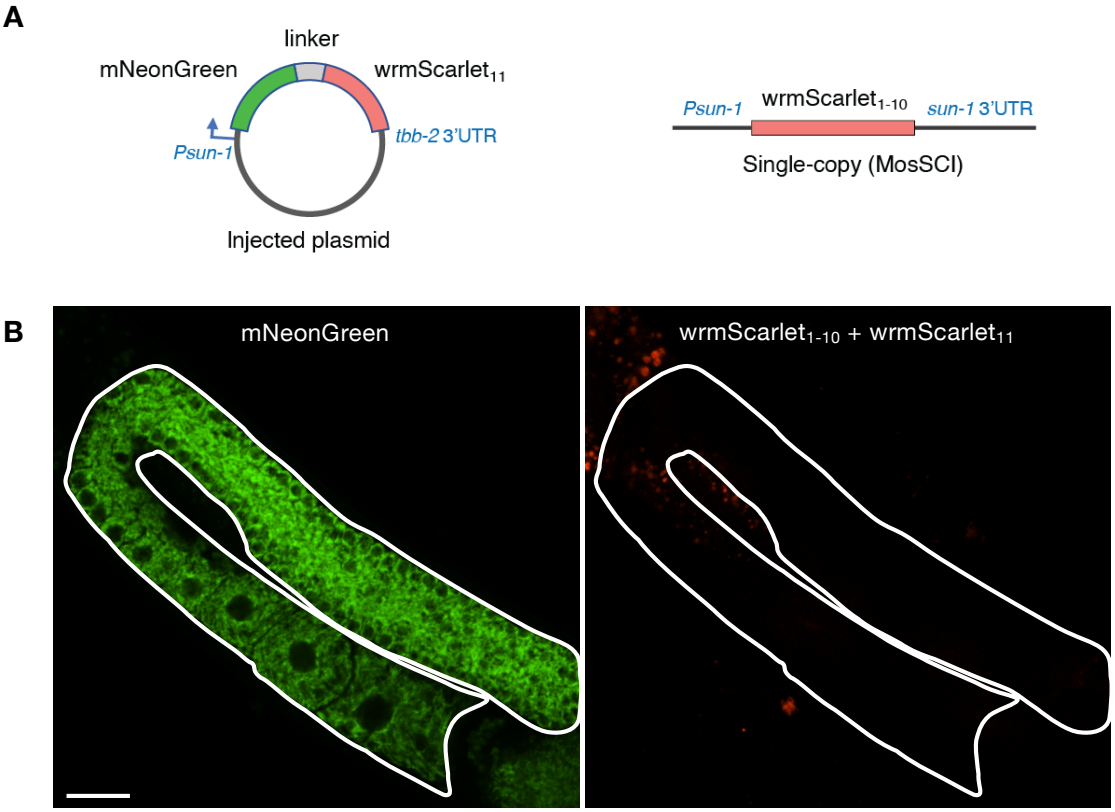
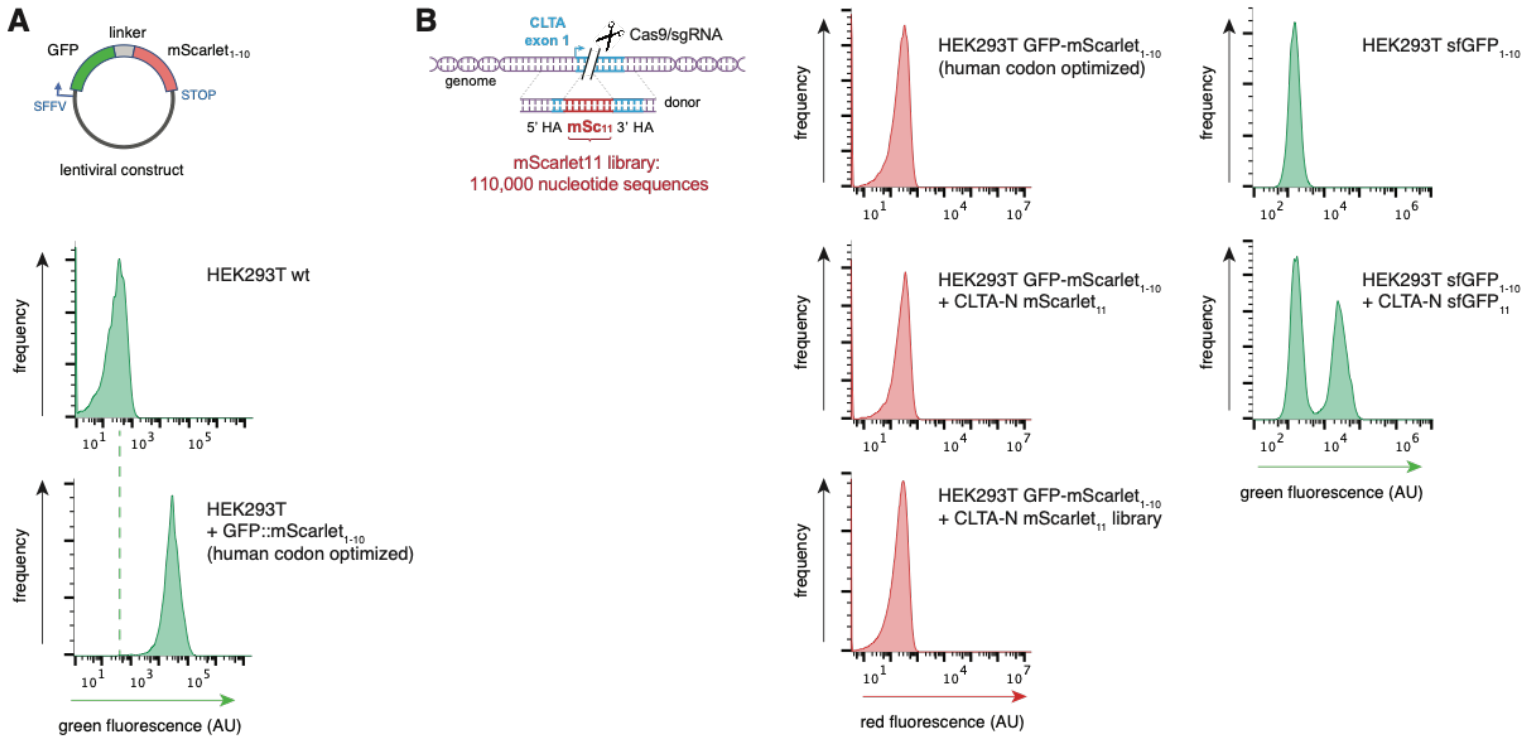



Figure S6. Screen for split-mScarlet fluorescence in mammalian cells.