Gene knock-ins in *Drosophila* using homology-independent insertion of universal donor plasmids

Justin A. Bosch[1], Ryan Colbeth[1], Jonathan Zirin[1], Norbert Perrimon[1,2]

[1] Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA

[2] Howard Hughes Medical Institute, Boston, MA, 02115, USA

Short Running Title: Homology-independent knock-ins in flies

Keywords: CRISPR-Cas9, knock-in, homology-independent, gene function, CRISPaint

100-word article summary: We report a new homology-independent genomic knock-in method in *Drosophila* to insert large DNA elements into any target gene. Using CRISPR-Cas9 and non-homologous end joining (NHEJ), an entire donor plasmid is inserted into the genome without the need for homology arms. This approach eliminates the burden associated with designing and constructing traditional donor plasmids. We demonstrate its usefulness in cultured cells and in vivo to fluorescently tag endogenous proteins, generate reporters of gene expression, and disrupt gene function.

Corresponding author:
Norbert Perrimon
Harvard Medical School
77 Avenue Louis Pasteur
Dept. of Genetics, NRB 336
Boston, MA 02115
617-432-7672

Email: perrimon@genetics.med.harvard.edu

Abstract

Site-specific insertion of DNA into endogenous genes (knock-in) is a powerful method to study gene function. However, traditional methods for knock-in require laborious cloning of long homology arms for homology-directed repair. Here, we report a simplified method in *Drosophila melanogaster* to insert large DNA elements into any gene using homology-independent repair. This method, known as CRISPaint, employs CRISPR-Cas9 and non-homologous end joining (NHEJ) to linearize and insert donor plasmid DNA into a target genomic cut site. The inclusion of commonly used elements such as GFP on donor plasmids makes them universal, abolishing the need to create gene-specific homology arms and greatly reducing user workload. Using this method, we show robust gene-specific integration of donor plasmids in cultured cells and the fly germ line. Furthermore, we use this method to analyze gene function by fluorescently tagging endogenous proteins, disrupting gene function, and generating reporters of gene expression. Finally, we assemble a collection of donor plasmids for germ line knock-in that contain commonly used insert sequences. This method simplifies the generation of site-specific large DNA insertions in *Drosophila* cell lines and fly strains, and better enables researchers to dissect gene function in vivo.

Introduction

Insertion of DNA into the animal genome is a powerful method to study gene function. This approach is multipurpose, and can be used to visualize protein localization (CRIVAT AND TARASKA 2012; HASSE *et al.* 2016; KANCA *et al.* 2017), to disrupt gene function (HOUSDEN *et al.* 2017), to assay gene expression (BRAND AND PERRIMON 1993; IVICS *et al.* 2009; BOUABE AND OKKENHAUG 2013), or to purify endogenous proteins (KIMPLE *et al.* 2013). Furthermore, the ability to insert large DNA elements such as promoters, protein coding sequences, or entire genes into the genome offers researchers endless options for genome modification. *Drosophila melanogaster* is an excellent animal model to analyze gene function because of its many genetic tools, fast generation time, and in vivo analysis (VENKEN *et al.* 2016; KORONA *et al.* 2017; BIER *et al.* 2018).

The two most commonly used methods in *Drosophila* to knock-in DNA into endogenous genes involve either transposable elements or homology directed-repair (HDR). Transposable DNA elements insert randomly in the genome by a Transposase enzyme (BELLEN *et al.* 2011), and cannot be used to target a user-specified gene. In contrast, HDR is used to insert DNA into a specific genomic location by cleavage at the genomic locus and precise homologous recombination of the DNA insert into the genome (BIER *et al.* 2018). Circular plasmids are commonly used as donor DNA for HDR because they can carry a large DNA insert (≤10kb) and homology arms corresponding to the target locus are added by traditional cloning techniques. While HDR is a useful method, the design and construction of unique plasmid donors for each gene is laborious. As a cloning-free alternative, synthesized single-stranded DNA (ssDNA) with short homology arms (~50-100 bp each) (BIER *et al.* 2018) or long

ssDNAs of ≤2kb (QUADROS *et al.* 2017) can be used as donors. However, ssDNA donors are limited to relatively small insertions such as epitope tags, and, like HDR plasmid donors, must be designed and produced for each gene that is targeted. Therefore, there is a need for easier, faster, and cheaper alternatives to knock-in large DNA elements into the *Drosophila* genome.

It was recently shown that large DNA elements could be knocked into a specific target locus without homology arms, known as homology-independent insertion (CRISTEA *et al.* 2013; MARESCA *et al.* 2013; AUER *et al.* 2014; KATIC *et al.* 2015; LACKNER *et al.* 2015; SCHMID-BURGK *et al.* 2016; SUZUKI *et al.* 2016; KATOH *et al.* 2017). In this method, simultaneous cutting of a circular donor plasmid and a genomic target-site by a nuclease such as Cas9 results in integration of the linearized donor plasmid into the genomic cut site by non-homologous end joining (NHEJ). This removes the need to construct homology arms, and only requires cloning or synthesizing a gene-specific single guide RNA (sgRNA). Furthermore, this approach is modular, since donor plasmids containing common insert sequences (e.g. GFP) can be targeted to different genomic locations and are thus "universal". Generating knock-ins by homology-independent insertion has been successfully applied in human cell lines (CRISTEA *et al.* 2013; MARESCA *et al.* 2013; LACKNER *et al.* 2015; SCHMID-BURGK *et al.* 2016; KATOH *et al.* 2017), mouse somatic cells (SUZUKI *et al.* 2016), zebrafish (AUER *et al.* 2014), and *C. elegans* (KATIC *et al.* 2015). However, this approach has not yet been applied in *Drosophila*.

Here, we show that homology-independent insertion functions effectively in *Drosophila* by using the CRISPaint method. We first characterize this method in cultured S2R+ cells, showing that a universal mNeonGreen donor plasmid can be used to fluorescently tag endogenous proteins at their C-terminus. We then demonstrate that

this approach works *in vivo*, using a universal *T2A-Gal4* donor plasmid in the fly germ line to obtain fly lines with insertions in a number of characterized genes. We show that these insertions can be used as expression reporters for the target gene and to generate loss-of-function phenotypes. Finally, we present a collection of different universal donor plasmids for the purpose of enabling the *Drosophila* research community to employ this method for their specific uses.

Materials and Methods

Plasmid cloning

*pCFD3-frame_selector_*(*0, 1, or 2*) plasmids (Addgene #s 127553-127555) were cloned by ligating annealed oligos encoding sgRNAs that target the CRISPaint target site (SCHMID-BURGK *et al.* 2016) into *pCFD3* (PORT *et al.* 2014), which contains the *Drosophila U6:3* promoter.

Additional sgRNA-encoding plasmids were generated by the TRiP (https://fgr.hms.harvard.edu/) or obtained from Filip Port (PORT *et al.* 2015). sgRNA plasmids targeting CDS close to the stop codon were *GP07595* (*Act5c*), *GP07596* (*His2Av*), *GP07609* (*alphaTub84B*), and *GP07612* (*Lam*). sgRNA plasmids targeting CDS close to the start codon were *GP06461* (*wg*), *GP02894* (*FK506-bp2*), *GP05054* (*alphaTub84B*), *GP00225* (*esg*), *GP00364* (*Myo1a*), *GP00400* (*btl*), *GP00583* (*Mhc*), *GP01881* (*hh*), *GP03252* (*Desat1*), *GP05302* (*ap*), *pFP545* (*ebony*), and *pFP573* (*ebony*). These sgRNAs were cloned into *pCFD3*, with the exception of those targeting *esg*, *Myo1a*, *btl*, and *Mhc*, which were cloned into *pl100* (KONDO AND UEDA 2013).

pCRISPaint-T2A-Gal4-3xP3-RFP (Addgene # 127556) was constructed using Gibson assembly (E2611, NEB) of three DNA fragments: 1) *Gal4-SV40-3xP3-RFP* was PCR amplified from *pHD-Gal4-DsRed* (Xu et al. 2019 submitted PNAS, (GRATZ *et al.* 2014); 2) linear plasmid backbone generated by digesting *pWalium10-roe* (PERKINS *et al.* 2015) with AscI/SacI; and 3) a synthesized double-stranded DNA fragment (gBlock, IDT) encoding the CRISPaint target site, linker sequence, T2A, and ends that overlap the other two fragments.

pCRISPaint-T2A-ORF-3xP3-RFP donor plasmids (Addgene #s 127557-127565) were cloned by PCR amplifying the ORFs and Gibson cloning into *CRISPaint-T2A-Gal4-3xP3-RFP* cut with NheI/KpnI. ORF sequences were amplified from templates as follows: *sfGFP* [amplified from *pUAS-TransTimer* (He et al. 2019 submitted)], *LexGAD* [amplified from p*CoinFLP-LexGAD/Gal4* (BOSCH *et al.* 2015)], *QF2* amplified from Addgene #80274, *Cas9-T2A-GFP* (amplified from template kindly provided by Raghuvir Viswanatha), *FLPo* (amplified from Addgene #24357), *Gal80* (amplified from Addgene #17748), *Nluc* (amplified from Addgene #62057), *Gal4DBD*, (amplified from Addgene #26233), and *p65* (amplified from Addgene #26234).

pCRISPaint-sfGFP-3xP3-RFP (Addgene # 127566) was cloned by PCR amplifying *sfGFP* coding sequence and Gibson cloning into *CRISPaint-T2A-Gal4-3xP3-RFP* cut with NotI/KpnI.

*pCRISPaint-CRIMIC_phase_*(*0,1, or 2*) (Addgene #s 127567-127569) donor plasmids were cloned by ligating annealed oligos containing the CRISPaint target site into *CRIMIC* [*pM37*, (LEE *et al.* 2018)] (frames 0,1,2) cut with NsiI.

*pCRISPaint-TGEM_phase_*(*0,1,or 2*) (Addgene #s 127570-127572) donor plasmids were cloned by ligating annealed oligos containing the CRISPaint target site into *T-GEM* (DIAO *et al.* 2015) (frames 0,1,2) cut with AgeI/NotI.

See Supplemental Table 6 for oligo and dsDNA sequences and Addgene for plasmid sequences.

Cell culture

Drosophila S2R+ cells stably expressing Cas9 and a mCherry protein trap in *Clic* (known as PT5/Cas9) (VISWANATHA *et al.* 2018) were cultured at 25˚C using Schneider's media (21720-024, ThermoFisher) with 10% FBS (A3912, Sigma) and 50 U/ml penicillin strep (15070-063, ThermoFisher). S2R+ cells were transfected using Effectene (301427, Qiagen) following the manufacturer's instructions. Plasmid mixes were composed of sgRNA-expressing plasmids (see above) and *pCRISPaint-mNeon-PuroR* (SCHMID-BURGK *et al.* 2016). Cells were transfected with plasmid mixes in 6-well dishes at 1.8x10^6 cells/ml, split at a dilution of 1:6 after 3-4 days, and incubated with 2 µg/ml Puromycin (540411 Calbiochem). Every 3-5 days, the media was replaced with fresh Puromycin until the cultures became confluent (~12-16 days). For single-cell cloning experiments, cultures were split 1:3 two days before sorting. Cells were resuspended in fresh media, triturated to break up cell clumps, and pipetted into a cell straining FACS tube (352235 Corning). Single cells expressing mNeonGreen were sorted into single wells of a 96 well plate containing 50% conditioned media 50% fresh media using an Aria-594 instrument at the Harvard Medical School Division of Immunology's Flow Cytometry Facility. Once colonies were visible by eye (3-4 weeks), they were expanded and screened for mNeonGreen fluorescence.

Fly genetics and embryo injections

Flies were maintained on standard fly food at 25˚C. Fly stocks were obtained from the Perrimon lab collection or Bloomington Stock center (indicated with BL#). Stocks used in this study are as follows: *yw* (Perrimon Lab), *yw/Y hs-hid* (BL8846), *yw; nos-Cas9attP40/CyO* (derived from BL78781), *yw;; nos-Cas9attP2* (derived from BL78782), *yw; Sp hs-hid/CyO* (derived from BL7757), *yw;; Dr hs-hid/TM3,Sb* (derived from BL7758), *UAS-2xGFP* (BL6874), *wg1-17/CyO* (BL2980), *wg1-8/CyO* (BL5351), *Df(2L)BSC291/CyO* (BL23676), *Mhc[k10423]/CyO* (BL10995), Df(2L)H20/CyO (BL3180), *Df(2L)ED8142/SM6a* (BL24135), *hh[AC]/TM3 Sb* (BL1749), *Df(3R)ED5296/TM3, Sb* (BL9338), *esgG66/CyO UAS-GFP* (BL67748), *Df(2R)Exel6069/CyO* (BL7551).

For embryo injections, each plasmid was column purified (Qiagen) twice, eluted in injection buffer (100 µM NaPO4, 5 mM KCl), and adjusted to 200ng/µl. Plasmids were mixed equally by volume, and mixes were injected into *Drosophila* embryos using standard procedures. For targeting genes on Chr. 2, plasmid mixes were injected into *yw;; nos-Cas9attP2* embryos. For targeting genes on Chr. 3, plasmid mixes were injected into *yw; nos-Cas9attP40/CyO* embryos. Approximately 500 embryos were injected for each targeted gene.

Injected G0 flies were crossed with *yw*. We used *yw/Y hs-hid* to facilitate collecting large numbers of virgin flies by incubating larvae and pupae at 37˚C for 1hr. G1 flies were screened for RFP expression in the adult eye on a Zeiss Stemi SVII fluorescence microscope. G1 RFP+ flies were crossed with the appropriate balancer stock (*yw; Sp hs-hid/CyO* or *yw;; Dr hs-hid/TM3,Sb)*. G2 RFP+ males that were *yellow-* (to remove the *nos-Cas9* transgene) and balancer+ were crossed to virgins of the appropriate balancer stock (*yw; Sp hs-hid/CyO* or *yw;; Dr hs-hid/TM3,Sb*). G3 larvae

and pupae were heat shocked at 37˚C for 1hr to eliminate the *hs-hid* chromosome, which generates a balanced stock (e.g. yw; [RFP+]/CyO).

Imaging

S2R+ cells expressing mNeonGreen were plated into wells of a glass-bottom 384 well plate (6007558, PerkinElmer). For fixed cell images, cells were incubated with 4% paraformaldehyde for 30min, washed with PBS with .1% TritonX-100 (PBT) 3x 5min each, stained with 1:1000 DAPI (D1306, ThermoFisher) and 1:1000 phalloidin-TRITC (P1951, Sigma), and washed with PBS. Plates were imaged on an IN Cell Analyzer 6000 (GE) using a 20x or 60x objective. Time-lapse videos of live mNeonGreen expressing single cell cloned lines were obtained by taking an image every minute using a 60x objective. Images were processed using Fiji software.

Wing imaginal discs from 3rd instar larvae were dissected in PBS, fixed in 4% paraformaldehyde, and permeabilized in PBT. For Wg staining, carcasses were blocked for 1hr in 5% normal goat serum (S-1000, Vector Labs) at room temp, and incubated with 1:50 mouse anti-wg (4D4, DSHB) primary antibody and 1:500 anti-mouse 488 (A-21202, Molecular Probes) secondary antibody. Primary and secondary antibody incubations were performed at 4˚C overnight. All carcasses were stained with DAPI and phalloidin-TRITC, and mounted on glass slides with vectashield (H-1000, Vector Laboratories Inc.) under a coverslip. Images of mounted wing discs were acquired on a Zeiss 780 confocal microscope.

Larvae, pupae, and adult flies were imaged using a Zeiss Axio Zoom V16 fluorescence microscope.

Quantification of mNeonGreen expressing S2R+ cells

For FACs-based cell counting, we collected cultures from each gene knock-in experiment before and after puromycin selection. Pre-selection cultures were obtained by collecting of 500ul of culture 3-4 days after transfection. Post-selection cultures were obtained after at least 2 weeks of puromycin incubation. Non-transfected cells were used as a negative control. 100,000 cells were counted for each sample and FlowJo software was used to analyze and graph the data. FSC-A vs GFP-A was plotted and we defined mNeonGreen+ cells by setting a signal intensity threshold where <0.02% of negative controls are counted due to autofluorescence.

For microscopy-based cell counting, the number of mNeonGreen cells was quantified by analyzing confocal images in Fiji using the manual Cell Counter Plugin (model). For transfected cells, 6 fields containing at least 200 cells were quantified (i.e. n=6). For puro-selected cells, 3 fields containing at least 200 cells were quantified (i.e. n=3).

Western blotting

Single cell-cloned cell lines were grown until confluent and 1ml of resuspended cells was centrifuged at 250g for 10min. The cell pellet was resuspended in 1ml ice cold PBS, re-centrifuged, and the pellet was lysed in 250ul 2x SDS-Sample buffer and boiled for 5min. 10ul was loaded on a 4-20% Mini-Protean TGX SDS-Page gel (4561096, BioRad), transferred to PVDF membrane (IPFL00010, Millipore), blocked in 5% non-fat dry milk, primary blotting using anti-mNeonGreen (1:1000, Chromtek 32F6) or hFAB™ Rhodamine Anti-Actin (12004164 BioRad), and secondary blotting using 1:3000 anti-

mouse HRP (NXA931, Amersham), imaging using ECL (34580, ThermoFisher) on a ChemiDoc MP Imaging System (BioRad).

PCR, sequencing, and sgRNA cutting assays

S2R+ cell genomic DNA was isolated using QuickExtract (QE09050, Lucigen). Fly genomic DNA was isolated by grinding a single fly in 50µl squishing buffer (10 mM Tris-Cl pH 8.2, 1 mM EDTA, 25 mM NaCl) with 200µg/ml Proteinase K (3115879001, Roche), incubating at 37˚C for 30 min, and 95˚C for 2 minutes. PCR was performed using Taq polymerase (TAKR001C, ClonTech) when running DNA fragments on a gel, and Phusion polymerase (M-0530, NEB) was used when DNA fragments were sequenced. DNA fragments corresponding to *mNeonGreen* or *T2A-Gal4* insertion sites were amplified using primer pairs where one primer binds to genomic sequence and the other primer binds to the insert. For amplifying non-knock-in sites, we used primers that flank the sgRNA target site. Primer pairs used for gel analysis and/or Sanger sequencing were designed to produce DNA fragments <1kb. Primer pairs used for next-generation sequencing of the insertion site were designed to produce DNA fragments 200-280bp. DNA fragments were run on a 1% agarose gel for imaging or purified on QIAquick columns (28115, Qiagen) for sequencing analysis. See Supplemental Table 6 for oligo sequences.

Sanger sequencing was performed at the DF/HCC DNA Resource Core facility and chromatograms were analyzed using Lasergene 13 software. Next-generation sequencing was performed at the MGH CCIB DNA Core. Fastq files were analyzed using CRISPresso2 (CLEMENT *et al.* 2019) by entering the PCR fragment sequence into the exon specification window and setting the window size to 10 bases. Quantification of

insertion types (seamless, in-frame in/del, and frameshift in/del) was taken from the allele plot and frame shift analysis outputs of CRISPresso2. The small proportion of "unmodified" reads that were not called by frameshift analysis were not included in the quantification.

T7 endonuclease assays (M0302L, NEB) were performed following the manufacturer instructions.

Data availability:

Donor plasmids and frame selector sgRNA plasmids will be deposited at Addgene. Fly strains, S2R+ cell lines, and sequence data are available on request. Oligo and dsDNA sequences are listed in Supplemental Table 6.

Results

To test homology-independent knock-in in *Drosophila*, we implemented a strategy known as CRISPaint (SCHMID-BURGK *et al.* 2016). This system is used to insert a protein tag or reporter gene into the coding sequence of an endogenous gene. Although it was originally designed for mammalian cell culture, CRISPaint has several advantages for use in *Drosophila*. First, this system uses CRISPR-Cas9 to induce double-strand breaks (DSBs), which is known to function efficiently in *Drosophila* cultured cells (BOTTCHER *et al.* 2014; VISWANATHA *et al.* 2018) and the germ line (KONDO AND UEDA 2013; REN *et al.* 2013; YU *et al.* 2013; BASSETT *et al.* 2014). Second, its use of a frame-selector gRNA target site makes insertion into the appropriate translation frame simple and modular (see below). Third, a collection of existing CRISPaint donor

plasmids (SCHMID-BURGK *et al.* 2016) containing common tags (e.g. GFP, RFP, Luciferase) are seemingly compatible for expression in *Drosophila*.

The CRISPaint system works by introducing three components into Cas9-expressing cells: 1) a single guide RNA (sgRNA) targeting a genomic locus; 2) a donor plasmid containing an insert sequence; and 3) a frame selector sgRNA targeting the donor plasmid (Figure 1A). This causes simultaneous cleavage of the genomic locus and donor plasmid, leading to the integration of linearized donor into the genomic cut site by non-homologous end joining (NHEJ). To ensure that the donor plasmid inserts in-frame with the endogenous gene, one of three frame selector sgRNAs are used. Importantly, these frame-selector sgRNAs do not target the *Drosophila* genome.

Homology-independent insertion functions efficiently in *Drosophila* S2R+ cells to produce endogenous protein tags

To test if the CRISPaint method functions in *Drosophila*, we set out to replicate the findings of (SCHMID-BURGK *et al.* 2016) in cultured S2R+ cells by genetically tagging endogenous proteins at their C-terminus. To accomplish this, we generated plasmids expressing frame-selector sgRNAs (frame 0,1, or 2) under the control of *Drosophila* U6 sequences (PORT *et al.* 2015) (Figure 1A). In addition, we generated plasmids expressing sgRNAs that target the 3' coding sequence of endogenous *Drosophila* genes. We chose to target *Actin5c*, *His2Av*, *alphaTub84B*, and *Lamin* because these genes are expressed in S2R+ cells (HU *et al.* 2017) and encode proteins with known subcellular localization (actin filaments, chromatin, microtubules, nuclear envelope, respectively). For donor plasmid, we used *pCRISPaint-mNeonGreen-T2A-PuroR* (SCHMID-BURGK *et al.* 2016), which contains a frame-selector sgRNA target site

upstream of coding sequence for the fluorescent mNeonGreen protein and Puromycin resistance protein (PuroR) linked by a cleavable T2A peptide sequence. Importantly, only integration of the donor plasmid in-frame with the target gene coding sequence will result in translation of mNeonGreen-T2A-PuroR (Figure 1A).

We transfected Cas9-expressing S2R+ cells (VISWANATHA *et al.* 2018) with a mix of three plasmids: *pCRISPaint-mNeonGreen-T2A-PuroR* donor, target-gene sgRNA, and the appropriate frame-selector sgRNA (Figure 1A, Supplemental Table 1). As an initial method to detect knock-in events, we used PCR to amplify the predicted insertion sites from transfected cells. Using primers that are specific to the target gene and *mNeonGreen* sequence, we successfully amplified *gene-mNeonGreen* DNA fragments for all four genes (Figure 1B). Furthermore, next-generation sequencing of these amplified fragments revealed that 34-50% of sense-orientation insertions are in frame with the target gene (Figure 1B, Supplemental Figure 1, Supplemental Table 1).

Next, we measured mNeonGreen fluorescence in transfected S2R+ cells as a more direct method of quantifying the frequency of in-frame knock-ins. Flow cytometry-based cell counting of transfected cells revealed that the number of mNeonGreen+ cells range from 0.19-2.4% (Figure 1C, Supplemental Table 1), in agreement with published results in human cultured cells (SCHMID-BURGK *et al.* 2016). These results were confirmed by confocal analysis of transfected cells, which showed mNeonGreen fluorescence in a small subset of cells (Figure 1C). Analysis of confocal images of *Act5c* and *His2Av* samples showed that 3.2% and 2.4% of transfected cells expressed mNeonGreen (Figure 1C, Supplemental Table 1), which roughly agreed with flow cytometry cell counting. Finally, mNeonGreen localized to the expected subcellular compartments, most obviously observed by His2Av-mNeonGreen and Lam-mNeonGreen co-localization with the nucleus, and Act5c-mNeonGreen and alphatub-

mNeonGreen exclusion from the nucleus (Figure 1C). These results suggest that a significant number of transfected S2R+ cells received in-frame insertion of mNeonGreen at their C-terminus using the CRISPaint homology-independent insertion method.

For knock-in cells to be useful in experiments, it is important to derive cultures where most cells, if not all, carry the insertion. Therefore, we enriched for in-frame insertion events using Puromycin selection (Figure 1D). After a two-week incubation of transfected S2R+ cells with Puromycin, flow-cytometry and confocal analysis revealed that most cells express mNeonGreen and exhibit correct subcellular localization (Figure 1E, Supplemental Table 1). For *alphaTub84B*, cell counting by flow-cytometry greatly underestimated the number of mNeonGreen+ cells counted by confocal analysis, likely because the mNeonGreen expression level was so low. Therefore, Puromycin selection is a fast and efficient method of selecting for mNeonGreen expressing knock-in cells after transfection.

A subset of cells in Puro-selected cultures had no mNeonGreen expression or unexpected localization (Figure 1E). Since each culture is composed of different cells with independent insertion events, we used FACs to derive single-cell cloned lines expressing mNeonGreen for further characterization (Figure 2A). At least 14 single-cell cloned lines were isolated for each target gene and imaged by confocal microscopy. Within a given clonal culture, every cell exhibited the same mNeonGreen localization (Figure 2B), confirming our single-cell cloning approach and demonstrating that the insertion is genetically stable over many cell divisions. Importantly, while many clones exhibited the predicted mNeonGreen localization, a subset of the clonal cell lines displayed an unusual localization pattern (Figure 2B). For example, three *Act5c-mNeonGreen* clones had localization in prominent rod structures, and 12 *Lamin-mNeonGreen* clones had asymmetric localization in the nuclear envelope (Figure 2B).

In addition, some clones had diffuse mNeonGreen localization in the cytoplasm and nucleus (Figure 2B).

To better characterize the insertions in single cell-cloned lines, we further analyzed three clones per gene (12 total), selecting different classes when possible (correct localization, unusual localization, and diffuse localization) (Supplemental Table 2). Using PCR amplification of the predicted insertion site (Figure 1A, Figure 2C) and sequencing of amplified fragments (Supplemental Table 2), we determined that all clones with correct or unusual mNeonGreen localization contained an in-frame insertion of mNeonGreen with the target gene. In contrast, we were unable to amplify DNA fragments from the expected insertion site in clones with diffuse mNeonGreen localization (Figure 2C). Western blotting of cell lysates confirmed that only clones with in-frame mNeonGreen insertion express fusion proteins that match the predicted molecular weights (Figure 2D). All together, these results suggest that clones with correct mNeonGreen localization are likely to contain an in-frame insert in the correct target gene.

Since S2R+ cells are polyploid (LEE *et al.* 2014), clones expressing mNeonGreen could bear one or more insertions. Furthermore, in/dels induced at the non-insertion locus could disrupt protein function. To explore these possibilities, we amplified the non-insertion locus in our single-cell cloned lines and used Sanger and next-generation sequencing to analyze the DNA fragments (Figure 2C, Supplemental Table 2). For each gene, we could find in/dels occurring at the non-insertion sgRNA cut site. For example, we could distinguish four distinct alleles in clone B11: a 3bp deletion, a 2bp deletion, a 1bp deletion, and a 27bp deletion. In addition, we identified an unusual mutation in clone C6, where a 1482bp DNA fragment inserted at the sgRNA cut site, which corresponds to a region from *alphatub84D*. We assume that this large insertion was

caused by homologous recombination, since *alphatub84D* and *alphatub84B* share 92% genomic sequence identity (Flybase). For *Act5c-mNeonGreen* clones A5 and A19, numerous in/del sequences were found, suggesting this region has an abnormal number of gene copies. We were unable to amplify a DNA fragment from *Lam-mNeonGreen* D9, despite follow-up PCRs using primers that bind genomic sequence further away from the insertion site (not shown).

One useful application of cell lines with fluorescently tagged endogenous proteins is to track their localization over time. Therefore, we used live confocal imaging of our single-cell cloned lines to capture mNeonGreen localization during cell division (Figure 2E, Supplemental Videos). Time-lapse images of dividing cells showed that Act5c-mNeonGreen localized to rod structures that asymmetrically or symmetrically distribute into daughter cells, His2Av-mNeonGreen localized to chromosomes that segregate into daughter cells, Lam-mNeonGreen showed disassembly and reassembly at the nuclear envelope, and alphaTub84B-mNeonGreen localized to mitotic spindles. These results demonstrate the usefulness of knock-in *Drosophila* cell lines to track the dynamic localization of endogenous proteins.

In vivo **germ line knock-in of T2A-Gal4 into endogenous genes using homology-independent insertion**

We next tested if homology-independent insertion could function in the *Drosophila* germ line for the purpose of generating knock-in fly strains. Compared to cultured cells, the isolation of flies bearing insertions that are in-frame with endogenous genes required additional considerations. As opposed to antibiotic selection, visible markers are commonly used to identify transgenic animals. In addition, since some

449 genes are expressed at low levels, target gene expression of an inserted reporter
450 element may be insufficient to identify in-frame insertion events.
451 To overcome these issues, we constructed the donor plasmid *pCRISPaint-T2A-*
452 *Gal4-3xP3-RFP* (Figure 3A). This donor contains a frame-selector sgRNA target site
453 upstream of the reporter gene *T2A-Gal4*, which encodes a form of the transcription
454 factor Gal4 that is cleaved from tagged endogenous protein (DIAO AND WHITE 2012).
455 Insertion of this element in-frame with genomic coding sequence would result in Gal4
456 translation, which can be detected using a *UAS-reporter* transgene (BRAND AND
457 PERRIMON 1993). In addition, this donor plasmid contains a *3xP3-RFP* selectable marker
458 gene that expresses bright red fluorescence in *Drosophila* larval tissues and the adult
459 eye (BERGHAMMER *et al.* 1999; GRATZ *et al.* 2014) (Figure 3B). Importantly, the
460 expression of *3xP3-RFP* is not dependent on in-frame insertion with the target gene.
461 Next, we tested for *pCRISPaint-T2A-Gal4-3xP3-RFP* insertion into 11
462 endogenous genes (Supplemental Table 3). These genes were selected based on their
463 known expression pattern, expression levels, or loss of function phenotype.
464 Furthermore, we targeted *pCRISPaint-T2A-Gal4-3xP3-RFP* to insert into the 5' portion
465 of the coding sequence (Figure 3A). This insertion location is designed to disrupt the
466 protein product by premature truncation. Plasmid mixes were injected into *nos-Cas9*
467 embryos, the resulting G0 progeny were outcrossed to *yw*, and G1 adults were
468 screened for RFP fluorescence (Figure 3C). Each RFP+ founder fly was outcrossed to
469 an appropriate balancer stock to establish a stable line. Figure 3D and Supplemental
470 Table 3 shows the integration efficiency results for each gene and Supplemental Table
471 4 has information on each balanced RFP+ line. From this data, we find that the
472 frequency of G0 crosses yielding RFP+ G1 progeny varies between 5% and 21%
473 (Figure 3D, Supplemental Table 3). For example, when targeting *ebony* with *pFP545*, 3

out of 16 G0 crosses produced ≥1 RFP+ G1 flies. Therefore, the *pCRISPaint-T2A-Gal4-3xP3-RFP* donor can insert into the genome of germ line cells in a homology-independent manner.

To gain insight into the genomic location of the insertions in our RFP+ lines, we first analyzed them by simple genetic crosses. During the fly stock balancing process, we determined that each insertion was located on the intended chromosome (Supplemental Table 4). In addition, flies that were homozygous for the insertion exhibited known phenotypes. For example, homozygous insertions in *ebony* produced flies with dark cuticle pigment (Figure 3E). Furthermore, flies with insertions targeting *wg*, *Mhc*, *hh*, and *esg* were homozygous lethal, which is consistent with known loss of function mutations in these genes (Supplemental Table 4). To test if the lethality of flies with homozygous insertions was due to on- or off-target gene disruption, we performed complementation tests by crossing RFP+ insertion lines with lines containing a known loss of function allele or genomic deletion spanning the gene. In all cases tested, trans-heterozygous combinations were lethal (Supplemental Table 4). Together, these results suggest that the *pCRISPaint-T2A-Gal4-3xP3-RFP* donor plasmid inserted into the intended target genes.

For T2A-Gal4 to be expressed by the target gene, the linearized donor plasmid must insert in the sense orientation relative to the target gene and in-frame with the coding sequence. As an initial screen for such events, we crossed RFP+ lines to a *UAS-GFP* line and assayed progeny for fluorescence. Through this approach, we identified Gal4-expressing lines for *ebony*, *myo1a*, *wg*, and *Mhc* (Figure 4A, Supplemental Tables 3, 4). *wg-T2A-Gal4* (#1 and 4), *Mhc-T2A-Gal4* (#1 and 2), and *Myo1a-T2A-Gal4* (#1) insertions express in the imaginal disc, larval muscle, and larval gut (Figure 4A), respectively, which matches the known expression patterns for these genes.

Furthermore, *wg-T2A-Gal4 #1 and #4* insertions were expressed in a distinctive Wg pattern in the wing disc pouch (Figure 4B). The expression pattern of *ebony* is less well understood. We find that *ebony-T2A-Gal4 pFP545 #2* is expressed in the larval brain (Figure 4A) and throughout the pupal body (Figure 4C), which is consistent with a previous study (HOVEMANN *et al.* 1998). However, *ebony-T2A-Gal4 pFP545 #2* is also expressed in the larval tracheal openings (Figure 4A), indicating that *ebony* may play a role in this tissue.

Next, we analyzed the insertion orientation and sequence structure in the RFP+ lines that express Gal4. We PCR amplified a region flanking the predicted insertion site from genomic DNA using primer pairs to distinguish sense and anti-sense insertions (Supplemental Figure 2). All RFP+ lines with Gal4 expression had insertions that were in the sense orientation (Supplemental Table 4, Supplemental Figure 2). Sequencing the resulting PCR fragments showed that the insert was present at the sgRNA cut site and each insertion contained an in/del between the target gene and *T2A-Gal4* sequence (Figure 4D). For example, *ebony-T2A-Gal4 pFP545 #2* contains a 15bp genomic deletion that is predicted to keep T2A-Gal4 in-frame with ebony. Similarly, *wg-T2A-Gal4* #1 contains an in-frame 45bp deletion and 21bp insertion. Remarkably, *wg-T2A-Gal4* #4 contains a frameshift in/del (Figure 4D), yet still expresses Gal4 in the Wg pattern, albeit at significantly lower levels than *wg-T2A-Gal4* #1 (Figure 4B). In similar cases, *Mhc-T2A-Gal4* lines #1, #2, and *Myo1a-T2A-Gal4* #1, each have in/dels that put *T2A-Gal4* out of frame with the target gene coding sequence. These findings confirm that our Gal4-expressing lines have *T2A-Gal4* inserted in the correct gene and orientation, but that in-frame insertion with the target gene is not necessarily a requirement.

To better characterize the insertion events in our collection of RFP+ lines, we analyzed those that did not produce fluorescence when crossed with *UAS-GFP* (Supplemental Table 4, Supplemental Figures 2,3). Using PCR and sequencing analysis, we found that some lines contained insertions in the correct target site but in the anti-sense orientation. In addition, we identified lines with insertions in the sense orientation, but were out of frame relative to the target gene. Unexpectedly, we found that *wg-T2A-Gal4* #6 contained a sense orientation in-frame insertion. Yet, unlike *wg-T2A-Gal4* #1, *wg-T2A-Gal4* #6 does not express Gal4. Importantly, our molecular analysis of every independently isolated RFP+ line (20 in total) revealed that each contained an insertion in the intended target site (Supplemental Table 4, Supplemental Figures 2,3).

We did not obtain RFP+ insertions when targeting *ap*, *alphaTub84B*, *btl*, or *Desat1*. Therefore, we investigated whether the sgRNAs targeting these genes were functional. All 4 sgRNAs used for germ line knock-ins have an acceptable efficiency score of >5, with the exception of the sgRNA targeting *btl* (Supplemental Table 5). We tested whether the sgRNAs were functional in transfected S2R+ cells by performing a T7 endonuclease assay that detects in/dels at the cut site. This test revealed that sgRNAs targeting *ap*, *alphaTub84B*, *btl* can cut at the target site, whereas the results with *desat1* were inconclusive (Supplemental Figure 4A). As an alternative functional test, we used PCR to detect knock-in events in S2R+ cells transfected with the *pCRISPaint-T2A-Gal4-3xP3-RFP* donor plasmid. This showed that sgRNAs targeting *ap*, *alphaTub84B*, *btl* and *desat1* can successfully knock-in *pCRISPaint-T2A-Gal4-3xP3-RFP* (Supplemental Figure 4B). Finally, we sequenced the sgRNA target sites in the *nos-Cas9* fly strains and found a SNP in the *btl* sgRNA binding site (not shown). The 10 remaining sgRNAs had no SNPs in the target site. In summary, we conclude that the

548 sgRNAs targeting *ap*, *alphaTub84B*, *btl*, and *Desat1* are able to induce cleavage at their
549 target site in S2R+ cells, but that the sgRNA targeting *btl* will not function in the germ
550 line using our *nos-Cas9* strains.
551
552 **A resource of CRISPaint donor plasmids for germ line knock-ins in Drosophila**
553
554 To facilitate the insertion of other sequences using the CRISPaint insertion
555 method, we generated 10 additional donor plasmids based on the same architecture as
556 *pCRISPaint-T2A-Gal4-3xP3-RFP* (Figure 5A). These include T2A-containing donors
557 with sequence encoding the alternative binary reporters LexGAD, QF2, and split-Gal4,
558 as well as Cas9 nuclease, FLP recombinase, Gal80 repression protein, NanoLuc
559 luminescence reporter, and super-folder GFP. Like *T2A-Gal4*, these can be used to
560 insert at 5' coding sequence, capturing endogenous gene expression and generating a
561 loss-of function. In addition, we generated *pCRISPaint-sfGFP-3xP3-RFP*, which can be
562 used to insert into 3' coding sequence, generating a C-terminal GFP fusion protein.
563 Several groups have demonstrated that coding sequence containing a splice
564 acceptor (SA) and inserted in a gene intron can produce a protein trap with the
565 preceding coding exon (MORIN *et al.* 2001; VENKEN *et al.* 2011). Recently, two studies
566 produced *SA-T2A-Gal4* donor plasmids for intron insertion by HDR, called CRIMIC and
567 T-GEM (DIAO *et al.* 2015; LEE *et al.* 2018). Therefore, we modified these two plasmids to
568 contain a CRISPaint target site upstream of the splice acceptor (Figure 5B).
569
570 **Discussion**
571

572 The insertion of large DNA elements into the genome by HDR requires a great
573 deal of expertise and labor for the design and construction of donor plasmids. Some
574 groups have developed strategies to improve the efficiency and scale at which
575 homology arms are cloned into donor plasmids (HOUSDEN *et al.* 2014; GRATZ *et al.*
576 2015), but the root problem still remains. Furthermore, each new gene-specific donor
577 plasmid requires the same amount of investment for their construction but is only used
578 once to achieve the desired knock-in. For these reasons, we believe that the current
579 methods for knock-in by HDR may act as a barrier to achieving widespread use by the
580 *Drosophila* community.

581 In this study, we addressed these challenges by demonstrating that large DNA
582 elements can insert into the *Drosophila* genome by a homology-independent
583 mechanism, using the previously established CRISPaint system. This approach has two
584 major advantages over HDR. 1) No construction of a donor plasmid is necessary, as
585 long as a suitable CRISPaint-compatible donor plasmid already exists. The only unique
586 reagent needed is an sgRNA that targets the endogenous gene (also required for HDR).
587 Cloning sgRNAs into expression plasmids, such as *pCFD3* (PORT *et al.* 2014), is simple,
588 fast, inexpensive, and works nearly every time. Furthermore, the availability of sgRNA-
589 encoding plasmids from public resources (e.g. TRiP, Addgene), and synthesized
590 sgRNA from commercial companies, means that researchers can increasingly order
591 their sgRNAs. 2) CRISPaint-compatible donor plasmids are “universal” and thus
592 modular. For example, different genes can be targeted by the same CRISPaint donor
593 plasmid, and different CRISPaint donor plasmids can be targeted to the same gene.
594 Publicly available collections of CRISPaint donor plasmids [(SCHMID-BURGK *et al.* 2016),
595 this study] ensure that researchers only need to select their insert of choice. Indeed, the
596 CRISPaint donor plasmids originally used for mammalian cell culture also function in

Drosophila S2R+ cells (Figure 1) and the *3xP3-RFP* marker in our germ line donor plasmids is compatible with other insects (BERGHAMMER *et al.* 1999).

An important step in obtaining correctly targeted knock-ins is molecularly validating the candidate insertions. Confirming an HDR insertion requires amplifying a large DNA fragment (~1.5kb-2kb) that encompasses part of the insert, an entire homology arm, and a portion of genomic sequence flanking the homology arm. This is necessary to verify that the donor did not insert off-target. These PCRs can sometimes fail or give inconclusive results due to the large fragment size. In contrast, CRISPaint knock-ins are easier to characterize by PCR analysis and sequencing because the amplified region is relatively small (~200-800bp) (Figure 1B, Figure 2C, Supplemental Figures 2,4). However, CRISPaint knock-ins require more work to screen since they can insert in two directions and in/dels occur at the insertion site. When possible, we recommend that researchers select for insert expression before molecular validation.

In this study, we generated knock-ins by inserting the entire linearized CRISPaint donor plasmid into the target gene. Since the backbone contains bacterial sequences, it may cause transgene silencing or impact neighboring gene expression (CHEN *et al.* 2004; SUZUKI *et al.* 2016). However, we note that thousands of transgenic fly lines contain bacterial sequences from phiC31 integration (PERKINS *et al.* 2015) with no reports of ill effects. Another issue is that insertion of the entire plasmid restricts the design of gene-tagging events to only append the insert 3' to the target insertion site. Different groups have used approaches that address these issues, such as providing donor plasmids as mini-circles (SCHMID-BURGK *et al.* 2016; SUZUKI *et al.* 2016), cutting donor plasmid twice to liberate the insert fragment (LACKNER *et al.* 2015; SUZUKI *et al.* 2016), or using PCR amplified inserts (Manna et al. 2019 BioRxiv). The first two modifications could in theory be made to our germ-line donor plasmids (e.g.

622 *pCRISPaint-T2A-Gal4-3xP3-RFP*), but for this study we opted to establish the simplest
623 protocol possible. Furthermore, we reasoned that cutting the donor twice would give rise
624 to two donor fragments and this could reduce knock-in efficiency.

625 Using the CRISPaint method in S2R+ cells, we readily identified cell lines with
626 endogenous proteins tagged with mNeonGreen at their C-terminus (Figure 2,
627 Supplemental Table 2). However, some lines exhibited unusual or unexpected protein
628 localization. In clones D6 and D9, Lamin-mNeonGreen localizes to the nuclear
629 envelope, but in D9 this localization is enriched asymmetrically in the direction of the
630 previous plane of cell division. Since these two clones contain the same seamless
631 mNeonGreen insertion, we speculate that mutations at non-knock-in loci account for this
632 difference. Indeed, clone D6 contained an in-frame 3bp deletion at the non-knock-in
633 locus, likely retaining wild-type function, whereas D9 had no remaining non-knock-in
634 locus. We saw a similar pattern for clones A3 and A5, where both had seamless
635 mNeonGreen insertions in *Actin5c*, but clone A5 exhibited distinct rod structures.
636 Finally, alphaTub84B-mNeonGreen fluorescence and protein levels were extremely low
637 in all cell lines, despite *alphaTub84B* being highly expressed in S2R+ cells (HU *et al.*
638 2017). We speculate that the alphaTub84B-mNeonGreen fusion protein is unstable and
639 previous studies in other organisms have highlighted problems with C-terminal tagging
640 of alpha-Tubulin (CARMINATI AND STEARNS 1997). Similarly, C-terminal tags can disrupt
641 Lamin and Actin function (DAVIES *et al.* 2009; NAGASAKI *et al.* 2017). These findings
642 illustrate the need for experimenters to consider the existing knowledge of the protein
643 when generating C-terminal protein fusions, and to carefully screen individual single cell
644 cloned lines.

We constructed a CRISPaint-compatible *T2A-Gal4* donor plasmid for use in the fly germ line and successfully identified insertion lines. Our knock-in efficiency, defined by the percentage of injected G0 flies that give RFP+ progeny, ranged from 5-21% (Figure 3B, Supplemental Table 3), which is roughly similar to knock-in efficiencies observed when using HDR [5-22% (GRATZ *et al.* 2014), 46-88% (PORT *et al.* 2015), 7-42% (GRATZ *et al.* 2015)]. In addition, all 20 of our RFP+ fly lines, which encompass 8 different sgRNA target sites, contain an insertion at the correct location. Though, we do not rule out the possibility of a second-site off-target event on the same chromosome.

To obtain *T2A-Gal4* insertions that express Gal4 under the control of the target gene (Figure 4), it was necessary to screen multiple independently derived insertions, due to the in/dels that occur at the insertion site and the two insertion orientations. However, we found for some genes the overall efficiencies were too low to obtain a successful Gal4-expressing line (*hh*, *esg, FK506-bp2*), or we did not obtain any RFP+ insertions (*ap*, *alphaTub84B*, *btl*, and *Desat1*). Additional steps could be taken to improve insertion efficiency, such as optimization of the injected plasmid concentrations, increasing the number of injected embryos, or simply reattempting with a different sgRNA. It is also possible that certain insertions are toxic to cells/animals during G0 germ-line development or in G1 progeny.

There were three unexpected findings with our germ-line insertions. First, some Gal4-expressing lines had *T2A-Gal4* inserted out of frame relative to the target gene. We speculate that it may be the result of ribosome frameshifting (KETTELER 2012), an internal ribosome entry site (IRES) (KOMAR AND HATZOGLOU 2005), or the presence of alternative open reading frames (altORFs) (MOUILLERON *et al.* 2016). However, we find no obvious evidence of these mechanisms by analyzing the sequence flanking *wg*, *Mhc*, and *Myo1a* insertion sites (not shown). Ultimately, we consider this a fortuitous effect as

long as Gal4 is expressed in the correct pattern. Second, we found an in-frame insertion in *wg* (#6) that that does not express Gal4. This finding highlights the importance of screening RFP+ insertions for Gal4 expression when possible. Third, we found that, unlike in cell culture, all of our RFP+ fly lines contain in/dels at the insertion site. Germ cells are known to differ in their NHEJ mechanisms compared to somatic cells (PRESTON *et al.* 2006; AHMED *et al.* 2015), but it is not clear why this would reduce the frequency of seamless insertions. Perhaps genetic or chemical manipulation of NHEJ regulators during embryo injection could address this issue in the future. This finding also suggests that the CRISPaint frame-selector approach may not be as useful in the fly germ line as it is in cell culture.

Our collection of donor plasmids (Figure 5) provides many options for inserting protein-coding sequence into target genes. However, other uses for homology-independent knock-in can be imagined, such as inserting enhancer sequences (e.g. *UAS*) upstream of endogenous genes to induce their overexpression (RORTH 1996), a reporter gene near non-coding regulatory sequences to capture the transcriptional expression pattern of neighboring genes (BRAND AND PERRIMON 1993), entire genes into intergenic sequence (SADELAIN *et al.* 2011), or sequences to be used for labeling DNA loci (ROBINETT *et al.* 1996). Furthermore, the donor plasmids described in this study could be used to simply knock out endogenous genes with a selectable marker. Indeed, all of our mNeonGreen-expressing single cell cloned lines contain mutations in the non-knock-in locus and our fly germ line insertions produced loss of function phenotypes. This approach could greatly increase the efficiency of selecting knock-out alleles, which are traditionally done by laborious PCR-based screening of frameshift in/dels. We also note that, similar to the *T2A-Gal4* reporters in vivo, cell lines could be targeted with translational reporters such as NanoLuciferase or GFP. Finally, since our collection of

CRISPaint donor plasmids contain enzyme restriction sites that flank the insert sequence, they are also useful as parental vectors for constructing traditional HDR donor plasmids.

In summary, our homology-independent knock-in approach enables researchers to focus more effort on screening for correct insertions in cells or flies than on designing and constructing donor plasmids. Furthermore, the techniques required for screening knock-ins are less specialized than those for constructing donor plasmids, making this trade off potentially attractive for labs with less molecular biology expertise or resources. Therefore, we hope that this method will put knock-in technology into the hands of more researchers due to its simplicity.

Acknowledgements

We thank Jonathan Schmid-Burgk for advice and the *CRISPaint-mNeonGreen* donor plasmid, Claire Hu and the TRiP for sgRNA design and construction, Ben Ewen-Campen for valuable comments on the manuscript, Stephanie Mohr, Oguz Kanca, and Hugo Bellen for helpful discussions, Raghuvir Viswanatha for the *Cas9-T2A-EGFP* template sequence, and Rich Binari and Cathryn Murphy for general assistance. J.A.B. was supported by the Damon Runyon Foundation. This work was supported by NIH grants R01GM084947, R01GM067761, R24OD019847. N.P. is an investigator of the Howard Hughes Medical Institute.

References:

Ahmed, E. A., H. Scherthan and D. G. de Rooij, 2015 DNA Double Strand Break Response and Limited Repair Capacity in Mouse Elongated Spermatids. Int J Mol Sci 16**:** 29923-29935.

Auer, T. O., K. Duroure, A. De Cian, J. P. Concordet and F. Del Bene, 2014 Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res 24**:** 142-153.

Bassett, A. R., C. Tibbit, C. P. Ponting and J. L. Liu, 2014 Mutagenesis and homologous recombination in Drosophila cell lines using CRISPR/Cas9. Biol Open 3**:** 42-49.

Bellen, H. J., R. W. Levis, Y. He, J. W. Carlson, M. Evans-Holm *et al.*, 2011 The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics 188**:** 731-743.

Berghammer, A. J., M. Klingler and E. A. Wimmer, 1999 A universal marker for transgenic insects. Nature 402**:** 370-371.

Bier, E., M. M. Harrison, K. M. O'Connor-Giles and J. Wildonger, 2018 Advances in Engineering the Fly Genome with the CRISPR-Cas System. Genetics 208**:** 1-18.

Bosch, J. A., N. H. Tran and I. K. Hariharan, 2015 CoinFLP: a system for efficient mosaic screening and for visualizing clonal boundaries in Drosophila. Development 142**:** 597-606.

Bottcher, R., M. Hollmann, K. Merk, V. Nitschko, C. Obermaier *et al.*, 2014 Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res 42**:** e89.

Bouabe, H., and K. Okkenhaug, 2013 Gene targeting in mice: a review. Methods Mol Biol 1064**:** 315-336.

Brand, A. H., and N. Perrimon, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118**:** 401-415.

Carminati, J. L., and T. Stearns, 1997 Microtubules orient the mitotic spindle in yeast through dynein-dependent interactions with the cell cortex. J Cell Biol 138**:** 629-641.

Chen, Z. Y., C. Y. He, L. Meuse and M. A. Kay, 2004 Silencing of episomal transgene expression by plasmid bacterial DNA elements in vivo. Gene Ther 11**:** 856-864.

Clement, K., H. Rees, M. C. Canver, J. M. Gehrke, R. Farouni *et al.*, 2019 CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol 37**:** 224-226.

Cristea, S., Y. Freyvert, Y. Santiago, M. C. Holmes, F. D. Urnov *et al.*, 2013 In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol Bioeng 110**:** 871-880.

Crivat, G., and J. W. Taraska, 2012 Imaging proteins inside cells with fluorescent tags. Trends Biotechnol 30**:** 8-16.

Davies, B. S., L. G. Fong, S. H. Yang, C. Coffinier and S. G. Young, 2009 The posttranslational processing of prelamin A and disease. Annu Rev Genomics Hum Genet 10**:** 153-174.

Diao, F., H. Ironfield, H. Luan, F. Diao, W. C. Shropshire *et al.*, 2015 Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. Cell Rep 10**:** 1410-1421.

Diao, F., and B. H. White, 2012 A novel approach for directing transgene expression in Drosophila: T2A-Gal4 in-frame fusion. Genetics 190**:** 1139-1144.

Gratz, S. J., C. D. Rubinstein, M. M. Harrison, J. Wildonger and K. M. O'Connor-Giles, 2015 CRISPR-Cas9 Genome Editing in Drosophila. Curr Protoc Mol Biol 111**:** 31 32 31-20.

Gratz, S. J., F. P. Ukken, C. D. Rubinstein, G. Thiede, L. K. Donohue *et al.*, 2014 Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196**:** 961-971.

Hasse, S., A. A. Hyman and M. Sarov, 2016 TransgeneOmics--A transgenic platform for protein localization based function exploration. Methods 96**:** 69-74.

Housden, B. E., S. Lin and N. Perrimon, 2014 Cas9-based genome editing in Drosophila. Methods Enzymol 546**:** 415-439.

Housden, B. E., M. Muhar, M. Gemberling, C. A. Gersbach, D. Y. Stainier *et al.*, 2017 Loss-of-function genetic tools for animal models: cross-species and cross-platform differences. Nat Rev Genet 18**:** 24-40.

Hovemann, B. T., R. P. Ryseck, U. Walldorf, K. F. Stortkuhl, I. D. Dietzel *et al.*, 1998 The Drosophila ebony gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression. Gene 221**:** 1-9.

Hu, Y., A. Comjean, N. Perrimon and S. E. Mohr, 2017 The Drosophila Gene Expression Tool (DGET) for expression analyses. BMC Bioinformatics 18**:** 98.

Ivics, Z., M. A. Li, L. Mates, J. D. Boeke, A. Nagy *et al.*, 2009 Transposon-mediated genome manipulation in vertebrates. Nat Methods 6**:** 415-422.

Kanca, O., H. J. Bellen and F. Schnorrer, 2017 Gene Tagging Strategies To Assess Protein Expression, Localization, and Function in Drosophila. Genetics 207**:** 389-412.

Katic, I., L. Xu and R. Ciosk, 2015 CRISPR/Cas9 Genome Editing in Caenorhabditis elegans: Evaluation of Templates for Homology-Mediated Repair and Knock-Ins by Homology-Independent DNA Repair. G3 (Bethesda) 5**:** 1649-1656.

Katoh, Y., S. Michisaka, S. Nozaki, T. Funabashi, T. Hirano *et al.*, 2017 Practical method for targeted disruption of cilia-related genes by using CRISPR/Cas9-mediated, homology-independent knock-in system. Mol Biol Cell 28**:** 898-906.

Ketteler, R., 2012 On programmed ribosomal frameshifting: the alternative proteomes. Front Genet 3**:** 242.

Kimple, M. E., A. L. Brill and R. L. Pasker, 2013 Overview of affinity tags for protein purification. Curr Protoc Protein Sci 73**:** Unit 9 9.

Komar, A. A., and M. Hatzoglou, 2005 Internal ribosome entry sites in cellular mRNAs: mystery of their existence. J Biol Chem 280**:** 23425-23428.

Kondo, S., and R. Ueda, 2013 Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics 195**:** 715-721.

Korona, D., S. A. Koestler and S. Russell, 2017 Engineering the Drosophila Genome for Developmental Biology. J Dev Biol 5.

Lackner, D. H., A. Carre, P. M. Guzzardo, C. Banning, R. Mangena *et al.*, 2015 A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat Commun 6**:** 10237.

Lee, H., C. J. McManus, D. Y. Cho, M. Eaton, F. Renda *et al.*, 2014 DNA copy number evolution in Drosophila cell lines. Genome Biol 15**:** R70.

Lee, P. T., J. Zirin, O. Kanca, W. W. Lin, K. L. Schulze *et al.*, 2018 A gene-specific T2A-GAL4 library for Drosophila. Elife 7.

Maresca, M., V. G. Lin, N. Guo and Y. Yang, 2013 Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res 23**:** 539-546.

Morin, X., R. Daneman, M. Zavortink and W. Chia, 2001 A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. Proc Natl Acad Sci U S A 98**:** 15050-15055.

Mouilleron, H., V. Delcourt and X. Roucou, 2016 Death of a dogma: eukaryotic mRNAs can code for more than one protein. Nucleic Acids Res 44**:** 14-23.

Nagasaki, A., T. K. S, T. Yumoto, M. Imaizumi, A. Yamagishi *et al.*, 2017 The Position of the GFP Tag on Actin Affects the Filament Formation in Mammalian Cells. Cell Struct Funct 42**:** 131-140.

Perkins, L. A., L. Holderbaum, R. Tao, Y. Hu, R. Sopko *et al.*, 2015 The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics 201**:** 843-852.

Port, F., H. M. Chen, T. Lee and S. L. Bullock, 2014 Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A 111**:** E2967-2976.

Port, F., N. Muschalik and S. L. Bullock, 2015 Systematic Evaluation of Drosophila CRISPR Tools Reveals Safe and Robust Alternatives to Autonomous Gene Drives in Basic Research. G3 (Bethesda) 5**:** 1493-1502.

Preston, C. R., C. C. Flores and W. R. Engels, 2006 Differential usage of alternative pathways of double-strand break repair in Drosophila. Genetics 172**:** 1055-1068.

Quadros, R. M., H. Miura, D. W. Harms, H. Akatsuka, T. Sato *et al.*, 2017 Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol 18**:** 92.

Ren, X., J. Sun, B. E. Housden, Y. Hu, C. Roesel *et al.*, 2013 Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc Natl Acad Sci U S A 110**:** 19012-19017.

Robinett, C. C., A. Straight, G. Li, C. Willhelm, G. Sudlow *et al.*, 1996 In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol 135**:** 1685-1700.

Rorth, P., 1996 A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. Proc Natl Acad Sci U S A 93**:** 12418-12422.

Sadelain, M., E. P. Papapetrou and F. D. Bushman, 2011 Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer 12**:** 51-58.

Schmid-Burgk, J. L., K. Honing, T. S. Ebert and V. Hornung, 2016 CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat Commun 7**:** 12338.

Suzuki, K., Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu *et al.*, 2016 In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature 540**:** 144-149.

Venken, K. J., A. Sarrion-Perdigones, P. J. Vandeventer, N. S. Abel, A. E. Christiansen *et al.*, 2016 Genome engineering: Drosophila melanogaster and beyond. Wiley Interdiscip Rev Dev Biol 5**:** 233-267.

Venken, K. J., K. L. Schulze, N. A. Haelterman, H. Pan, Y. He *et al.*, 2011 MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8**:** 737-743.

Viswanatha, R., Z. Li, Y. Hu and N. Perrimon, 2018 Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in Drosophila cells. Elife 7.

Yu, Z., M. Ren, Z. Wang, B. Zhang, Y. S. Rong *et al.*, 2013 Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics 195**:** 289-291.

Figure 1

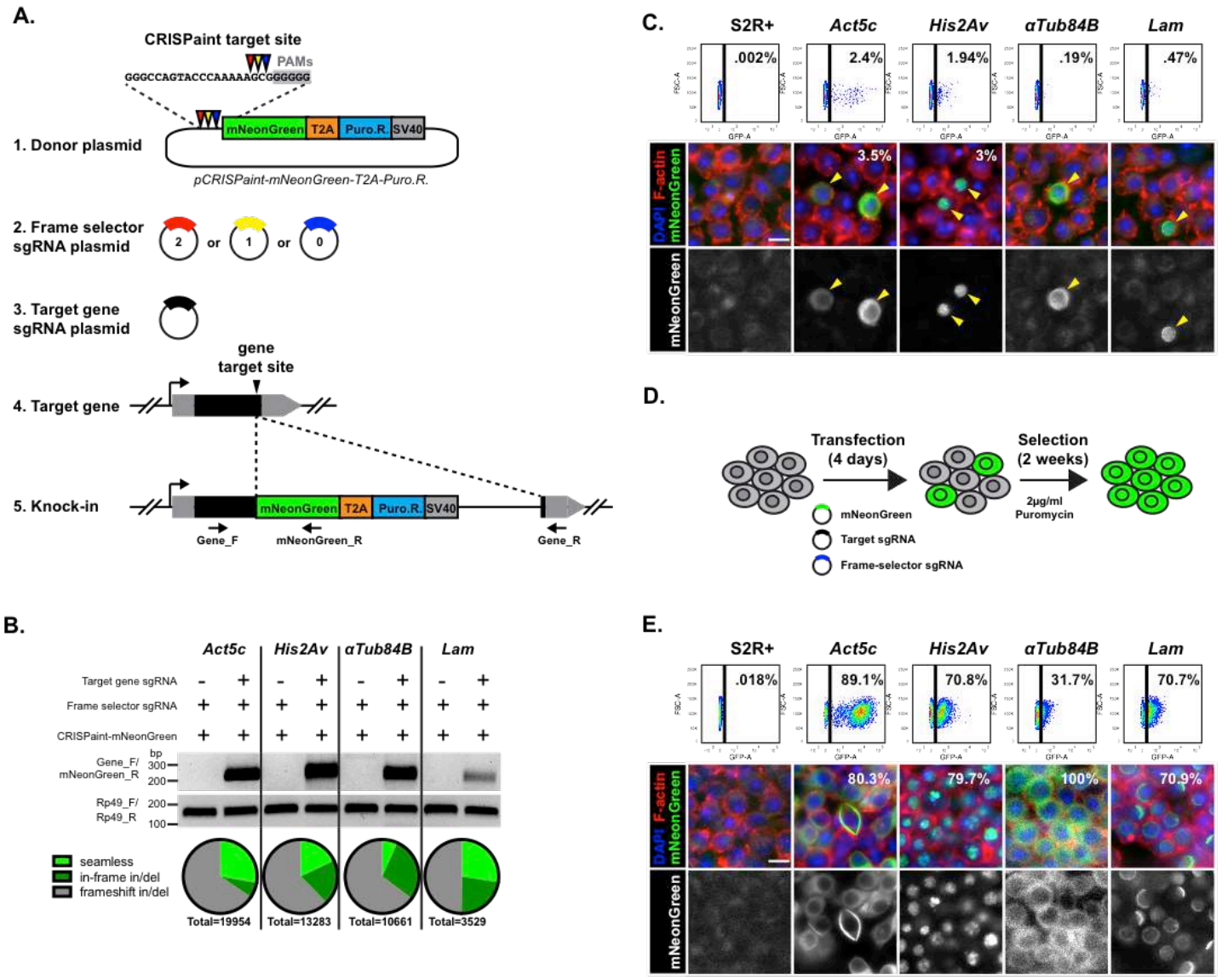


Figure 1 – Knock-in of *mNeonGreen-T2A-PuroR* into Drosophila S2R+ cells using homology-independent insertion. (**A**) Schematic of CRISPaint knock-in approach. *mNeonGreen-T2A-PuroR* is inserted into 3' coding sequence. (**B**) Analysis of knock-in efficiency of transfected cells by diagnostic PCR (DNA gel image) and next-generation sequencing (pie charts). (**C**) Analysis of knock-in efficiency of transfected cells by FACs and confocal microscopy. Numbers indicate percentage of cells with fluorescence. F-actin stained using Phalloidin-TRITC (red), nuclei labeled with DAPI (blue), mNeonGreen signal is in green. Scale bar 10μm. (**D**) Schematic of Puromycin selection of mNeonGreen-expressing cells. (**E**) Analysis of knock-in frequency of puromycin-selected cells using FACs and confocal microscopy. Numbers indicate percentage of cells with green fluorescence. Scale bar 10μm.

Figure 2

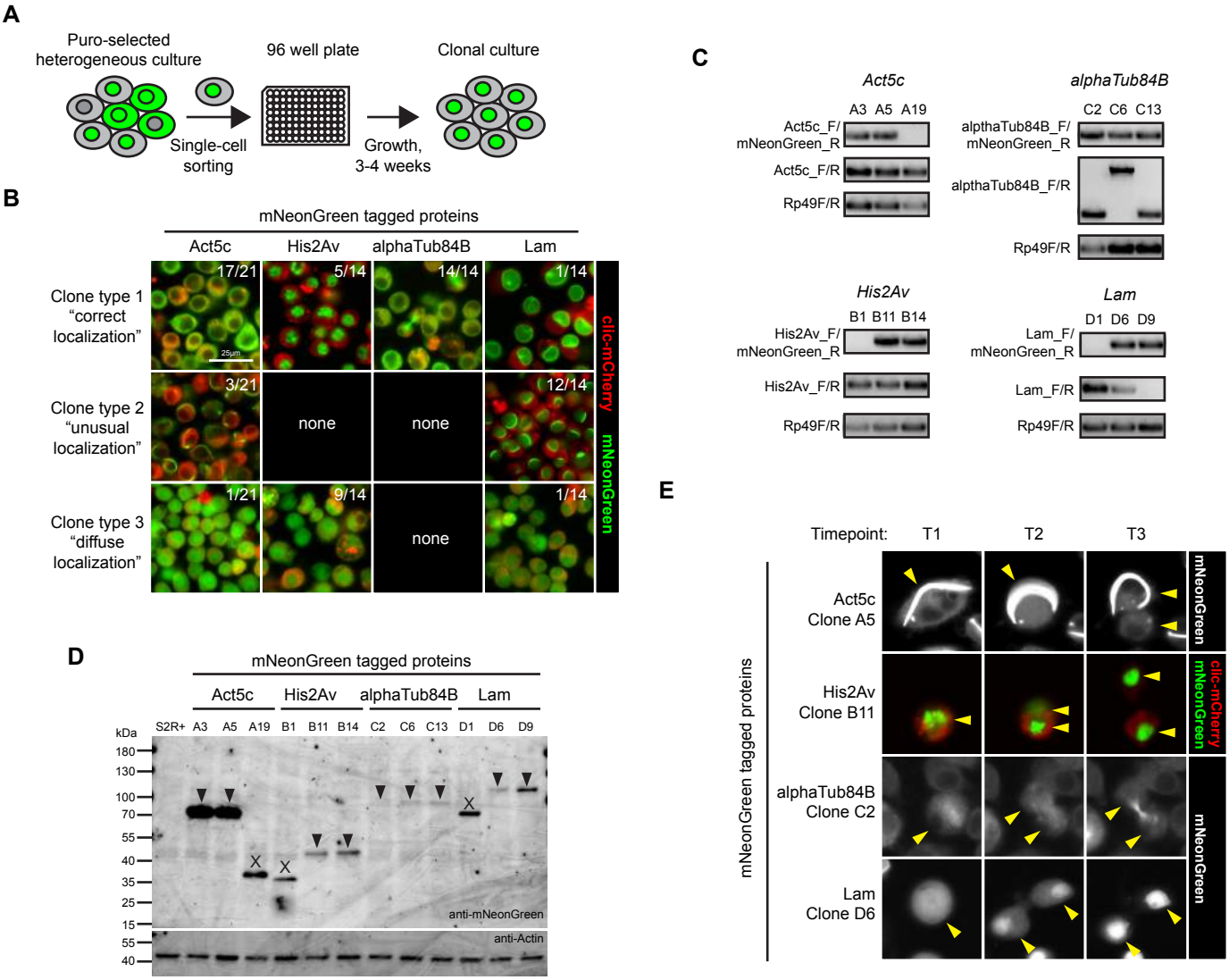


Figure 2. Analysis of S2R+ mNeonGreen-expressing single-cell cloned lines. (**A**) Schematic of FACS isolation of single-cell clones expressing mNeonGreen. (**B**) Confocal images of live mNeonGreen-expressing cell lines, categorized into three clone types. Numbers indicate the frequency of each clone type for each gene targeted. Images show fluorescence from Clic-mCherry (red) and mNeonGreen (green). Scale bar 25µm. (**C**) Agarose gel with PCR fragments amplified from knock-in (Gene_F/mNeonGreen_R) and non-knock-in loci (Gene_F/R). Positive control bands were amplified from *Rp49* genomic sequence. (**D**) Western blot detecting mNeonGreen protein fusions. Arrowheads indicate expected molecular weight. X's indicate incorrect molecular weight. (**E**) Confocal images of live S2R+ cells expressing mNeonGreen protein fusions during cell division at three timepoints. Arrowheads indicate cells before/after cell division.

Figure 3

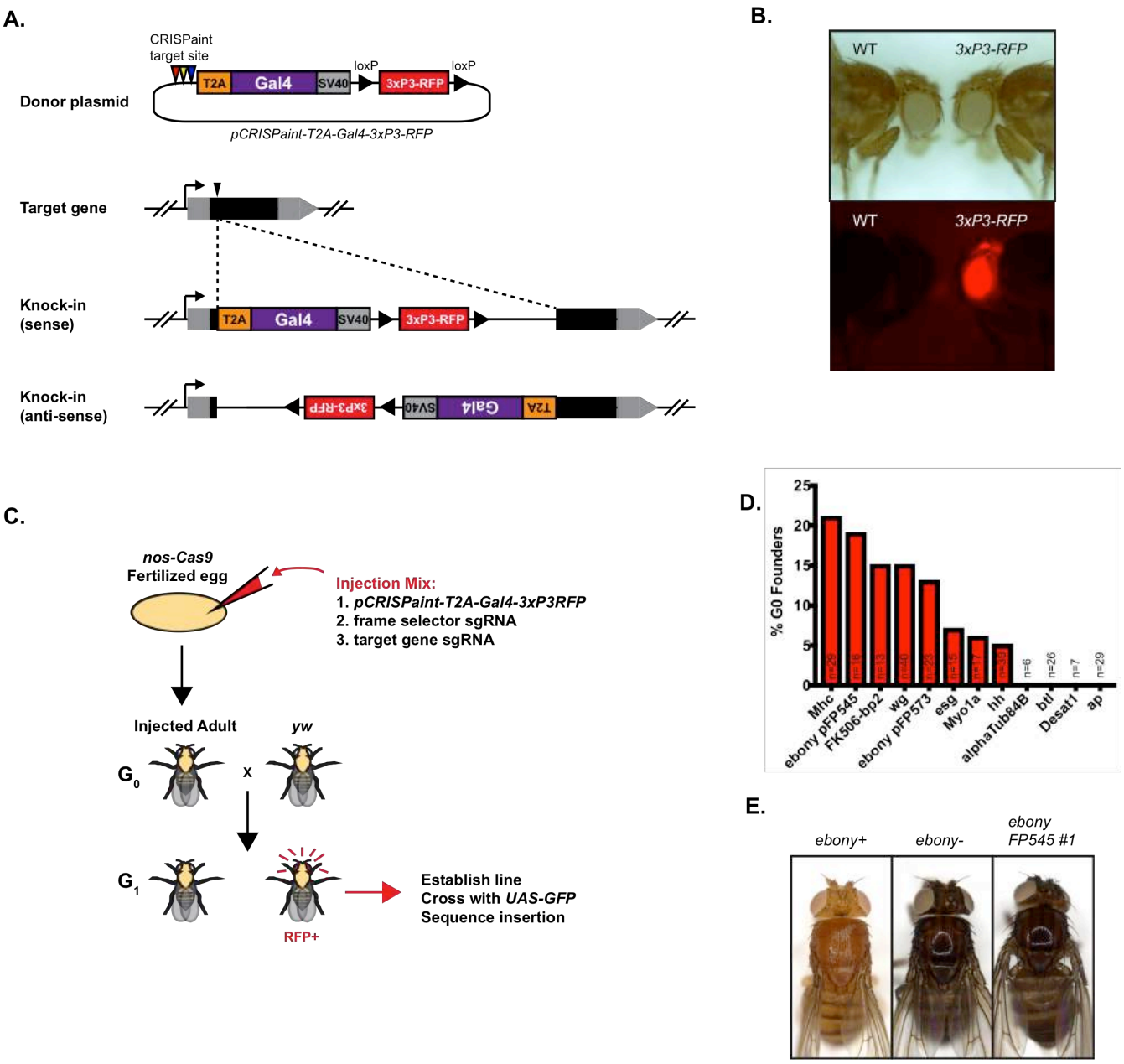


Figure 3 – Knock-in of *T2A-Gal4* into the *Drosophila* germ line using homology-independent insertion. (**A**) Schematic of knock-in approach. *pCRISPaint-T2A-Gal4-3xP3-RFP* is inserted into 5' coding sequence. (**B**) Images of adult flies with 3xP3-RFP fluorescence in the eye. Top panel is brightfield, bottom panel is fluorescence. (**C**) Schematic of plasmid injections, fly crosses, and analysis of insertions. (**D**) Graph with results of knock-in efficiency for 12 sgRNA target sites and 11 genes. (**E**) Image of adult flies. Homozygous *ebony-T2A-Gal4 FP545* #1 flies have dark cuticle pigment.

Figure 4

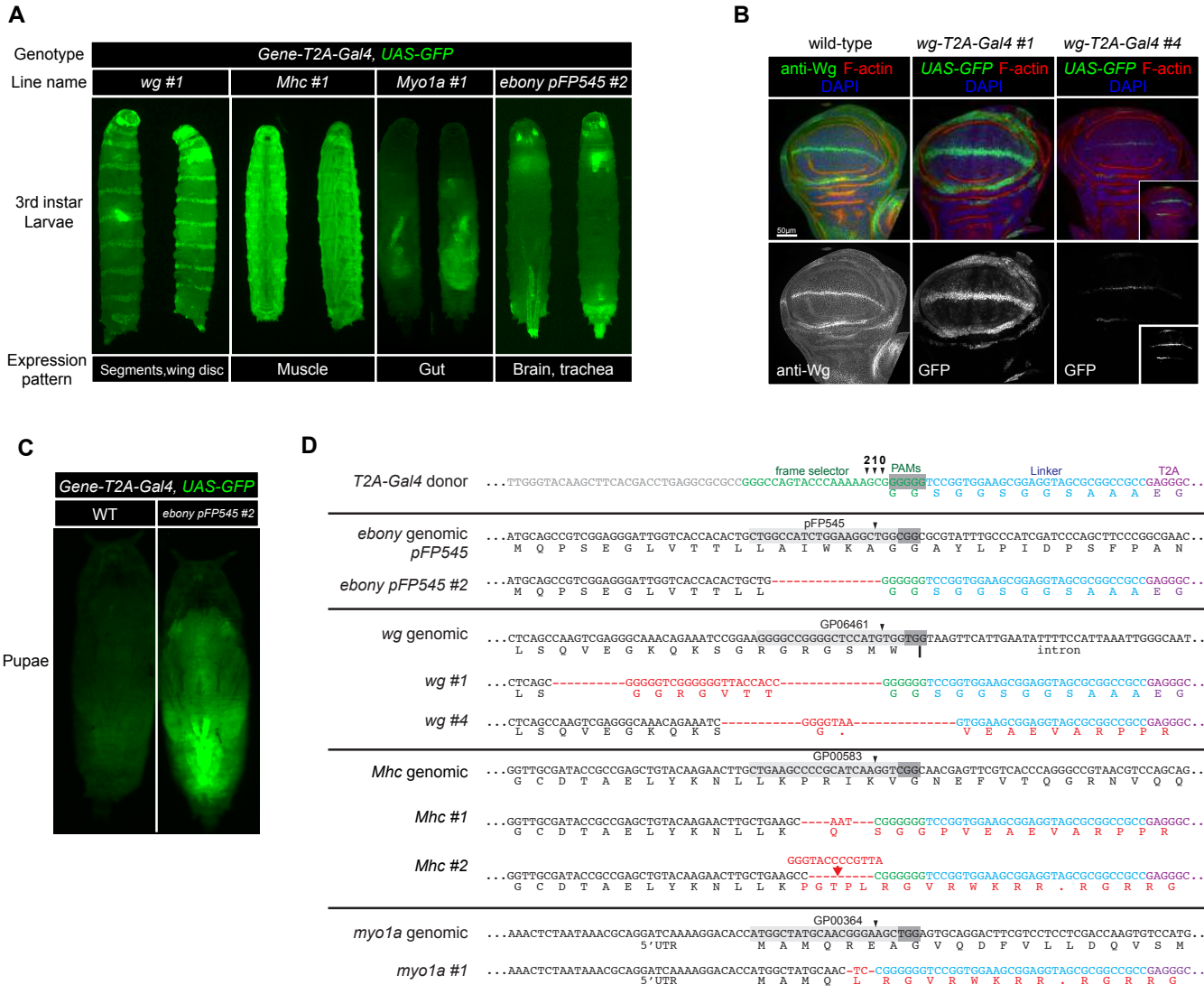


Figure 4 – Germ line insertions that express *T2A-Gal4* under the control of the target gene. (**A**) Fluorescence images of 3rd instar larvae with indicated genotypes. Expression of Gal4 under control of the target gene drives expression of the *UAS-GFP* reporter. (**B**) Confocal images of wing imaginal discs showing protein staining of Wg protein (anti-wg, green) or *UAS-GFP* expression (green). GFP fluorescence was recorded at identical exposure settings for lines *wg-T2A-Gal4* #1 and #4. Inset shows digitally increased GFP signal. Scale bar 50µm. (**C**) Fluorescence image of pupae, *ebony-T2A-Gal4 pFP545* #2 expression is visible throughout the cuticle. (**D**) Sequence structure of *T2A-Gal4* insertions that express Gal4.

Figure 5

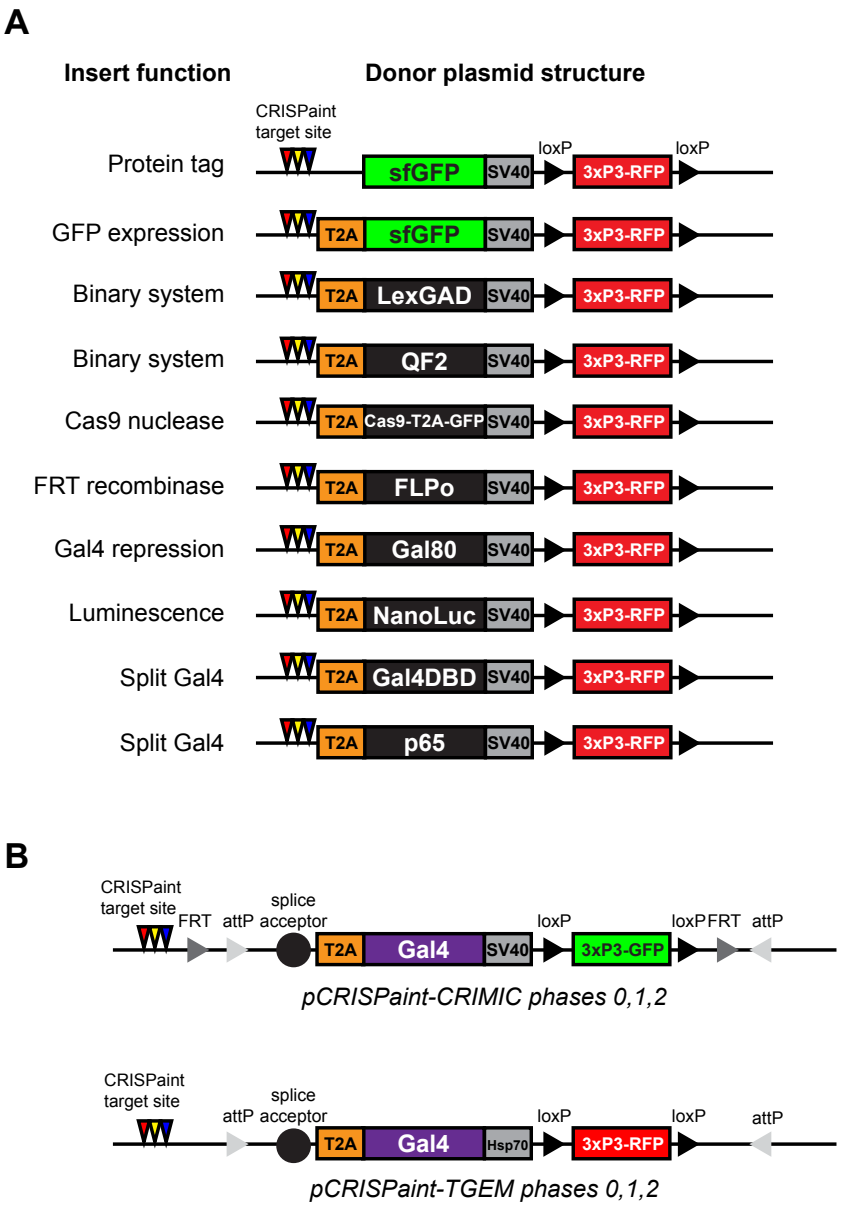
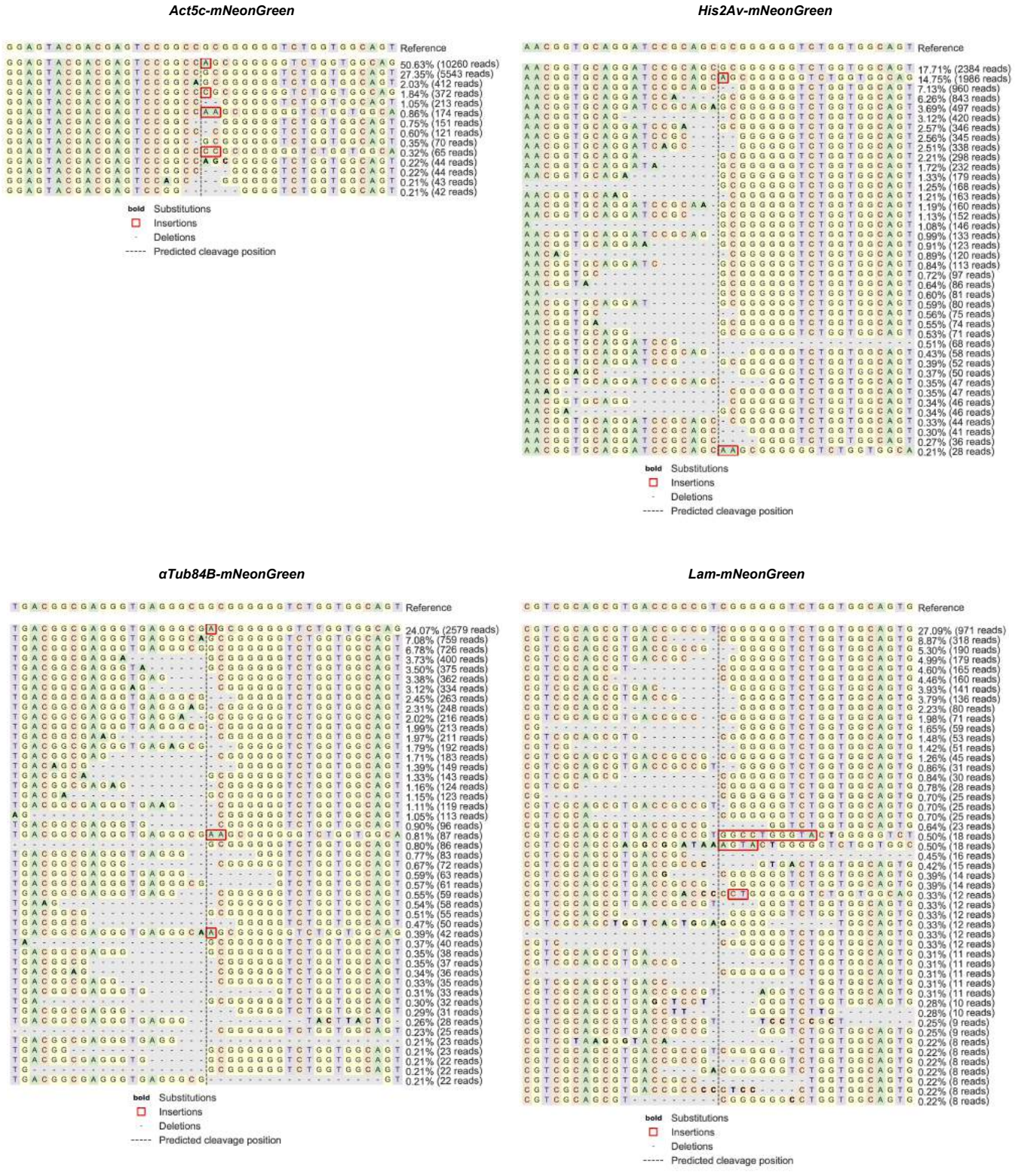


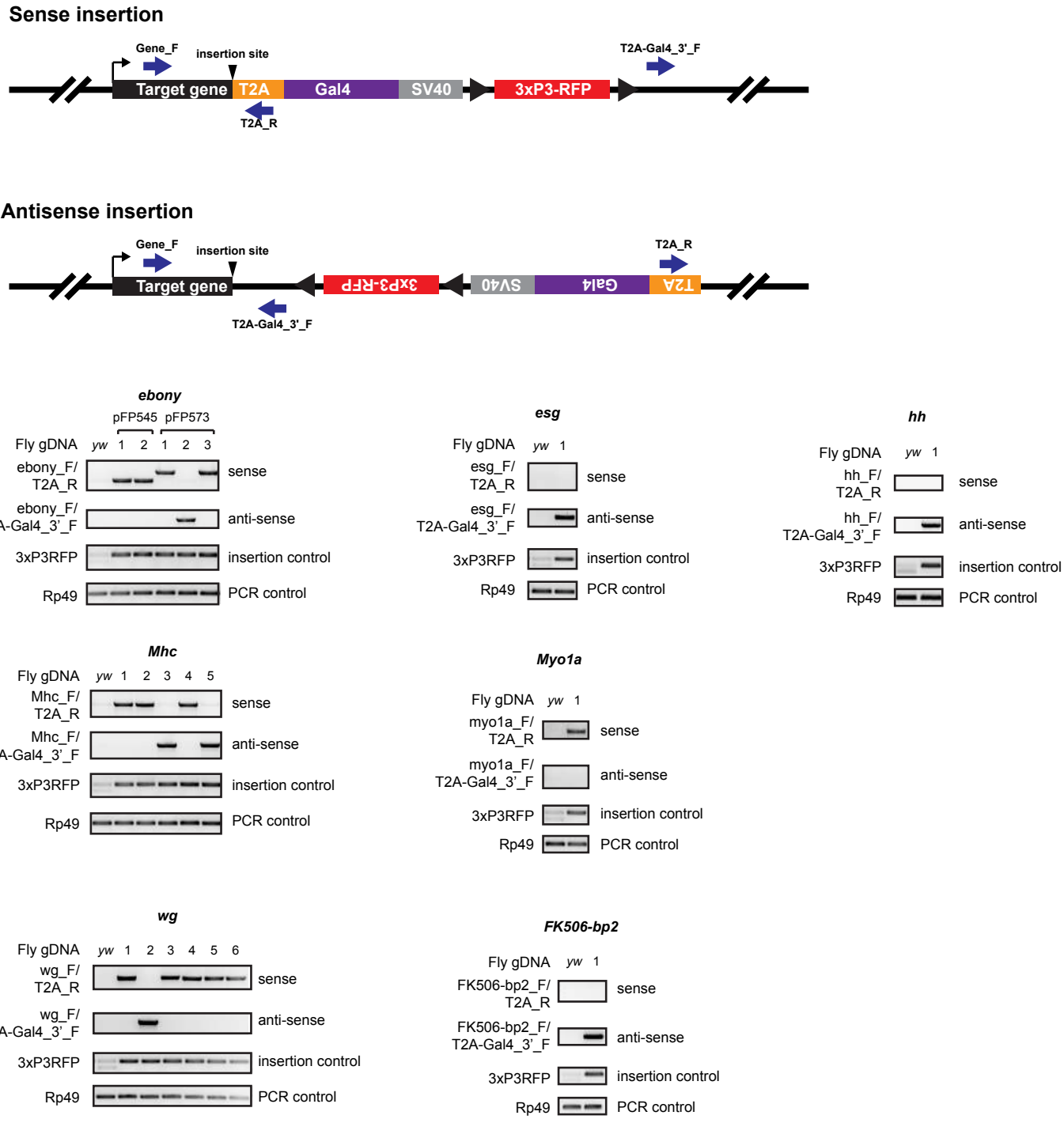
Figure 5 – A collection of CRISPaint-compatible donor plasmids for germ line knock-ins. (**A**) Donor plasmids for insertion into 5' coding sequence. (**B**) Donor plasmids for insertion into intronic sequence, modified from CRIMIC and T-GEM HDR donor plasmids.

Supplemental Figure 1



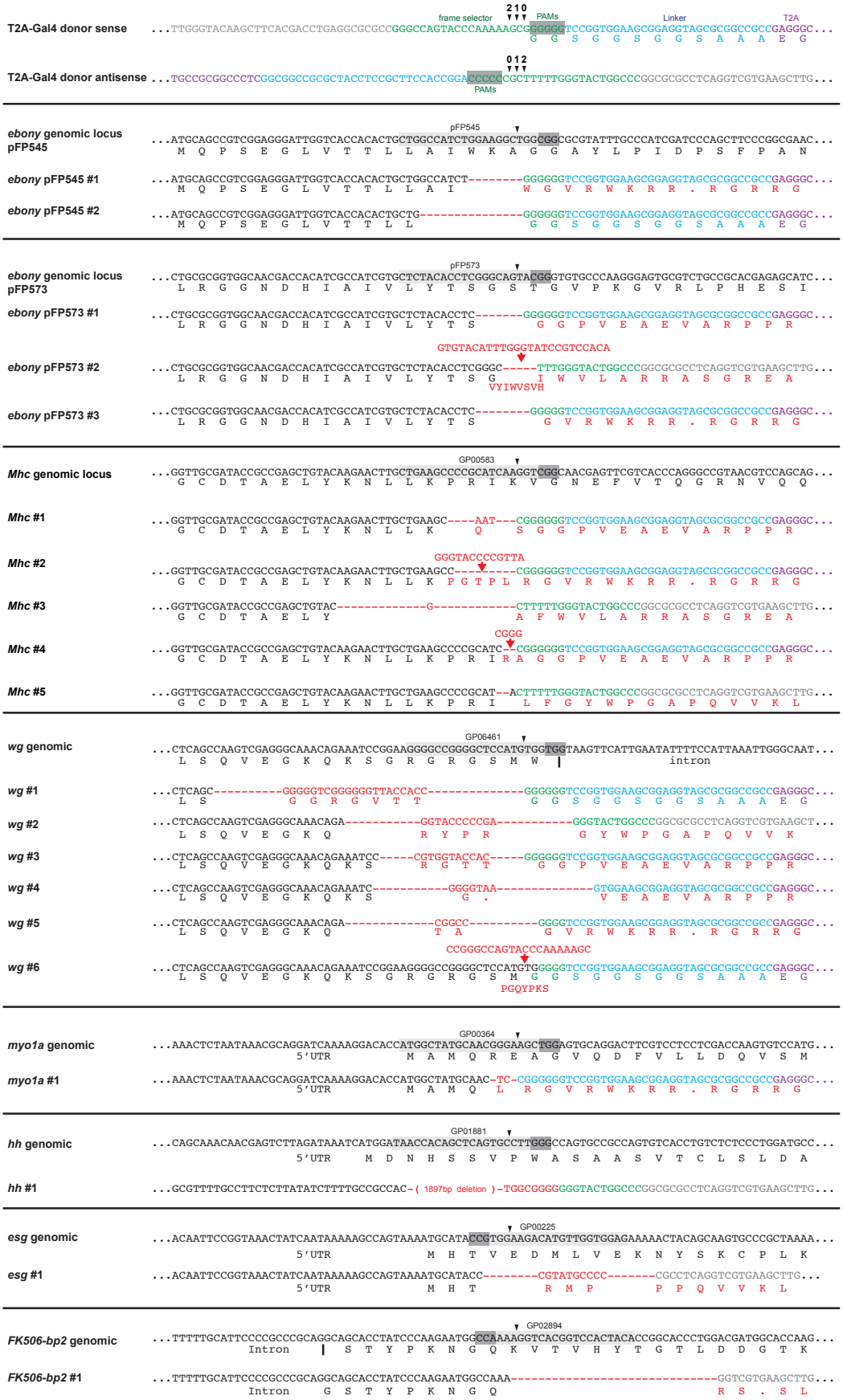
Supplemental Figure 1 – Next-generation sequencing analysis of mNeonGreen insertion sites in transfected S2R+ cells using CRISPresso2

Supplemental Figure 2



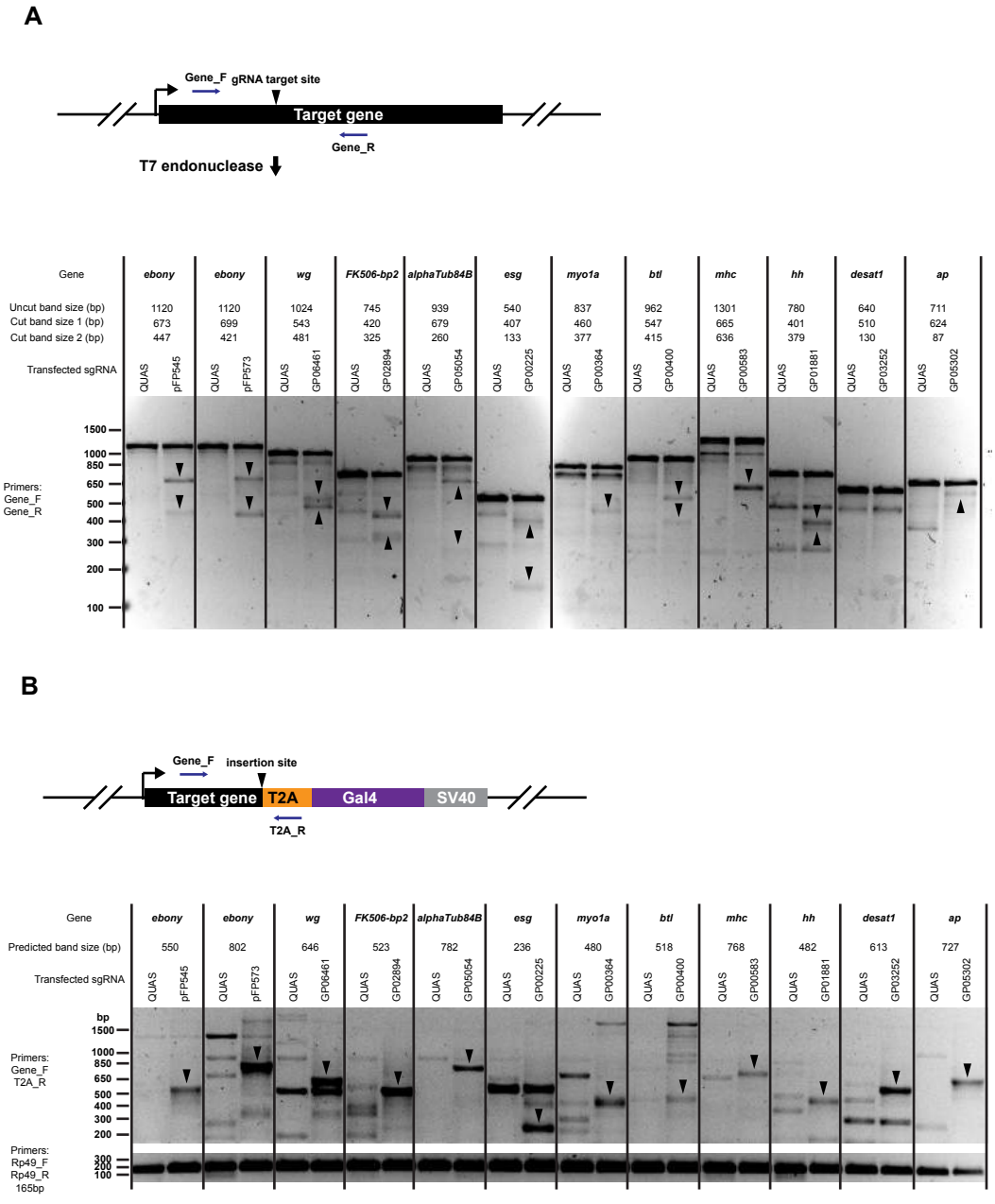
Supplemental Figure 2 – Diagnostic PCR of *CRISPaint-T2A-Gal4* insertions in RFP+ fly lines to confirm their insertion site and orientation.

Supplemental Figure 3



Supplemental Figure 3 – Sequence structure of *CRISPaint-T2A-Gal4* insertions in RFP+ fly lines

Supplemental Figure 4



Supplemental Figure 4 – Cutting efficiency of 12 sgRNAs in transfected S2R+ cells. (**A**) T7 endonuclease assay. (**B**) Diagnostic PCR to detect for presence of sense orientation *CRISPaint-T2A-Gal4* insertion events.

	Gene and sgRNA information			
Gene Name	**Protein localization**	**sgRNA plasmid (TRiP)**	**sgRNA target site with PAM**	**frame selector sgRNA**
Act5c	actin	GP07595	GCGGTGCACAATGGAGGGGCCGG	2
His2Av	chromatin	GP07596	GGTGCAGGATCCGCAGCGGAAGG	2
αTub84B	microtubules	GP07609	CGGCGAGGGTGAGGGCGCTGAGG	2
Lamin	nuclear envelope	GP07612	CGCAGCGTGACCGCCGTGGACGG	1
Untransfected				

	CRISPresso2 output			
Gene Name	**seamless**	**in-frame in/del**	**frameshift in/del**	**% in-frame**
Act5c	5543	1343	13068	34.50937155
His2Av	2384	2620	8279	37.6722126
αTub84B	726	3034	6901	35.26873652
Lamin	971	801	1757	50.21252479
Untransfected				

	FACs Transfected			FACs Puro-selected		
Gene Name	**mNeonGreen+**	**Single cells total**	**% mNeonGreen+**	**mNeonGreen+**	**Single cells total**	**% mNeonGreen**
Act5c	1963	80589	2.435816302	45355	50883	89.1358607
His2Av	1554	80291	1.935459765	32780	46298	70.80219448
αTub84B	150	78752	0.190471353	16682	52643	31.6889235
Lamin	385	81688	0.471305455	34826	49234	70.73567047
Untransfected	2	83869	0.002384671	10	55656	0.017967515

	Cell counting transfected (n=6)			Cell counting Puro-selected (n=3)		
Gene Name	**mNeonGreen+**	**Total cells**	**Average % mNeonGreen+**	**mNeonGreen+**	**Total cells**	**Average % mNeonGreen+**
Act5c	17/17/26/18/22/25	571/679/577/599/574/613	3.5	342/280/318	397/398/377	80.3
His2Av	26/21/12/19/13/19	634/634/568/525/669/612	3	314/327/315	397/402/400	79.7
αTub84B				412/400/403	412/400/403	100
Lamin				270/282/259	382/388/374	70.9
Untransfected						

Supplemental Table 1 – Quantification of CRISPaint-mNeonGreen insertion events in transfected and puro-selected S2R+ cells by CRISPresso2, FACs, and cell counting of confocal images.

					endogenous gene			
Clone name	**Gene-mNeonGreen**	**predicted mNeonGreen fusion (kDa)**	**Observed Localization**	**mNeon insertion site**	**allele 1**	**allele 2**	**allele 3**	**allele 4**
A3	Act5c	71.4	very infrequent small rods	seamless	1bp deletion	1bp insertion		
A5	Act5c	71.4	very frequent long rods	seamless	numerous in/dels			
A19	Act5c	71.4	no rods, cyto and nuclear	N/A	numerous in/dels			
B1	His2Av	44.5	cyto and nuclear	N/A	4bp deletion	19bp deletion		
B11	His2Av	44.5	nuclear	seamless	3bp deletion	2bp deletion	1bp deletion	27bp deletion
B14	His2Av	44.5	nuclear	3 bp deletion, 1bp change	19bp deletion	3bp deletion	1bp deletion	
C2	alphatub84B	79.5	mitotic spindles	seamless	wt	6bp deletion	1bp deletion	
C6	alphatub84B	79.5	mitotic spindles	seamless	insertion of 1482bp from alphatub84D			
C13	alphatub84B	79.5	mitotic spindles	3bp deletion	wt	6bp insertion		
D1	Lam	100.9	cytoplasmic, nuclear	N/A	8bp deletion	4bp deletion	3bp insertion	
D6	Lam	100.9	nuclear lamina	seamless	3bp deletion			
D9	Lam	100.9	nuclear lamina, polarized	seamless	N/A			

Supplemental Table 2 – Molecular characterization of single cell cloned mNeonGreen-expressing S2R+ lines.

Gene	Chr.	target gene sgRNA	target site (no PAM)	Frame selector sgRNA	# G0 crosses	# G0 founders	% founders/total	# G1 progeny screened	# G1 RFP+ progeny	% RFP G1/total G1
Mhc	2	GP00583	CTGAAGCCCCGCATCAAGGT	1	29	6	21	2086	21	1.01
ebony	3	pFP545	TGGCCATCTGGAAGGCTGG	1	16	3	19	1298	11	0.85
FK506-bp2	2	GP02894	TGTAGTGGACCGTGACCTTT	1	13	2	15	901	2	0.22
wg	2	GP06461	GGGGCCGGGGCTCCATGTGG	0	40	6	15	4423	26	0.59
ebony	3	pFP573	TCTACACCTCGGGCAGTAC	1	23	3	13	2121	27	1.27
esg	2	GP00225	CTCCACCAACATGTCTTCCA	2	15	1	7	772	1	0.13
Myo1a	2	GP00364	ATGGCTATGCAACGGGAAGC	1	17	1	6	860	1	0.12
hh	3	GP01881	TAACCACAGCTCAGTGCCTT	2	39	2	5	1975	3	0.15
alphaTub84B	3	GP05054	GCTGACGGTAGGTTCCGGTA	1	6	0	0	506	0	0.00
btl	2	GP00400	GGCAAAAGTGCCGATCACGC	2	26	0	0	2085	0	0.00
Desat1	3	GP03252	AAGCTGCAGGAGGACTCCAC	1	7	0	0	386	0	0.00
ap	2	GP05302	CACCACCTGTAGCACATCAA	0	29	0	0	2928	0	0.00

Supplemental Table 3 – Germ line knock-in efficiency of *CRISPaint-T2A-Gal4*

#	Gene targeted	sgRNA	G0 cross #	Insert orientation	Insert frame	Insertion site sequence	Gal4 expression	Homozygous phenotype	Complementation test
1	ebony	pFP545	1	Sense	Out	8bp del	none	dark cuticle pigment	fail to complement e1 on TM3
2	ebony	pFP545	2	Sense	In	15bp del	anterior and posterior trachae in larvae, whole body cuticle in pupae and adult	dark cuticle pigment	fail to complement e1 on TM3
3	ebony	pFP573	1	Sense	Out	7bp del	none	dark cuticle pigment	fail to complement e1 on TM3
4	ebony	pFP573	2	Antisense	N/A	5bp del, 25bp ins	none	dark cuticle pigment	fail to complement e1 on TM3
5	ebony	pFP573	3	Sense	Out	8bp deletion	none	dark cuticle pigment	fail to complement e1 on TM3
6	hh	GP01881	1	Antisense	N/A	1897bp del, 8bp ins	none	lethal	fail to complement hh[AC], Df(3R)ED5296
7	Mhc	GP00583	1	Sense	Out	10bp del, 3bp ins	muscle	lethal	fail to complement P{lacW}Mhck10423, Df(2L)H20
8	Mhc	GP00583	2	Sense	Out	9bp del, 13bp ins	muscle	lethal	fail to complement P{lacW}Mhck10423, Df(2L)H20
9	Mhc	GP00583	3	Antisense	N/A	27bp del, 1bp ins	none	lethal	fail to complement P{lacW}Mhck10423, Df(2L)H20
10	Mhc	GP00583	4	Sense	Out	2bp del, 4bp ins	none	lethal	fail to complement P{lacW}Mhck10423, Df(2L)H20
11	Mhc	GP00583	5	Antisense	N/A	2bp deletion	none	lethal	fail to complement P{lacW}Mhck10423, Df(2L)H20
12	Myo1a	GP00364	1	Sense	Out	4bp del, 2bp ins	gut	viable	complement Df(2L)ED8142
13	esg	GP00225	1	Antisense	N/A	25bp del, 10bp	none	lethal	fail to complement P{enG}esg[G66]
14	FK506-bp2	GP02894	1	Antisense	N/A	30bp del	none	viable	complement Df(2R)Exel6069
15	wg	GP06461	1	Sense	In	45bp del, 21bp ins	wingless, strong expression	lethal	fail to complement wg[l-17], wg[l-8], and Df(2L)BSC291
16	wg	GP06461	2	Antisense	N/A	33bp del, 11bp ins	none	lethal	fail to complement wg[l-17], wg[l-8], and Df(2L)BSC291
17	wg	GP06461	3	Sense	Out	21bp del, 11bp ins	none	lethal	fail to complement wg[l-17], wg[l-8], and Df(2L)BSC291
18	wg	GP06461	4	Sense	Out	32bp del, 7bp ins	wingless, weak expression	lethal	fail to complement wg[l-17], wg[l-8], and Df(2L)BSC291
19	wg	GP06461	5	Sense	Out	28bp del, 5bp ins	none	in progress	fail to complement wg[l-17], wg[l-8], and Df(2L)BSC291
20	wg	GP06461	6	Sense	In	21bp insertion	none	in progress	in progress

Supplemental Table 4 – Molecular and phenotypic characterization of 20 RFP+ fly strains, each carrying a distinct *CRISPaint-T2A-Gal4* insertion.

gene	sgRNA name	sgRNA sequence	DRSC efficiency score
ebony	pFP545	gTGGCCATCTGGAAGGCTGG	5.56968
ebony	pFP573	gTCTACACCTCGGGCAGTAC	5.08877
wg	GP06461	GGGGCCGGGGCTCCATGTGG	5.61857
FK506-bp2	GP02894	TGTAGTGGACCGTGACCTTT	6.5446
alphaTub84B	GP05054	GCTGACGGTAGGTTCCGGTA	6.20359
esg	GP00225	CTCCACCAACATGTCTTCCA	9.35069
Myo1a	GP00364	ATGGCTATGCAACGGGAAGC	7.98579
btl	GP00400	GGCAAAAGTGCCGATCACGC	3.52487
Mhc	GP00583	CTGAAGCCCCGCATCAAGGT	8.00559
hh	GP01881	TAACCACAGCTCAGTGCCTT	6.99912
Desat1	GP03252	AAGCTGCAGGAGGACTCCAC	6.60024
ap	GP05302	CACCACCTGTAGCACATCAA	5.73411

Supplemental Table 5 – Efficiency scores for 12 sgRNAs used in germ line knock-ins

Name	Sequence	How used
JB880_frame+0_gRNA_top	GTCGgccagtacccaaaaagcggg	for cloning pCFD3-CRISPaint_frameselector
JB881_frame+0_gRNA_bot	AAACcccgctttttgggtactggc	for cloning pCFD3-CRISPaint_frameselector
JB882_frame+1_gRNA_top	GTCGggccagtacccaaaaagcgg	for cloning pCFD3-CRISPaint_frameselector
JB883_frame+1_gRNA_bot	AAACccgctttttgggtactggcc	for cloning pCFD3-CRISPaint_frameselector
JB884_frame+2_gRNA_top	GTCGgggccagtacccaaaaagcg	for cloning pCFD3-CRISPaint_frameselector
JB885_frame+2_gRNA_bot	AAACcgcttttgggtactggccc	for cloning pCFD3-CRISPaint_frameselector
JB886_target-T2A-Gal4overlap_gBlock	ttcacgacctgaggcgcgccgggccagtacccaaaaagcgggggtccggtggaagcggagg tagcgcggccgccgagggccgcggcagcctgctgacctgcggcgatgtggaggagaaccccg ggcccGCTAGCatgaagctactgtcttctatcgaacaagcatgcgatatttgccgacttaaa aagctcaagtgctccaaagaaaaccgaagtgcgccaagtgtct	gBlock containing CRISPaint site and T2A and overlap sequence
JB877_Gal4_F	atgaagctactgtcttctatcgaaca	to amplify Gal4-3xP3RFP
JB878_target-T2A-Gal4_P3RFP_R	gggaacaaaagctggagctcataacttcgtatagcatacattatacgaagttatCGTATGGG CCTTCGCTGCTTACAG	to amplify Gal4-3xP3RFP
JB915_sfGFP_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcGTGTCCAAGGGCGAGGAG	for cloning into JAB290 cut with NheI/KpnI
JB916_sfGFP_HIKI_R	caaagatcctctagaggtaccCTACTTGTACAGCTCATCCATGC	for cloning into JAB290 cut with NheI/KpnI
JB1000_LexGAD_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGCCACCCAAGAAGAAGC	for cloning into JAB290 cut with NheI/KpnI
JB1001_LexGAD_HIKI_R	caaagatcctctagaggtaccCTACTCCTTCTTTGGGTTCGG	for cloning into JAB290 cut with NheI/KpnI
JB998_QF2_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGCCACCCAAGCGCAAA	for cloning into JAB290 cut with NheI/KpnI
JB999_QF2_HIKI_R	caaagatcctctagaggtaccTCACTGTTCGTATGTATTAATGTCG	for cloning into JAB290 cut with NheI/KpnI
JB1055_Cas9-T2A-EGFP_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGGATTACAAGGATCACGATG	for cloning into JAB290 cut with NheI/KpnI
JB1056_Cas9-T2A-EGFP_HIKI_R	caaagatcctctagaggtaccTTAGTACAGCTCGTCCATGCC	for cloning into JAB290 cut with NheI/KpnI
JB1002_FLPo_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcatgagccagttcgacatcct	for cloning into JAB290 cut with NheI/KpnI
JB1003_FLPo_HIKI_R	caaagatcctctagaggtacctcagatccgcctgttgatgtagc	for cloning into JAB290 cut with NheI/KpnI
JB1004_Gal80_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGGACTACAACAAGAGATCTTCG	for cloning into JAB290 cut with NheI/KpnI
JB1005_Gal80_HIKI_R	caaagatcctctagaggtaccTTATAAACTATAATGCGAGATATTGCTAA	for cloning into JAB290 cut with NheI/KpnI
JB1008_Nluc_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGGTCTTCACACTCGAAGATTT	for cloning into JAB290 cut with NheI/KpnI
JB1009_Nluc_HIKI_R	caaagatcctctagaggtaccTTACGCCAGAATGCGTTC	for cloning into JAB290 cut with NheI/KpnI
JB1032_Gal4DBD_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGCTGGAGATCCGCGCC	for cloning into JAB290 cut with NheI/KpnI
JB1033_Gal4DBD_HIKI_R	caaagatcctctagaggtaccTTACGATACCGTCAGTTGCC	for cloning into JAB290 cut with NheI/KpnI
JB1030_p65_HIKI_T2A_F	tgtggaggagaaccccgggcccgctagcATGGATAAAGCGGAATTAATTCC	for cloning into JAB290 cut with NheI/KpnI
JB1031_p65_HIKI_R	caaagatcctctagaggtaccTTACTTGCCGCCGCCCAG	for cloning into JAB290 cut with NheI/KpnI
JB917_sfGFP_HIKI_F	ggtggaagcggaggtagcgcggccgccGTGTCCAAGGGCGAGGAG	for cloning into JAB290 cut with NotI/KpnI
JB916_sfGFP_HIKI_R	caaagatcctctagaggtaccCTACTTGTACAGCTCATCCATGC	for cloning into JAB290 cut with NotI/KpnI
JB969_crispaintsite_NsiI_top	agggccagtacccaaaaagcgggggtTGCA	for cloning into CRIMIC pM37
JB970_crispaintsite_NsiI_bottom	acccccccgcttttgggtactggccctTGCA	for cloning into CRIMIC pM37
JB971_crispaintsite_AgeI-NotI_top	ccggtgggccagtacccaaaaagcgggggggc	for cloning into T-GEM
JB972_crispaintsite_AgeI-NotI_bottom	ggccgccccccgcttttgggtactggccca	for cloning into T-GEM
JB1355_FK506-bp2_geno_1F	acgcgccaaaatacaaaaac	amplification of endogenous target gene and knock-in
JB1356_FK506-bp2_geno_1R	GCCTATTCGACCTTGAGCAG	amplification of endogenous target gene
JB1357_alphaTub84B_geno_1F	tttgtgtgggcaaaattcaa	amplification of endogenous target gene and knock-in
JB1358_alphaTub84B_geno_1R	GCTTGGACTTCTTGCCGTAG	amplification of endogenous target gene
JB1359_esg_geno_1F	CGTTTGGTATTTGTGCATCG	amplification of endogenous target gene and knock-in
JB1360_esg_geno_1R	GTAGGGCGACATGTGGAAGT	amplification of endogenous target gene
JB1361_btl_geno_1F	aactaagggaggggcaaaaa	amplification of endogenous target gene and knock-in
JB1362_btl_geno_1R	CGTCCACCAAGGATTTGAGT	amplification of endogenous target gene
JB1363_Desat1_geno_1F	aatccacctggtgcttgttc	amplification of endogenous target gene and knock-in
JB1364_Desat1_geno_1R	GTAACCGAAGGCGATGATGT	amplification of endogenous target gene
JB1365_ap_geno_1F	ttgcaaatctgtcaggaacg	amplification of endogenous target gene and knock-in
JB1366_ap_geno_1R	ATCTGGACACGAGGATGAGG	amplification of endogenous target gene
JB959_ebony_geno_F	gcattagcctgcattgcata	amplification of endogenous target gene and knock-in
JB960_ebony_geno_R	CACGCCCTCATCGAAATAGT	amplification of endogenous target gene
JB961_wg_geno_F	CAGTTAAGCGTTGGCACTGA	amplification of endogenous target gene and knock-in
JB962_wg_geno_R	ttgttgcatctctgcggtag	amplification of endogenous target gene
JB963_Myo1a_geno_F	TCGTCGTCATCAACAGAAGC	amplification of endogenous target gene and knock-in
JB964_Myo1a_geno_R	tctggagtggaaccgaaaac	amplification of endogenous target gene
JB965_Mhc_geno_F	cggctaaagactgacccaaa	amplification of endogenous target gene and knock-in
JB966_Mhc_geno_R	CTCTTGCTCCATGACGAACA	amplification of endogenous target gene
JB967_hh_geno_F	TCGTACTCGCACTCGAACAC	amplification of endogenous target gene and knock-in
JB1388_hh_geno_6F	aaatcaaagctggaccaaatc	amplification of endogenous target gene and knock-in, used with hh #1 insertion
JB968_hh_geno_R	GTTGTAGTTGGGCACGAGGT	amplification of endogenous target gene
JB900_T2A_R	cggggttctcctccacat	reverse primer for amplifying T2A-Gal4 insert in sense orientation
JB958_T2A-Gal4_3'_F	gttttcccagtcacgacgtt	reverse primer for amplifying T2A-Gal4 insert in antisense orientation
JB659_3P3dsred_seq1F	ACTCCAAGCTGGACATCACC	to amplify 3xP3-dsred as a control
JB660_3P3dsred_seq1R	CGAGGGTTCGAAATCGATAA	to amplify 3xP3-dsred as a control
JB713_Rp49_F	ATCGGTTACGGATCGAACAA	to amplify endogenous Rp49 gene as a control
JB714_Rp49_R	GACAATCTCCTTGCGCTTCT	to amplify endogenous Rp49 gene as a control
JB1051_alphaTub84B_C-term_F	CCTTCGTCCACTGGTACGTT	amplification of endogenous target gene and knock-in
JB1052_Act5c_C-term_F	CGTCGACCATGAAGATCAAG	amplification of endogenous target gene and knock-in
JB1053_His2Av_C-term_F	CTCCTCGCCACTTACAGCTC	amplification of endogenous target gene and knock-in
JB1054_Lam_C-term_F	GCCGACAACACTAGGACGAT	amplification of endogenous target gene and knock-in
JB1050_mNeonGreen_R	GGGAGAGAGGCGTTATCCTC	reverse primer for amplifying gene-mNeonGreen insert
JB1120_act5c_seqnextgenR	CGACTTCTCCTCCTCCTCCT	amplification of endogenous target gene
JB1121_His2av_seqnextgenR	TCGTCGGTGTTTTAGCTTGTC	amplification of endogenous target gene
JB1122_alphatub_seqnexgenR	GCGATTGGAAGCGTAAACAC	amplification of endogenous target gene
JB1123_Lam_seqnextgenR	GTGTTGTGCTGCGTTTGATT	amplification of endogenous target gene
JB1192_Lam_2F	GCGGCTAATCAACGAGAAAG	amplification of non-knock-in Lamin gene
JB1193_Lam_2R	TCTGTTGTCAGGAGCGTTTG	amplification of non-knock-in Lamin gene
JB1194_Lam_3F	ACGAGGAGCAGATGCAGATT	amplification of non-knock-in Lamin gene
JB1195_Lam_3R	GGTCTAAACCGGGAGAAAGC	amplification of non-knock-in Lamin gene
JB1196_Lam_4F	AGCTGCAGAACCTGAACGAT	amplification of non-knock-in Lamin gene
JB1197_Lam_4R	ACTAGCCGAACCCAGGATTT	amplification of non-knock-in Lamin gene
JB1198_Lam_5F	cccattacaagcgacgattt	amplification of non-knock-in Lamin gene
JB1199_Lam_5R	GAACAGCTCCACTCCTCCAG	amplification of non-knock-in Lamin gene
JB1200_Lam_6F	GTCTCGGTCTGCTCCTCATC	amplification of non-knock-in Lamin gene
JB1201_Lam_6R	GAATGGCATAGCCACCACTT	amplification of non-knock-in Lamin gene
JB1202_Lam_7F	CATTTGCAAGATGGTGGTTG	amplification of non-knock-in Lamin gene
JB1203_Lam_7R	ccaattaggccaacactgct	amplification of non-knock-in Lamin gene
JB1050_mNeonGreen_R	GGGAGAGAGGCGTTATCCTC	amplification and sequencing primer, 5' end of mNeonGreen

Supplemental Table 6 – Oligo and dsDNA sequences