An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms

Oguz Kanca[1,2], Jonathan Zirin[3], Jorge Garcia-Marques[4], Shannon Knight[3], Donghui Yang-Zhou[3], Gabriel Amador[3], Hyunglok Chung[1,2], Zhongyuan Zuo[1,2], Liwen Ma[1], Yuchun He[1,5], Wen-Wen Lin[1], Ying Fang[1], Ming Ge[1], Shinya Yamamoto[1,2,6,7], Karen L. Schulze [1,5] , Yanhui Hu [3], Allan C. Spradling[8], Stephanie E. Mohr[3], Norbert Perrimon[3], Hugo J. Bellen[1,2,5,6,7] *

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX USA
[2]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX USA
[3]Howard Hughes Medical Institute and Department of Genetics, Harvard Medical School, Boston, MA USA
[4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA USA
[5]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX USA
[6]Program in Developmental Biology, Baylor College of Medicine, Houston, TX USA
[7]Department of Neuroscience, Baylor College of Medicine, Houston, TX USA
[8]Department of Embryology, Howard Hughes Medical Institute, Carnegie Institution for Science, Baltimore, MD USA

Keywords: Cas9, CRISPR, ssDNA donors, short homology arms, S2 cells

*Hugo J. Bellen: Author for correspondence: hbellen@bcm.edu

1

Abstract

We previously reported a CRISPR-mediated knock-in strategy into introns of *Drosophila* genes, generating an *attP-FRT-SA-T2A-GAL4-polyA-3XP3-EGFP-FRT-attP* transgenic library for multiple uses (Lee et al., 2018b). The method relied on double stranded DNA (dsDNA) homology donors with ~1 kb homology arms. Here, we describe three new simpler ways to edit genes in flies. We create single stranded DNA (ssDNA) donors using PCR and add 100 nt of homology on each side of an integration cassette, followed by enzymatic removal of one strand. Using this method, we generated GFP-tagged proteins that mark organelles in S2 cells. We then describe two dsDNA methods using cheap synthesized donors flanked by 100 nt homology arms and gRNA target sites cloned into a plasmid. Upon injection, donor DNA (1 to 5 kb) is released from the plasmid by Cas9. The cassette integrates efficiently and precisely *in vivo*. The approach is fast, cheap, and scalable.

Introduction

A main goal of the Drosophila Gene Disruption Project (GDP) is to create genetic tools that facilitate an integrated approach to analyze the function of each gene in detail. This involves assessment of the loss of function phenotype, identification of the cells that express the gene, determination of the subcellular protein localization, selective removal of the transcript or protein in any tissue, the ability to perform immunoprecipitation of the protein and its interacting proteins or DNA, rescue of the induced fly mutant phenotypes with fly or human cDNAs and assessment of the consequences of amino acid variants *in vivo*. These elegant and precise manipulations are made possible by the integration of a Swappable Integration Cassette (SIC) in the gene of interest (GOI) using transposon mediated integration [*Minos*-mediated Integration Cassette (MiMIC) (Venken et al., 2011; Nagarkar-Jaiswal et al., 2015a)] or homologous recombination mediated by CRISPR, a technique we named CRIMIC (CRISPR-mediated Integration Cassette) (Zhang et al., 2014; Diao et al., 2015; Lee et al., 2018b). A SIC is typically flanked with *attP* sites and can be replaced using Recombination Mediated Cassette Exchange (Bateman et al., 2006; Venken et al., 2011). The CRIMIC variety of SIC currently used by the GDP is an artificial exon consisting of *attP-FRT-SA-T2A-GAL4-polyA-3XP3-EGFP-FRT-attP* inserted in a coding intron (intron flanked by two coding exons) of the GOI (Lee et al., 2018b). This insert typically creates a severe loss of function allele and generates a GAL4 protein that is expressed in the target gene's spatial and temporal expression pattern (Diao et al., 2015; Gnerer et al., 2015; Lee et al., 2018b). The resulting GAL4 can then be used to drive a *UAS-nuclear localization signal (NLS)::mCherry* to determine which cells express the gene or a *UAS-membrane (CD8)::mCherry* to outline the cell projections (Brand and Perrimon, 1993; Shaner et al., 2004). Alternatively, a *UAS-GOI cDNA* can be used to test for rescue of the loss of function phenotype induced by the insertion cassette. This provides a means for rigorous quality assessment of the genetic reagent and, when combined with mutant and/or truncated forms of the *UAS-GOI cDNA,* facilitates structure-function analysis. In addition, a *UAS-human-homologue cDNA* of the GOI permits humanization of the flies and assessment of human variants (Bellen and

Yamamoto, 2015; Kanca et al., 2017; Şentürk and Bellen, 2018; Chao et al., 2017; Yoon et al., 2017).

The SIC can also be replaced by an artificial exon that consists of *SA-Linker-EGFP-FlAsH-StrepII-TEV-3xFlag-Linker-SD*, abbreviated *SA-GFP-SD,* which adds an integral Green Fluorescent Protein (GFP) and other tags to the gene product (Venken et al., 2011; Nagarkar-Jaiswal et al., 2015a). This tag does not disrupt protein function in 75% of cases examined and permits the determination of the subcellular protein localization (Venken et al., 2011; Nagarkar-Jaiswal et al., 2015a; Yoon et al., 2017; Lee et al., 2018b) as well as removal of the protein in any tissue using specific GAL4 drivers (Jenett et al., 2012) to drive a DeGradFP protein that leads to polyubiquitination and degradation of the protein of interest (Caussinus et al., 2011; Nagarkar-Jaiswal et al., 2015a; Lee et al., 2018a). The GFP tag can also be used as an epitope for immunoprecipitation to determine interaction partners of the tagged protein (Neumüller et al., 2012; Zhang et al., 2013; David-Morrison et al., 2016; Yoon et al., 2017).

Additionally, SICs can be replaced by other RMCE vectors that enable integration of additional binary or tertiary systems (*e.g. SA-T2A-LexA, SA-T2A-split GAL4;* Gnerer et al., 2015; Diao et al., 2015) to obtain finer tools to express transgenes in specific cell populations. SICs can also be used to generate conditional alleles of targeted genes (Flip-flop and FLPstop (Fisher et al., 2017; Nagarkar-Jaiswal et al., 2017). Finally, strategies have been developed to convert SICs through genetic crosses rather than by injection (*Trojan Exons* and *Double Header;* Nagarkar-Jaiswal et al., 2015b; Diao et al., 2015; Li-Kroeger et al., 2018).

Although the above reagents form a powerful toolset, the generation of libraries of many thousands of genes based on these methods is labor-intensive and costly. The cost for reagents and labor for the generation of a single CRIMIC fly line is $1,000-2,000. Indeed, to create each CRIMIC construct we need to amplify two 1 kb homology arms, clone these arms on either site of a SIC in a plasmid, sequence verify the constructs, amplify and inject the DNA with a target-specific gRNA into 600 embryos, screen to obtain several independent transgenic flies, establish several fly stocks for each construct, PCR-verify the insertions, and cross each line with *UAS-mCherry* to determine expression patterns. Since the production of each transgenic line involves multiple steps, low failure rates at each step accumulate and decrease the overall success rate to ~50%. Given that we are in the process of tagging ~5,000 genes that

contain suitable introns, it is highly desirable to develop a more efficient, less labor-intensive, and cheaper alternative. One of the main bottlenecks is the production of large (5 kb) SIC homology donor plasmids containing a visible dominant marker and flanked by two ~1kb homology arms to promote homologous recombination (Beumer et al., 2008; 2013; Bier et al., 2018, Lee et al., 2018b)**.** We therefore explored a series of alternative strategies to reduce the construct size and facilitate cloning.

Here, we report the development of methods, using either a PCR-generated, single stranded DNA donor (ssDNA drop-in) or a synthesized double stranded homology donor (dsDNA drop-in) that greatly simplify the generation of homology donor constructs and improve the transgenesis rate. We tested both methods *in vivo* in *Drosophila* and targeted the same 10 genes with 5 different constructs to assess transformation efficiency and accuracy of integration. We show that the ssDNA drop-in method works efficiently in *Drosophila* S2R+ cells for constructs that are less than 2 kb and we used this method to mark several cellular organelles with GFP tagged proteins. The dsDNA drop-in strategy is based on short homology arms flanking SICs of up to 5 kb. The success rate for tagging the tested genes was 70-80%. The dsDNA drop-in donor vector is easy and cheap to produce, transformation efficiency is high, and the insertions are precise. Hence, these changes significantly decrease the costs of generating a transgenic CRIMIC library and make the CRIMIC technique more accessible to others. We anticipate that the drop-in approaches will also be useful in other species.

Results and Discussion

ssDNA homology donors

dsDNA homology donors for insertion of large cassettes in *Drosophila* typically require stretches of 500 nt to 1 kb of homology to the target site on either side of the SIC (Rong and Golic, 2000; Beumer et al., 2013; Zhang et al., 2014; Diao et al., 2015; Bier et al., 2018; Lee et al., 2018b). The large size of the homology regions affects cloning efficiency of the donor constructs (Lee et al., 2018b). Single stranded homology donors typically rely on much shorter homology arms (50-100 nt) to successfully integrate short DNA segments in *Drosophila* (~200 nt) based on

homology-directed repair (Beumer et al., 2013; Gratz et al., 2013; Wissel et al., 2016; Bier et al., 2018). To facilitate and speed up the preparation of homology donor constructs, we decided to test ssDNA donors for CRISPR-mediated homologous recombination of donors that are 1 to 2 kb.

To produce ssDNA homology donors, we established a cloning-free method based on PCR (Higuchi and Ochman, 1989). We generated SIC containing PCR templates that can be amplified with an M13 universal primer-derived primer pair (26 nt forward primer and 24 nt reverse primer). Gene-specific 100 nt homology arms are incorporated into these primers as 5' overhangs (left homology arm to the forward primer and reverse complement of right homology arm to the reverse primer) (Figure 1A). One of the primers is phosphorylated at its 5' end. The resulting PCR product contains a 5' phosphorylated strand and a non-phosphorylated strand. When this PCR product is treated with Lambda Exonuclease, a 5'-to-3' nuclease with a preference for 5' phosphorylated DNA (Little, 1967; Mitsis and Kwagh, 1999), the 5'-phosphorylated strand is degraded, leaving the non-phosphorylated strand as a ssDNA homology donor. This PCR-based approach has several advantages. First, the homology arms are included in the ~125 nt primers as 5' overhangs such that a single PCR generates the complete donor without any cloning steps. Second, the same primers can be used to amplify many different SIC templates for the same GOI. Third, the priming sequences and overall construct length do not change between genes. Hence, the same protocol can be applied to create homology donors for different genes. Fourth, the PCR-based method does not require bacterial transformation, eliminating possible rearrangements associated with propagation of DNA in bacteria (Figure 1B). Finally, the risk of inserting PCR-induced mutations in the SIC is negligible when a proofreading polymerase is used.

We first tested the efficacy of ssDNA drop-in constructs as a substrate for homology-directed repair in *Drosophila* S2R+ cells stably transfected with Cas9 (S2R+-MT::Cas9; Viswanatha et al., 2018). Specifically, we were interested in generating a collection of *Drosophila* S2R+ cells in which different organelles are marked by a protein tagged with a superfolder GFP (sfGFP) in each cell line (Pédelacq et al., 2006). We therefore amplified a SIC consisting of *attP-SA-Linker-sfGFP-Linker-SD-attP* (1392 nt including homology arms) using 20 different gene-specific primer pairs (Figure 2A, Table S1). Each gene-specific ssDNA homology donor was electroporated into Cas9-positive cells along with a corresponding gene-

specific gRNA and then subjected to fluorescence-activated cell sorting FACS (Figure 2B).

The frequency of GFP+ cells was determined by FACS (Figure S1) to range between ~1-4%. GFP-positive cells were clonally isolated, cultured, and analyzed. We observed that for any given clone, all cells had the same subcellular GFP localization, indicating that they were derived from a single cell and that the insertion was stably integrated. For 19 out of 20 genes targeted we observed GFP signal by FACS (Table 1). For 12 of 19 genes, we could establish GFP+ clones, verify correct integration by PCR, and determine subcellular localization by immunostaining (Table 1, Figure 2C, Figure S2). With the exception of Ref2P, all correct insertions of GFP resulted in fusion proteins with the expected subcellular distribution. The genes for which we observed GFP signal by FACS but could not successfully isolate GFP+ clones tended to be expressed at low levels in S2R+ cells based on modEncode expression profiling (Cherbas et al., 2011) (Table 1). For these genes it is possible that the signal-to-noise ratio for GFP was insufficient to robustly select GFP+ clones, leading to the loss of positive cells in the population.

Sequencing of the SIC insertion sites (Table 1) for the 12 cell lines showed that the insertions are precise. Western blotting of cell lysates confirmed that the inserted tags lead to GFP fusion proteins of the expected molecular weights (Figure S3). Given the dynamic localization of Polo protein during mitosis (Llamazares et al., 1991) we recorded the Polo subcellular localization pattern in live cells through mitosis. Time-lapse confocal imaging of Polo-GFP showed that the protein is localized to centrosomes, spindle, and midbody during cell division, in agreement with the data obtained by immunofluorescence (Llamazares et al., 1991) or using *polo-GFP* transgenes (Moutinho-Santos et al., 1999) (Supplemental movie 1).

Previously, PCR generated double-stranded constructs containing GFP and an antibiotic resistance gene have been used for homologous recombination in S2 cells (Bottcher et al., 2014; Kunzelmann et al., 2016). However, selection with a drug resistance gene was used to enrich the population and GFP integration frequency was about 2% as judged by FACS. We were able to generate GFP protein traps in S2 cells using long ssDNA in up to 4% transfected cells and obtain clones without using drug-based selection (Figure 2, Figure S2).

In summary, ssDNA drop-in constructs are simple to produce and provided an efficient homology substrate in S2R+ cells for about 60% of the tested genes. The

generated cell lines with marked organelles—endoplasmic reticulum (2), recycling endosomes, cis and trans Golgi, kinetochores, lysosomes (2), mitochondria, nucleoli and nuclear envelopes (Table 1)—are listed at the *Drosophila* RNAi Screening Center (DRSC) website (https://fgr.hms.harvard.edu/crispr-modified-cell-lines) and available through the Drosophila Genomics Resource Center (DGRC; https://dgrc.bio.indiana.edu). These cell lines will be a useful resource for small- and large-scale studies of organelle biogenesis, organelle function, and/or subcellular distribution of organelles or specific fusion proteins, and genetic and pharmacological screens to identify regulators of organelle homeostasis

Integration of ssDNA drop-in donors in the germline

We next generated a drop-in SIC for *in vivo Drosophila* transformation that can be amplified using PCR primers. Our initial experiments showed that the empirical size limit for ssDNA production was ~2 kb, significantly smaller than the ~5 kb SIC present in the CRIMIC cassettes currently used by the GDP (Lee et al., 2018b). A minimal SIC should contain *attP* sites to enable downstream RMCE applications and a dominant marker that allows detection of positive insertions in the targeted locus. The smallest self-sufficient visible dominant marker in *Drosophila* is *3XP3-EGFP* (Horn et al., 2000). Hence, in order to remain under the size limit, we generated an *attP-3XP3-EGFP-attP* PCR template that is 1242 nt long and can be used for ssDNA donor (1442 nt including two 100 nt homology arms) generation (Figure 3A). The *attP-3XP3-EGFP-attP* cassette is not mutagenic unless inserted in exons and only generates a landing site for RMCE in the targeted region.

To test the efficiency of this new ssDNA drop-in construct *in vivo*, we compared the efficiency of the ssDNA drop-in donor (with 100 nt homology arms) and the current CRIMIC dsDNA donor (*attP-FRT-SA-T2A-GAL4-polyA-3XP3-EGFP-FRT-attP* with ~1kb homology arms) for the same 10 genes using the same gRNAs by injecting ~400-600 embryos for each gene with each construct. Surviving adults were crossed with *y w* flies as single fly crosses and the numbers of independent targeting events were quantified. We targeted a set of 10 genes with each construct to have a large enough set to reduce the possibility of locus-specific properties skewing the results. We were successful in targeting four genes with both approaches and obtained insertions in *CG5009*, *CG9527, Khc,* and *NLaz* with the

8

CRIMIC approach and *CG5009, endoB, Cp1,* and *Lst* with the ssDNA drop-in construct. We found that the use of CRIMIC constructs was more efficient, as judged by the number of independent fly lines produced per successful gene (Figure 3B).

Short ssDNA homology donors were previously used in *Drosophila* to integrate small epitope tags or site-specific recombination sites (Gratz et al., 2013; Wissel et al., 2016). Due to size constraints, these constructs lacked a dominant marker and detection of successful events relied on labor-intensive PCR strategies. More recently, novel methods were developed to generate and use longer ssDNAs (~1000 nt) as homology donors *in vivo* in mice (Miura et al., 2015; Quadros et al., 2017; Lanza et al., 2018). Our results show that longer ssDNA constructs with visible dominant markers integrate in the fly genome *in vivo* albeit with lower efficacy than double stranded homology donors with large homology arms (i.e. CRIMIC donors).

In summary, the ssDNA constructs are efficient donors for S2R+ cells for protein tagging by knock-in but for fly transgenesis they are not more efficient than the standard CRIMIC method. Nevertheless, the ease and low cost of producing these constructs and the transformation efficiency in S2 cells may justify their use to integrate GFP tags or landing sites in cultured cells or the germline.

dsDNA drop-in donors of <2 kb are efficient homology donors for transgenesis

Given that the ssDNA drop-in constructs did not increase the success rate of fly transformation, we next attempted to optimize dsDNA homology donors for production ease and transformation efficiency. One means of increasing the targeting rate is to shorten the SIC, since homologous recombination is dependent on the size of the inserted cassette (Beumer et al., 2013). We used three strategies to shorten SICs (Figure 4A). First, we used a shortened GAL4 construct, referred to as miniGAL4 (1200 nt), which is about half the size of full-length GAL4 (2646 nt) and has about 50% of the transcriptional activity of full-length GAL4 in yeast (Ding and Johnston, 1997). Second, we used a very short dominant marker. Because *3XP3-EGFP* is ~1 kb in length, we opted to use an alternative marker that is only ~200 nt long and contains a ubiquitous U6 promotor expressing a gRNA sequence (gRNA1) that does not have a target in the fly genome. This strategy is based on the single-strand annealing pathway of DNA repair to reconstitute expression of a transgenic ubiquitous mCherry that is interrupted by the gRNA1 target sequence (Figure S4)

9

(Garcia Marques et al., 2019). Upon Cas9-mediated double strand break, the modified non-functional *mCherry* gene is repaired and becomes functional, providing a visible marker. Third, we shortened the polyA sequence from 135 nt to 35 nt *(minipolyA)* (McFarland et al., 2006). This resulted in *attP-SA-T2A-miniGAL4-minipolyA-U6gRNA1-attP*, which functions as a gene trap and is 1968 nt in length, including homology arms. We also generated a smaller minimal mutagenic construct, *attP-SA-3XSTOP-minipolyA-U6gRNA1-attP,* that is 715 nt in length including homology arms (Figure 4A). Both dsDNA drop-in constructs are small enough to be commercially synthesized at a low cost (less than $250).

To further improve the insertion efficiency, we decided to induce *in vivo* linearization of the plasmid constructs. Linearization has previously been shown to boost homologous recombination rates in cell culture, zebrafish and mouse transgenesis (Cristea et al., 2013; Hisano et al., 2015; Suzuki et al., 2016; Yao et al., 2017). Hence, use of short 100 nt left and right homology arms flanked by the gRNA1 target sites to linearize the construct *in vivo* upon injection may significantly increase the frequency of homologous recombination (Figure 4B).

It is worth noting that higher transgenesis rates have been reported in mice and worms using donor constructs linearized *in vitro* (Paix et al., 2014; 2015; 2017; Dokshin et al., 2018; Yao et al., 2018). However, *in vitro* linearized DNA or direct PCR products are very poor substrates for homologous recombination for germ line transformation in *Drosophila* (Beumer et al., 2008; Bottcher et al., 2014). We therefore opted to use dsDNA donors that are linearized *in vivo* by the Cas9 expressed in germ cells.

Upon injection of gene-specific gRNA coding plasmids and donor plasmids we isolated multiple transgenic lines for each of the 10 targeted genes. We obtained 65 independent transgenic lines for the *attP-SA-3XSTOP-minipolyA-U6gRNA1-attP* construct, and 119 independent transgenic lines for the *attP-SA-T2A-miniGAL4-minipolyA-U6gRNA1-attP* construct. We verified the insertion sites by genomic PCR in 7/10 genes for both methods. Analyses of the genomic DNA sequences of the three genes for which we could not verify the integration site revealed that these genes contain an unanticipated variation in the gRNA target site in the isogenized Cas9 injection stocks when compared to the FlyBase reference genome sequence (Santos et al., 2015; Goodman et al., 2018). This suggests that the precise

homologous recombination rate is dependent on efficient cutting of the target site and is likely higher than 70%. Our current gRNA design platform considers these strain variations which can be substantial in introns.

Because we shortened the *polyA* tail from 135 nt (Okada et al., 1999) to 35 nt (McFarland et al., 2006), we assessed the mutagenic potential of new constructs that contain *minipolyA*. We used Western blotting with an antibody that recognizes the gene product of *CG5009* (*Drosophila* ortholog of *ACOX1*) to compare protein levels in animals homozygous for the CRIMIC allele (with 135 nt polyA tail) with protein levels of animals homozygous for either of the two constructs that carry the *minipolyA*. As shown in Figure 5A, the artificial exon *attP-SA-3XSTOP-minipolyA-U6gRNA1-attP* leads to ~40% decrease in protein levels and creates a much less severe allele than the CRIMIC allele. However, the *attP-SA-T2A-miniGAL4-minipolyA-U6gRNA1-attP* cassette leads to ~80-90% decrease in protein levels, similar to a CRIMIC insertion in the same locus (Figure 5A).

The miniGAL4 construct had never been tested in flies. We therefore compared full length versus miniGAL4 induction of a *UAS-LacZ* reporter by Western blot. As shown in Figure 5B, we observed a ~60% reduction of reporter levels with T2A-miniGAL4 as compared to CRIMIC T2A-GAL4 driven *UAS-LacZ* expression for the *CG5009* locus (Figure 5B). Hence, the *attP-SA-T2A-miniGAL4-minipolyA-U6gRNA1-attP* drives lower levels of reporter expression when compared to CRIMIC, which may limit the use of miniGAL4 for genes expressed at low levels.

In summary, the decrease of SIC size and linearization increased the transgenesis rate. We used novel selection markers, smaller *polyA* tails, and new GAL4 variants to decrease the size of the integrated construct. However, this decrease in size came with trade-offs. The U6-gRNA1 dominant marker is very easy to detect and is smaller than previously established dominant markers. However, this marker requires the presence of other transgenes for detection (Figure S4) and the reconstituted dominant marker transgene segregates independently from the targeted gene. These limitations make establishing and maintaining stocks more challenging. Moreover, T2A-miniGAL4 may not be strong enough to drive robust reporter expression in genes with low expression levels.

Large double stranded drop-in dsDNA donors (~5 kb) are efficient homology donors to integrate CRIMIC-like SICs

11

To avoid the issues raised in the previous section, we designed a strategy to integrate full length CRIMIC cassettes with short homology arms. To achieve this, we tested whether large DNA fragments with short homology arms (100 nt) could be integrated in target genes upon linearization using the gRNA1 *in vivo*. This would allow the use of dominant markers like *3XP3-EGFP* as well as integration of the full length *GAL4* gene with an extended *polyA* tail, i.e. the CRIMIC cassette SIC. Unfortunately, synthesis of a full length 5 kb CRIMIC cassette would be cost prohibitive (>$1,000) especially for thousands of genes. Hence, we developed a modified cloning strategy in which we first synthesize a *gRNA1 target-100nt homology arm-Restriction cassette-100nt homology arm-gRNA1 target* inserted into a pUC57 vector for each target gene (cost is $80). The SIC containing *attP-FRT-T2A-GAL4-polyA-3XP3-EGFP-FRT-attP* is then subcloned directionally into this plasmid in a single straightforward cloning step, replacing the restriction cassette with the SIC of interest (Figure 6A). We refer to these constructs *drop-in int100-CRIMIC* constructs. The SIC can be replaced by any other effector (e.g. *SA-GFP-SD*) to generate other homology donor constructs.

We injected vectors containing the full length CRIMIC cassette for seven of the genes in which we previously successfully inserted dsDNA drop-in cassettes. Using the *drop-in int100 CRIMIC* cassette we obtained multiple knock-in alleles in five genes, as verified by PCR (Figure 6B). Hence, 100 nt homology arms are sufficient to integrate large SICs into target sites. In addition, increasing homology arm length to 200 nt should not increase synthesis costs, as the total length of the construct remains less than 500 nt. Whether use of 200 nt may improve efficiency remains to be tested.

To ensure functionality of knock-in alleles, we compared expression patterns of drop-in int100 CRIMIC with the corresponding CRIMIC for *CG5009* (Figure 6C) as well as drop-in int100 CRIMIC with the *T2A-miniGAL4* for *Khc* (Figure S5) by crossing the flies to *UAS-NLS::mCherry* reporter lines. In both cases the constructs lead to mCherry expression in very similar patterns.

In vivo linearization was previously shown to lead to knock-ins in zebrafish and cell culture even when there is no homology arm, provided that the homology donor and target site are cut concomitantly (Cristea et al., 2013; Auer et al., 2014; Schmid-Burgk et al., 2016; Suzuki et al., 2016). This process is more prone to

generating small deletions in the target site and the insertions are non-directional. Recently, a homology independent knock-in method was established for *Drosophila* cell culture and germline transformation (Bosch et al., 2019). By simultaneously cutting the donor construct and target region in the absence of homology arms, Bosch et al. (2019) integrated a CRIMIC cassette in 4 of 11 genes targeted. However, these insertion events were error-prone and non-directional. Note that for drop-in constructs, the presence of short homology arms allows the donor vectors to be synthesized cheaply and introducing larger SICs is accomplished by a single straightforward cloning step, providing a good balance between ease of construct generation and efficient *in vivo* use.

In summary, we have developed efficient pipelines for CRISPR knock-in using ssDNA in *Drosophila* cells or a dsDNA approach in the germline. With respect to the dsDNA donors, we significantly improved the overall efficiency compared to the method described in Lee et al. (2018b). Key features that improved efficiency are: 1) shortening the homology arms to 100 nt, which allows synthesis followed by a simple cloning step, thus eliminating cloning failure and reducing the cost from ~$300 to ~$100 per construct; and 2) integration of target gRNA1 sites on either side of the SIC to linearize the donor with Cas9 in the germ cells. In conclusion, the methods we describe are efficient, simple, and precise.

Materials and Methods

Generation of templates for ssDNA production

The sequences of scaffold vectors can be found in Supplementary Materials and Methods. Briefly, to generate the *attP-3XP3-EGFP-attP,* the scaffold vector 3XP3-EGFP cassette was amplified by PCR using the long primers 3XP3-EGFP-RI-for and 3XP3-EGFPrev-NotI and cloned between the *Eco*RI and *Not*I sites in pCasper3 (Thummel and Pirrotta, 1992) . To generate *attP-SA-sfGFP-SD-attP*, a scaffold vector that we named pScaffold was produced by integrating annealed oligonucleotides with sequences M13For-attPfor-SbfI-AvrII-attPrev-M13rev in pCasper3 backbone in *Eco*RI-*Not*I sites. SA-sfGFP-SD was cloned as a three-fragment ligation in pScaffold with linker-SA (amplified from pDoubleHeader (Li-Kroeger et al., 2018) with primers SA-for-Sbf and Linker-SA-rev_BamHI), sfGFP

(amplified from pUAST-NLS-sfGFP-3XMyc-PEST with primers sfGFP-for_BamHI and sfGFP-rev_KpnI), and Linker-SD (amplified from pDoubleHeader with primers Linker-SD-for_KpnI and Linker-SD-rev_NotI).

Production of ssDNA drop-in constructs

Gene-specific homology donors were produced by incorporating the homologous sites (regions spanning 100 nt upstream of 3 nt prior to the PAM (protospacer adjacent motif) for left homology arm or 97 nt downstream of that region for the right homology arm (Supplementary materials and methods)) as 5' overhangs to primers with complementarity to the template (26 bases for the forward primer and 24 bases for the reverse primer). The long primers were ordered from IDT (Coralville, Iowa) as Ultramers®. The reverse primer was 5' phosphorylated. Four 50 μl PCR reactions were set up with Q5 hot start high fidelity 2X master mix (NEB #M0494L). PCR conditions were optimized using gradient PCR (Bio-Rad C1000 Touch). The optimal annealing temperature was 70°C. The elongation time was 1 minute for *attP-3XP3-EGFP-attP* and *attP-SA-sfGFP-SD-attP.* Resulting PCR amplicons were pooled in two samples and isolated using Qiaquick spin columns (Qiagen #28106) following the manufacturer's protocol. Samples were eluted in 2 X 50μl elution buffer from the kit. Two lambda exonuclease (NEB #M0262L) digestion reactions of 100 μl were set up using 4-6 μg DNA and 8μl enzyme each. Digestion conditions were 37°C for 1 hour, followed by 10 minutes at 75°C for heat inactivation. Reaction products were pooled in two samples and ssDNA isolated using NEB Monarch DNA purification kit following the manufacturer's instructions (NEB #T1030L). Samples were isolated in 2 X 10μl prewarmed (55°C) elution buffer from the kit and the DNA concentration was measured using NanoDrop One (ThermoFisher Scientific).

Generation of GFP Knock-In Cell Lines

Cell Culture and Regular Media

Drosophila cells stably expressing Cas9 (S2R+-MT::Cas9; Drosophila Genomics Resource Center cell stock #268; Viswanatha et al. 2018) were cultured in Schneider's Drosophila Medium 1X (ThermoFisher Scientific #21720024) with 10% FBS and 1% penicillin/streptomycin (referred to as regular media). We note that this

cell line is a derivative of S2R+ NPT005 (DGRC #229) and thus contains an mCherry tag in the *Clic* locus (Neumüller et al., 2012).

Conditioned Media

Conditioned media were prepared as previously described by Housden et al. (2017).

Cellular Transfection

Electroporation of S2R+-MT::Cas9 cells was performed using a Lonza 4D Nucleofector (Lonza #AAF-1002B) following the manufacturer's protocol. For each transfection, 1 µg of sgRNA (100 ng) and 2 µL of the sfGFP donor (100 ng) were used with 4×10^5 sub-confluent S2R+-MT::Cas9 cells. See Table S1 for all sgRNA and insert sequences. After electroporation, cells were immediately placed in regular media and health was monitored. Cell cultures were then maintained and expanded for FACS.

Isolation of Single-Cell Clones

Prior to FACS, 10^6 cells were filtered through a 40 µM Falcon Cell Strainer (Corning #21008-949) into 15 mL conical tubes (ThermoFisher Scientific #14-959-70C). Single cells were then FACS-isolated on a BD Aria IIu, based on the presence of GFP with an expression level greater than 2×10^2 (see Figure S1). Single cells were sorted into wells of a 96 well plate (VWR #29444-010) filled with 100 µL of conditioned medium. Cells were observed ~14 days later and any viable clones were expanded.

Analysis of GFP Knock-In cell lines

Image Analysis

Multiple clonal cell lines were generated per gene and image analysis was used to measure the expression of the GFP knock-in marker. Several images of each live clone were taken using an InCell Analyzer 6000 automated confocal fluorescence microscope (GE Healthcare Lifesciences) using the dsRed channel to detect mCherry fluorescence present in all cells and the FITC channel to detect GFP fluorescence. Images were analyzed using CellProfiler (version 2.1.1). As mCherry is present throughout each cell, this image was used by CellProfiler to determine the

15

outline of individual cells using the command "Identify Primary Objects." The outlines generated from this step were then applied to the GFP image through the process "Identify Secondary Objects." GFP fluorescence intensity was then measured using the command "Measure Object Intensity," which averages the GFP fluorescence within all of the cellular outlines (see Supplementary Figure S2). Mean GFP intensities for each clone were noted, and the three clones per gene with the highest means were chosen for molecular analysis. All other clones were pooled and stored for use in a subsequent FACS if needed. Image data were managed using OMERO, as supported by the Harvard Medical School Image Data Management Core.

Genomic DNA Extraction

Genomic DNA was extracted from individual clones using a Quick-DNA MiniPrep Kit (Zymo Research #D3024) according to the manufacturer's protocol. Concentrations were measured using a NanoDrop 8000 Spectrophotometer (ThermoFisher Scientific #ND-8000-GL).

Molecular Analysis

Due to the large size of the insertion, the 5' and 3' integration sites of each insertion were analyzed by PCR amplifying and sequencing a fragment spanning the junction of each end of the insertion with flanking DNA (see Table S1 for a list the two primer sets per gene). These flanking sites were PCR amplified with High Fidelity Phusion Polymerase (NEB #M0530) using the following program: 1) 98ºC for 30 sec, 2) 98ºC for 10 sec, 3) 56ºC for 30 sec, 4) 72ºC for 30 sec (35 cycles), 5) 72ºC for 10 min, 6) 4ºC hold. PCR products were excised from a 2% TAE agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen #28704) and Sanger sequenced at the Dana-Farber/Harvard Cancer Center DNA Resource Core.

Immunostaining

Cells were fixed in 4% paraformaldehyde in phosphate-buffered saline with 0.1% Triton X-100 for 30 min. A standard staining protocol was used. Primary antibodies were used as indicated: Chicken anti-GFP (1:1000, Abcam ab13970), Mouse anti-Lamin (1:500, DSHB #ADL84.12), Mouse anti-Calnexin99A (1:5, DSHB #Cnx99A 6-2-1), Mouse anti-ATP5A (1:100, Abcam ab14748), Rabbit anti-Arl8 (1:500, DSHB

16

Arl8), Goat anti-GMAP (1:2000, DSHB, #GMAP), Goat anti-Golgin245 (1:2000, DSHB, #Golgin245), Rabbit anti-Ref2P (1:500, Abcam #ab178440), Rabbit anti-Myc-tag (1:1000, Cell Signaling Technology #2278S). Secondary antibodies with Alexa Fluor conjugates and DAPI (Molecular probes #D-1306) were used at 1:1000. Images were obtained using a GE IN Cell 6000 automated confocal microscope with a 60x objective. Time-lapse videos were generated by imaging every 30 sec over a 2 hr period.

Western Blotting

Clones that are generated using phase 1 template contained a 5 bp deletion in the splice donor portion of the insert. To determine if full-length fusion proteins were generated, we analyzed the following: S2R+-MT::Cas9 (unmodified control), Actin::GFP (a GFP-positive control), *Rab11* GFP knock-in (no deletion in the insert control), *spin* GFP knock-in (5 bp deletion), two *Lam* GFP knock-in clones (5 bp deletion in the first, 206 bp deletion in the second), *Ref2p* GFP knock-in (5 bp deletion), and *Cnx99a* GFP knock-in (5 bp deletion). Cells were spun down and lysed by resuspension in Pierce RIPA Buffer (ThermoFisher Scientific #89901) and Halt Protease Inhibitor Cocktail, EDTA-Free (ThermoFisher Scientific #87785). Cells were agitated for 30 minutes at 4ºC and then centrifuged at 15,000 rpm for 20 minutes at 4ºC. The lysates (supernatant) were removed and held at -20ºC. Protein concentrations were determined using a Pierce BCA Protein Assay Kit (ThermoFisher Scientific #23227) according to the manufacturer's protocol. Appropriate volumes of each lysate were added to 4X Laemmli Sample Buffer (BioRad #1610747), vortexed, held at 100ºC for 10 minutes, and then spun down at 13,000 rpm for 3 minutes before loading into a Mini-PROTEAN TGX Precast Gel (BioRad #4561095) and running at 100 V for 1 hour. The gel was then transferred onto a PVDF membrane (BioRad #1620177) using the Trans-Blot Turbo Transfer System (BioRad #1704150). After blocking with 5% blocking solution and washing in TBST, the membrane was incubated in the primary antibody rabbit anti-GFP (1:5000; Molecular Probes #A6455) shaking at 4ºC overnight. The membrane was then washed four times with TBST and incubated with Donkey anti-rabbit HRP (1:3000; GE Healthcare #NA934) for 1 hour at room temperature, washed with TBST, and prepared for imaging using the SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific #34580) according to

17

manufacturer's protocol. The blot was imaged using a ChemiDoc MP Imaging System (see Supplementary Figure 3 top panel; BioRad #17001402). The blot was then stripped with Restore PLUS Western Blot Stripping Buffer (ThermoFisher Scientific #46430), reprobed with mouse anti-tubulin (1:2000; SigmaT #5168) and Sheep anti-mouse HRP (1:3000; GE Healthcare #NXA931), and re-imaged (see Supplementary Figure # middle panel). The blot was then stripped again and reprobed with mouse anti-lamin (1:500; DSHB #ADL84.12) and Sheep anti-mouse Horse Radish Peroxidase (HRP) (1:3000; GE Healthcare #NXA931), and re-imaged (see Supplementary Figure # bottom panel).

For western analysis in adult flies, flies were dissected and lysed in 0.1% CHAPS buffer [50mM NaCl, 200mM HEPES, 1mM EDTA and protease inhibitor cocktail (Roche)] Tissue or cell debris were removed by centrifugation. Isolated lysates were subjected to electrophoresis using a 4%–12% gradient SDS-PAGE gel and transferred to Immobilon-FL polyvinylidene difluoride membranes. Loading input was adjusted for protein concentration. Primary antibodies used were as follows: Rabbit anti-ACOX1 (1:1000; HPA021195, Sigma), Rabbit anti-β-galactosidase (1:1000; MP Biomedicals #55976), and Mouse anti-Actin-c4 (1:5000; Millipore Sigma #MAB1501,). Secondary antibodies include Jackson ImmunoResearch HRP conjugated (1:5000). Blots were imaged on a Bio-Rad ChemiDocMP. The intensity of each band was measured and normalized to a loading control using Imagelab software (Bio-RAD). 3 biological repeats were performed and ordinary one way ANOVA was used to compare expression levels of ACOX1 in different conditions. Two technical repeats were performed for β-galactosidase measurements.

Fly injections

ssDNA constructs were injected at 50-100 ng/µl concentration with 25ng/µl gene specific gRNA encoding pCFD3 vector (Port et al., 2014). *attP-SA-3XStop-minipolyA-U6gRNA1-attP* and *attP-SA-T2A-miniGAL4-minipolyA-U6gRNA1-att* dsDNA drop-in constructs were injected at ~150ng/µl concentration together with 25ng/µl gene specific gRNA. dsDNA drop-in int100-CRIMIC constructs were injected at 300-400 ng/µl along with 25ng/µl gene specific gRNA and 25 ng/ µl pCFD3-gRNA1. Injections were performed as described in (Lee et al., 2018b). 400-600 *yw; iso; attP2(y+){nos-Cas9(v+)}* embryos per genotype were injected. Resulting G0 males and females were crossed to *yw* flies as single fly crosses for *3XP3-EGFP*

detection and with *actin5C-Cas9; actin5C-GF-gRNA2-FP; actin5C-mcherr-#1-ry* flies for gRNA1 detection (Figure S4B) (Garcia Marques et al., 2019). Up to 5 independent lines were generated per construct per gene. *actin5C-GF-#2-FP* is an internal control that would detect non-specific activation of dominant markers. We have not detected GFP in any of the screened flies, showing the specificity of dominant marker detection through gRNA1.

PCR validation

PCR primers that flank the integration site were designed for each targeted gene (Table S1 for primer sequences). These primers were used in combination with insert-specific primers that bind 5' of the inserted cassette in reverse orientation and 3' of the insert in forward orientation (pointing outwards from the insert cassette). 200-800 nt amplicons were amplified from genomic DNA from individual insertion lines through single fly PCR (Gloor et al., 1993) using OneTaq PCR master mix (NEB #M0271L). PCR conditions were denaturation at 95°C for 30 seconds, 95°C 30 seconds, 58°C 30 seconds, 68°C 1 minute for 34 cycles and 68°C 5 minutes.

dsDNA drop-in constructs production

Templates for ordering the dsDNA drop-in constructs can be found in Supplementary Materials and Methods. dsDNA drop-in constructs were ordered for production from Genewiz ("ValueGene" option) in pUC57 Kan vector backbone at 4 μg production scale. When lyophilized samples arrived from production, samples were resuspended in 25 μl of ddH$_2$O at 55°C for 30 minutes. 19 μl was mixed with 1 μl gene-specific gRNA plasmid (25 ng/ul final concentration of gRNA plasmid). The rest was stored at -20°C for back-up purposes.

Confocal imaging of transgenic larval brains

Dissection and imaging were performed following the protocols in (Lee et al., 2018b). In brief, fluorescence-positive 3rd instar larvae were collected in 1x PBS solution and then cut in half and inverted to expose the brain. Brains were transferred into 1.5mL centrifuge tubes and fixed in 4% PFA in 1xPBS buffer for 20 minutes. Brains were then washed for 10 minutes three times in 0.2% PBST. Finally, samples were

mounted on glass slides with 8µL of VectaShield (VectorLabs #H-1000) and imaged at 20x zoom with a Nikon W1 dual laser spinning-disc confocal microscope.

Figure legends

Figure 1. ssDNA pipeline is faster than cloning CRIMIC constructs. (A) Schematic of PCR-based generation of drop-in ssDNA constructs. Gray boxes, UTRs; orange boxes, coding exons, yellow line, coding introns, black line, outside coding introns and exons (B) Comparison of donor generation pipelines for CRIMIC and PCR-based drop-in ssDNA homology donors. Making ssDNA donors is about 5X times faster than making CRIMIC donors.

Figure 2. ssDNA homology donors are effective in S2 cells to tag organelles. (A) Schematic of drop-in cassette encoding for sfGFP artificial exon. Size of the construct including the homology arms is indicated in the right. sfGFP: superfolderGFP; SA: Splice Acceptor of *mhc*; SD: Splice Donor of *mhc*; L: flexible linker that consists of four copies of Gly-Gly-Ser. (B) Diagram of steps to isolate cell clones resulting from successful homologous recombination events. (C) Examples of S2R+ cells with organelles marked with GFP. Left panel, antibody staining; middle panel, GFP signal; right panel, the merge.

Figure 3. ssDNA constructs are not as efficient as double stranded CRIMIC constructs for fly transformation. (A) Schematic of drop-in cassette used for fly transformation. Size of the construct including homology arms is indicated on the right. (B) Injection results for the 10 genes selected for comparison of transformation efficacy. Numbers indicate positive events / fertile G0 single fly crosses. Red, no positive events; light green, positive non-confirmed events; dark green, genes with PCR-confirmed events.

Figure 4. Double stranded DNA synthetic constructs are efficient for fly transformation. (A) Schematic of synthesized plasmid drop-in donors. mPA indicates *minipolyA*. (B) Schematic of the donor plasmid followed by linearization by Cas9 in germ cells and integration of donors *in vivo*. gRNA[target] is gene specific gRNA. (C) Injection results for the 10 genes selected for comparison of

20

transformation efficacy. Numbers indicate positive events / fertile G0 single fly crosses. Red, no positive events; light green, positive non-confirmed events; and dark green, genes with PCR-confirmed events.

Figure 5. T2A-miniGAL4 gene trap cassette is mutagenic and expresses an active GAL4. (A) Western blot and quantification of the level of ACOX1 protein in flies homozygous for *SA-T2A-miniGAL4-minipolyA*, *SA-3XSTOP-minipolyA* or CRIMIC construct for *CG5009*. **** P < 0.0001, * P < 0.01. (B) Western blot of β-galactosidase and quantification of heterozygous flies carrying a copy of the miniGAL4 construct compared to heterozygous flies carrying a CRIMIC.

Figure 6. Single step cloning method allows efficient insertion of the CRIMIC cassette in coding introns. (A) Schematic of a single step cloning vector pUC57. LHA Left Homology Arm, RE1 Restriction Enzyme 1, RE2 restriction Enzyme 2, RHA Right Homology Arm. (B) Injection results for the 7 genes selected to estimate transformation efficacy. Numbers indicate positive events / G0 single fly crosses. (C) Third instar larval brain expression domain of *CG5009* as determined by crossing conventional CRIMIC or drop-in int100-CRIMIC flies to *UAS-NLS-mCherry* reporter lines. Scale bar is 100μm.

Table 1. Summary of ssDNA drop-in mediated homologous recombination in S2R+ cells

Supplementary Figure 1. **FACS data of control cells (left) and ssDNA knock-in cells (right).** All cell lines express mCherry::Clic, which is present in the parental cell line and thus in these derivatives. Single cell clones with GFP expression levels greater than 2×10^2 were retained.

Supplementary Figure 2. Detection of subcellular localization of GFP tagged proteins in S2 cells. Left panels, confocal fluorescence detection of GFP fusion proteins in single-cell isolated clones. The specific proteins tagged by GFP knock-in are indicated. Center panels, confocal fluorescence detection of mCherry::Clic,

21

which is present in the parental cell line and thus in these derivatives. Right panels, merged image.

Supplementary Figure 3. Western blot analysis of tagged proteins observed in S2R+ cells. Western blot analysis of selected clones using anti-GFP antibody shows expected protein sizes.

Supplementary Figure 4. Schematic and crossing scheme for CRISPR gRNA-based dominant marker strategy. (A) Schematics of the action of gRNA1-based detection of dominant marker. (B) Crossing scheme for screening the transgenics using *U6gRNA1* as dominant marker. The crosses for a second chromosome insertion are shown; for insertions on other chromosomes, an appropriate balancer stock would be used in the second and third crosses.

Supplementary Figure 5. Comparison of expression domain obtained by T2A-miniGAL4 and drop-in int100-CRIMIC. Third instar larval brain expression domain of *Khc* as determined by crossing conventional T2A-miniGAL4 or drop-in int100-CRIMIC flies to *UAS-NLS-mCherry* reporter lines. Scale bars are 100µm.

Table S1. List of oligos and ultramers used in this study.

Acknowledgements

22

References

Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res. 2014 Jan 1;24(1):142–53.

Bateman JR, Lee AM, Wu C-T. Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics. Genetics; 2006 Jun;173(2):769–77. PMCID: PMC1526508

Bellen HJ, Yamamoto S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. Cell. 2015 Sep 24;163(1):12–4. PMCID: PMC4783153

Beumer KJ, Trautman JK, Bozas A, Liu J-L, Rutter J, Gall JG, et al. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. U.S.A. National Academy of Sciences; 2008 Dec 16;105(50):19821–6. PMCID: PMC2604940

Beumer KJ, Trautman JK, Mukherjee K, Carroll D. Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. G3 (Bethesda). G3: Genes, Genomes, Genetics; 2013 Apr 9;3(4):657–64. PMCID: PMC3618352

Bier E, Harrison MM, O'Connor-Giles KM, Wildonger J. Advances in Engineering the Fly Genome with the CRISPR-Cas System. Genetics. Genetics; 2018 Jan;208(1):1–18. PMCID: PMC5753851

Bottcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, et al. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res. 2014 Jun 19;42(11):e89–9.

Bosch, J. A., Colbeth, R., Zirin, J., & Perrimon, N. (2019). Gene knock-ins in Drosophila using homology-independent insertion of universal donor plasmids, 16, 29923–47. http://doi.org/10.1101/639484

Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993 Jun;118(2):401–15.

Caussinus E, Kanca O, Affolter M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat. Struct. Mol. Biol. Nature Publishing Group; 2011 Dec 11;19(1):117–21.

Chao, H.-T., Davids, M., Burke, E., Pappas, J. G., Rosenfeld, J. A., McCarty, A. J., et al. (2017). A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. American Journal of Human Genetics, 100(1), 128–137. http://doi.org/10.1016/j.ajhg.2016.11.018

Cherbas L, Willingham A, Zhang D, Yang L, Zou Y, Eads BD, et al. The transcriptional diversity of 25 Drosophila cell lines. Genome Res. 2011 Feb 1;21(2):301–14.

Cristea S, Freyvert Y, Santiago Y, Holmes MC, Urnov FD, Gregory PD, et al. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol. Bioeng. John Wiley & Sons, Ltd; 2013 Mar;110(3):871–80.

David-Morrison G, Xu Z, Rui Y-N, Charng W-L, Jaiswal M, Yamamoto S, et al. WAC Regulates mTOR Activity by Acting as an Adaptor for the TTT and Pontin/Reptin Complexes. Dev. Cell. 2016 Jan 25;36(2):139–51. PMCID: PMC4730548

Diao F, Ironfield H, Luan H, Diao F, Shropshire WC, Ewer J, et al. Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. Cell Rep. 2015 Mar 3;10(8):1410–21. PMCID: PMC4373654

Ding WV, Johnston SA. The DNA binding and activation domains of Gal4p are sufficient for conveying its regulatory signals. Mol. Cell. Biol. American Society for Microbiology (ASM); 1997 May;17(5):2538–49. PMCID: PMC232103

Dokshin GA, Ghanta KS, Piscopo KM, Mello CC. Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in Caenorhabditis elegans. Genetics. 2018 Nov 6;210(3):781–7.

Fisher YE, Yang HH, Isaacman-Beck J, Xie M, Gohl DM, Clandinin TR. FlpStop, a tool for conditional gene control in Drosophila. Elife. eLife Sciences Publications Limited; 2017 Feb 17;6:796. PMCID: PMC5342825

Garcia Marques J, Yang C-P, Espinosa-Medina I, Mok K, Koyama M, Lee T. Unlimited Genetic Switches for Cell-Type-Specific Manipulation. Neuron. Elsevier Inc; 2019 Aug 9;:1–20.

Gloor, G. B., Preston, C. R., Johnson-Schlitz, D. M., Nassif, N. A., Phillis, R. W., Benz, W. K., et al. (1993). Type I repressors of P element mobility. *Genetics*, *135*(1), 81–95.

Gnerer JP, Venken KJT, Dierick HA. Gene-specific cell labeling using MiMIC transposons. Nucleic Acids Res. 2015 Apr 30;43(8):e56. PMCID: PMC4417149

Goodman JL, Strelets VB, Marygold SJ, Matthews BB, Millburn G, Trovisco V, et al. FlyBase 2.0: the next generation. Nucleic Acids Res. Oxford University Press; 2018 Oct 26;47(D1):D759–65.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, et al. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics. Genetics; 2013 Aug;194(4):1029–35. PMCID: PMC3730909

Higuchi RG, Ochman H. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. Oxford University Press; 1989 Jul 25;17(14):5865. PMCID: PMC318226

Hisano Y, Sakuma T, Nakade S, Ohga R, Ota S, Okamoto H, et al. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Nature Publishing Group. Nature Publishing Group; 2015 Mar 5;5(1):8841. PMCID: PMC4350073

Horn C, Jaunich B, Wimmer EA. Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. Dev. Genes Evol. 2000 Dec;210(12):623–9.

Housden, B. E., Nicholson, H. E., & Perrimon, N. (2017). Synthetic Lethality Screens Using RNAi in Combination with CRISPR-based Knockout in Drosophila Cells. *Bio-Protocol*, 7(3). http://doi.org/10.21769/BioProtoc.2119

Jenett A, Rubin GM, Ngo T-TB, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. 2012 Oct 25;2(4):991–1001. PMCID: PMC3515021

Kanca O, Bellen HJ, Schnorrer F. Gene Tagging Strategies To Assess Protein Expression, Localization, and Function in Drosophila. Genetics. 2017 Oct;207(2):389–412. PMCID: PMC5629313

Kunzelmann S, Böttcher R, Schmidts I, Förstemann K. A Comprehensive Toolbox for Genome Editing in Cultured Drosophila melanogaster Cells. G3 (Bethesda). G3: Genes, Genomes, Genetics; 2016 Jun 1;6(6):1777–85. PMCID: PMC4889673

Lanza DG, Gaspero A, Lorenzo I, Liao L, Zheng P, Wang Y, et al. Comparative analysis of single-stranded DNA donors to generate conditional null mouse alleles. BMC Biology; 2018 Jun 15;:1–18.

Lee P-T, Lin G, Lin W-W, Diao F, White BH, Bellen HJ. A kinase-dependent feedforward loop affects CREBB stability and long term memory formation. Elife. 2018a Feb 23;7:2743. PMCID: PMC5825208

Lee P-T, Zirin J, Kanca O, Lin W-W, Schulze KL, Li-Kroeger D, et al. A gene-specific T2A-GAL4 library for Drosophila. Elife. eLife Sciences Publications Limited; 2018b Mar 22;7:e35574. PMCID: PMC5898912

Li-Kroeger D, Kanca O, Lee P-T, Cowan S, Lee MT, Jaiswal M, et al. An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in Drosophila. Elife. 2018 Aug 9;7. PMCID: PMC6095692

Little JW. An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. J. Biol. Chem. 1967 Feb 25;242(4):679–86.

Llamazares S, Moreira A, Tavares A, Girdham C, Spruce BA, Gonzalez C, et al. polo encodes a protein kinase homolog required for mitosis in Drosophila. Genes Dev. Cold Spring Harbor Lab; 1991 Dec;5(12A):2153–65.

McFarland TJ, Zhang Y, Atchaneeyaskul L-O, Francis P, Stout JT, Appukuttan B. Evaluation of a novel short polyadenylation signal as an alternative to the SV40 polyadenylation signal. Plasmid. 2006 Jul;56(1):62–7.

Mitsis PG, Kwagh JG. Characterization of the interaction of lambda exonuclease with the ends of DNA. Nucleic Acids Res. Oxford University Press; 1999 Aug 1;27(15):3057–63. PMCID: PMC148530

Miura H, Gurumurthy CB, Sato T, Sato M, Ohtsuka M. CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. Nature Publishing Group. Nature Publishing Group; 2015 Jul 27;:1–11.

Moutinho-Santos T, Sampaio P, Amorim I, Costa M, Sunkel CE. In vivo localisation of the mitotic POLO kinase shows a highly dynamic association with the mitotic apparatus during early embryogenesis in Drosophila. Biol. Cell. 1999 Nov;91(8):585–96.

Nagarkar-Jaiswal S, Lee P-T, Campbell ME, Chen K, Anguiano-Zarate S, Gutierrez MC, et al. A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. Elife. eLife Sciences Publications Limited; 2015 Mar 31;4:e05338. PMCID: PMC4379497

Nagarkar-Jaiswal S, DeLuca SZ, Lee P-T, Lin W-W, Pan H, Zuo Z, et al. A genetic toolkit for tagging intronic MiMIC containing genes. Elife. eLife Sciences Publications Limited; 2015 Jun 23;4:166. PMCID: PMC4499919

Nagarkar-Jaiswal S, Manivannan SN, Zuo Z, Bellen HJ. A cell cycle-independent, conditional gene inactivation strategy for differentially tagging wild-type and mutant cells. Elife. eLife Sciences Publications Limited; 2017 May 31;6:1385. PMCID: PMC5493436

Neumüller RA, Wirtz-Peitz F, Lee S, Kwon Y, Buckner M, Hoskins RA, et al. Stringent analysis of gene function and protein-protein interactions using fluorescently tagged genes. Genetics. Genetics; 2012 Mar;190(3):931–40. PMCID: PMC3296255

Okada A, Lansford R, Weimann JM, Fraser SE, McConnell SK. Imaging cells in the developing nervous system with retrovirus expressing modified green fluorescent protein. Exp. Neurol. 1999 Apr;156(2):394–406.

Paix A, Folkmann A, Goldman DH, Kulaga H, Grzelak MJ, Rasoloson D, et al. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. Proc. Natl. Acad. Sci. U.S.A. National Academy of Sciences; 2017 Dec 12;114(50):E10745–54. PMCID: PMC5740635
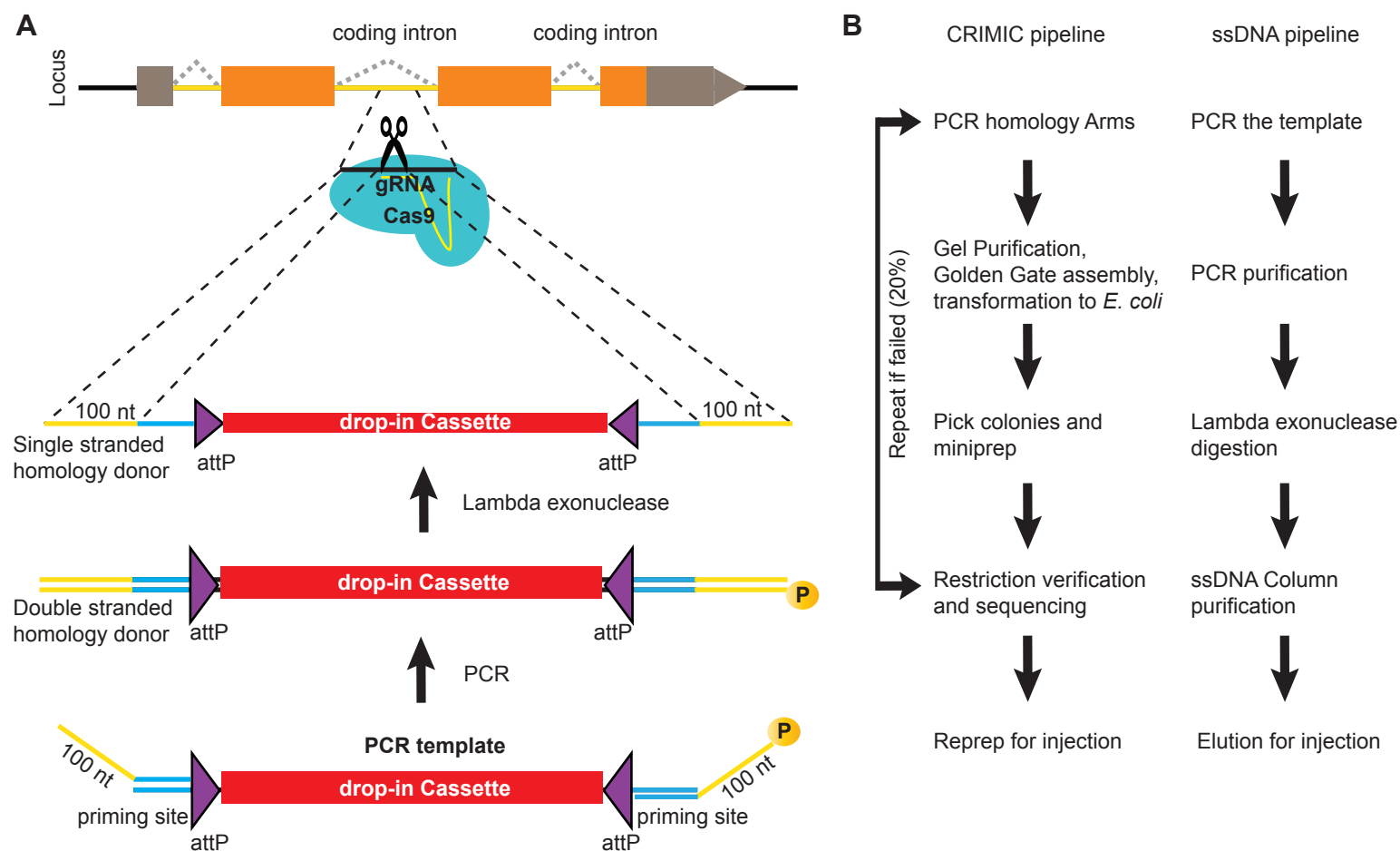
Paix A, Folkmann A, Rasoloson D, Seydoux G. High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. Genetics. Genetics; 2015 Sep;201(1):47–54. PMCID: PMC4566275

Paix A, Wang Y, Smith HE, Lee C-YS, Calidas D, Lu T, et al. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in Caenorhabditis elegans. Genetics. Genetics; 2014 Dec;198(4):1347–56. PMCID: PMC4256755

Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. Nature Publishing Group; 2006 Jan;24(1):79–88.

Port F, Chen H-M, Lee T, Bullock SL. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc. Natl. Acad. Sci. U.S.A. National Academy of Sciences; 2014 Jul 22;111(29):E2967–76. PMCID: PMC4115528

Quadros RM, Miura H, Harms DW, Akatsuka H, Sato T, Aida T, et al. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biology; 2017 May 16;:1–15.

Rong YS, Golic KG. Gene targeting by homologous recombination in Drosophila. Science. American Association for the Advancement of Science; 2000 Jun 16;288(5473):2013–8.

Santos dos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, et al. FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 2015 Jan 15;43(D1):D690–7.

Schmid-Burgk JL, Höning K, Ebert TS, Hornung V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat Commun. Nature Publishing Group; 2016 Jul 28;7:12338. PMCID: PMC4974478

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. Nature Publishing Group; 2004 Nov 21;22(12):1567–72.

Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature. Nature Publishing Group; 2016 Dec 1;540(7631):144–9. PMCID: PMC5331785

Şentürk M, Bellen HJ. ScienceDirect Genetic strategies to tackle neurological diseases in fruit flies. Current Opinion in Neurobiology. Elsevier Ltd; 2018 Jun 1;50:24–32. PMCID: PMC5940587

Thummel C, Pirrotta V. New pCaSpeR P element vectors. Drosoph Inf Serv 1992;71:150

Venken KJT, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M, et al. MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods. 2011 Sep;8(9):737–43. PMCID: PMC3191940

Viswanatha R, Li Z, Hu Y, Perrimon N. Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in Drosophila cells. Elife. eLife Sciences Publications Limited; 2018 Jul 27;7:705. PMCID: PMC6063728

Wissel S, Kieser A, Yasugi T, Duchek P, Roitinger E, Gokcezade J, et al. A Combination of CRISPR/Cas9 and Standardized RNAi as a Versatile Platform for the Characterization of Gene Function. G3 (Bethesda). G3: Genes, Genomes, Genetics; 2016 Aug 9;6(8):2467–78. PMCID: PMC4978900

Yao X, Wang X, Liu J, Hu X, Shi L, Shen X, et al. CRISPR/Cas9 - Mediated Precise Targeted Integration In Vivo Using a Double Cut Donor with Short Homology Arms. EBioMedicine. The Authors; 2017 Jun 1;20(C):19–26.

Yao X, Zhang M, Wang X, Ying W, Hu X, Dai P, et al. Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells. Dev. Cell. Elsevier Inc; 2018 May 21;45(4):526–536.e5.

Yoon WH, Sandoval H, Nagarkar-Jaiswal S, Jaiswal M, Yamamoto S, Haelterman NA, et al. Loss of Nardilysin, a Mitochondrial Co-chaperone for α-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. Neuron. 2017 Jan 4;93(1):115–31. PMCID: PMC5242142

Zhang K, Li Z, Jaiswal M, Bayat V, Xiong B, Sandoval H, et al. The C8ORF38 homologue Sicily is a cytosolic chaperone for a mitochondrial complex I subunit. J. Cell Biol. Rockefeller University Press; 2013 Mar 18;200(6):807–20. PMCID: PMC3601355

Zhang X, Koolhaas WH, Schnorrer F. A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in Drosophila. G3 (Bethesda). G3: Genes, Genomes, Genetics; 2014 Oct 15;4(12):2409–18. PMCID: PMC4267936

Kanca et al. Figure 1. Comparison of ssDNA pipeline to CRIMIC

Kanca et al. Figure 2. ssDNA homology donors are effective in S2 cells to tag organelles
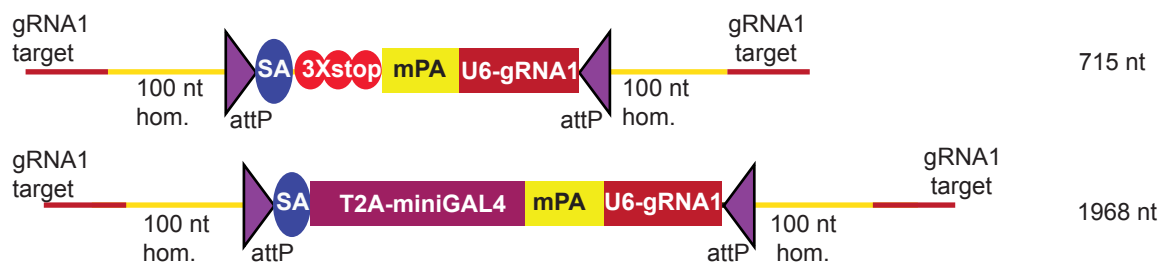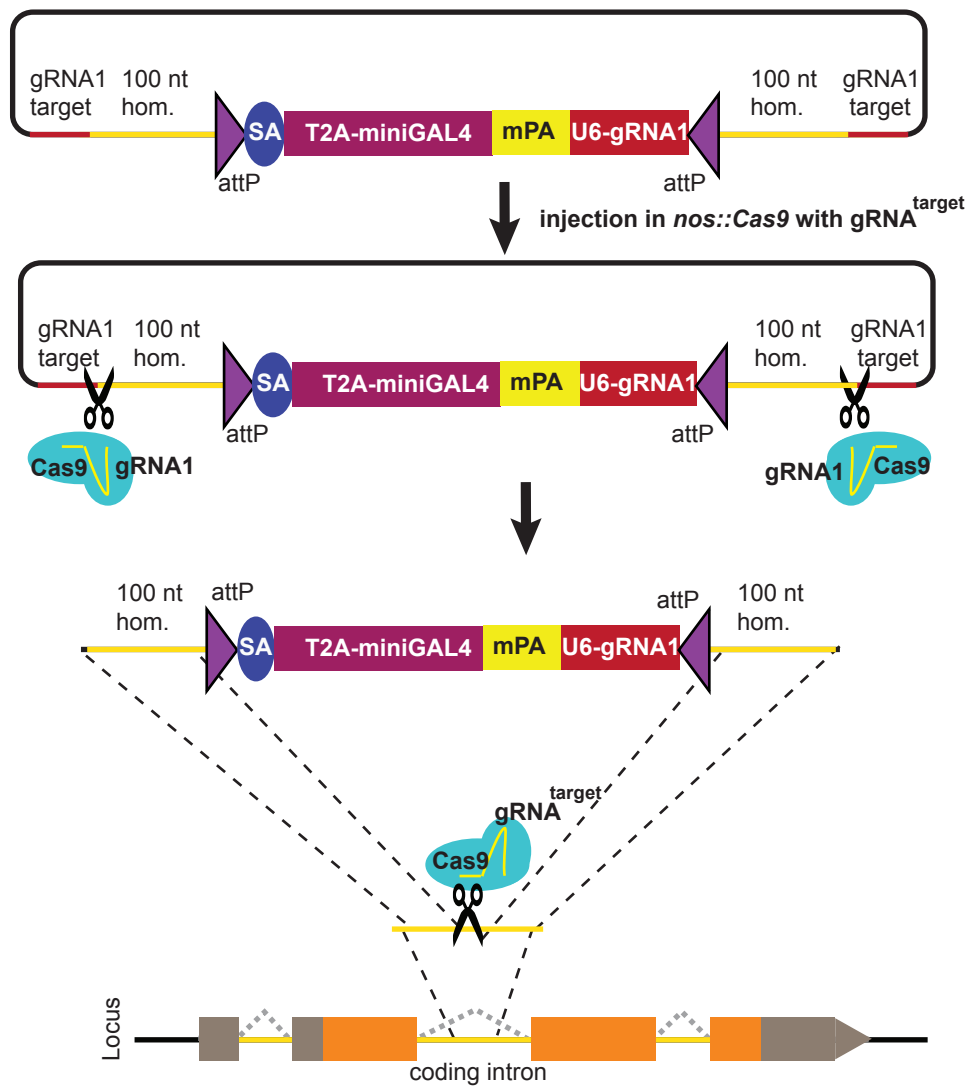
Kanca et al. Figure 3. Comparison of ssDNA drop-in donor to CRIMIC construct for fly transformation
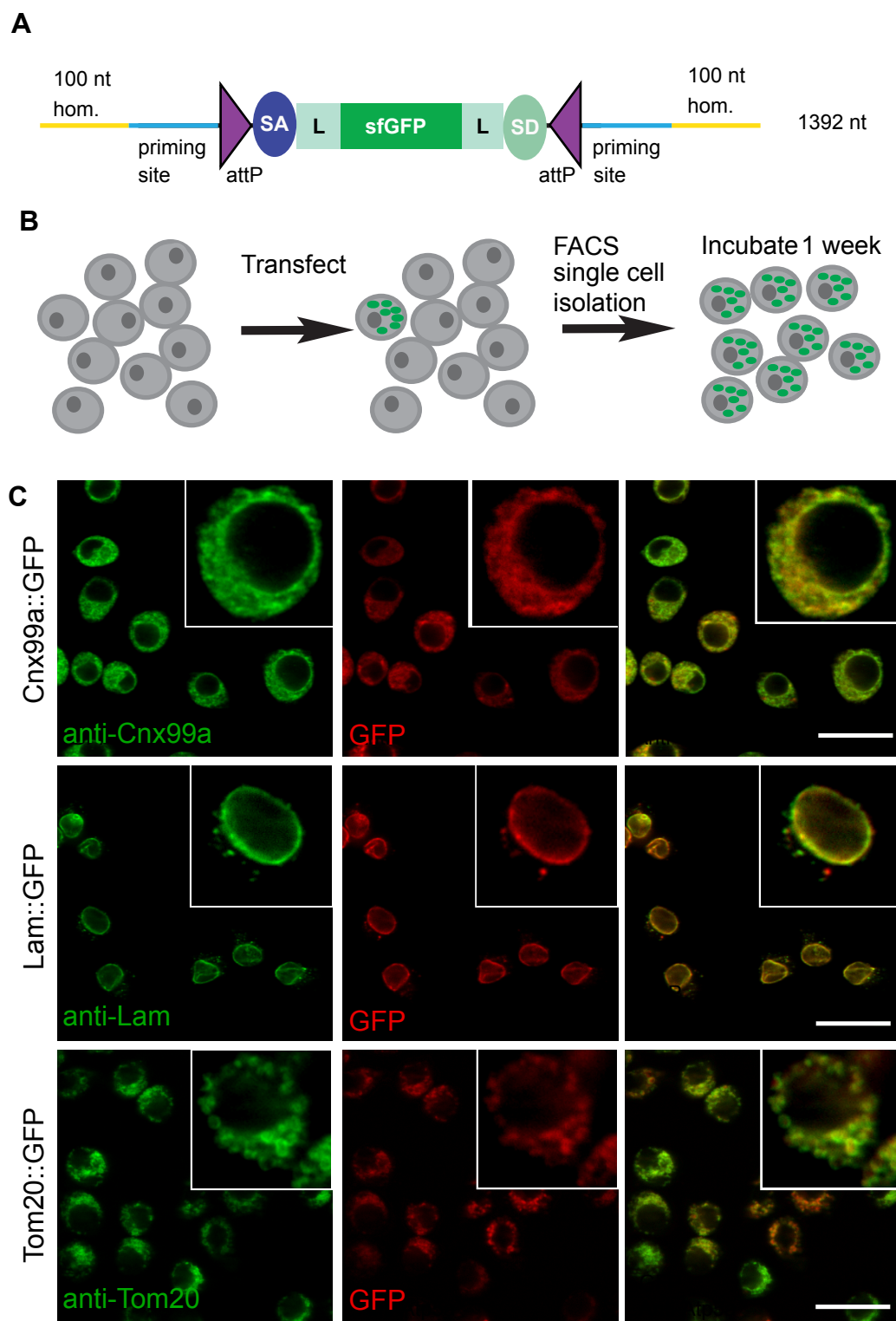
A

B

Gene	CRIMIC	attP-3XP3-EGFP-attP_ssDNA
Acox57D	0/ 102 crosses	0/ 120 crosses
CG3376	Failed at cloning	0/ 104 crosses
CG5009	10/ 160 crosses	1/ 82 crosses
CG9527	10/ 225 crosses	0/ 73 crosses
CG17544	0/ 185 crosses	0/ 77 crosses
Cp1	0/ 181 crosses	1/ 86 crosses
endoB	Failed at cloning	3/ 175 crosses
Khc	10/ 229 crosses	0/ 180 crosses
Lst	Failed at cloning	1/ 107 crosses
NLaz	7/ 162 crosses	0/ 84 crosses

Kanca et al. Figure 4. Double stranded synthetic constructs are efficient for fly transformation

Kanca et al. Figure 5. T2A-miniGAL4 is a mutagen and expresses an active GAL4
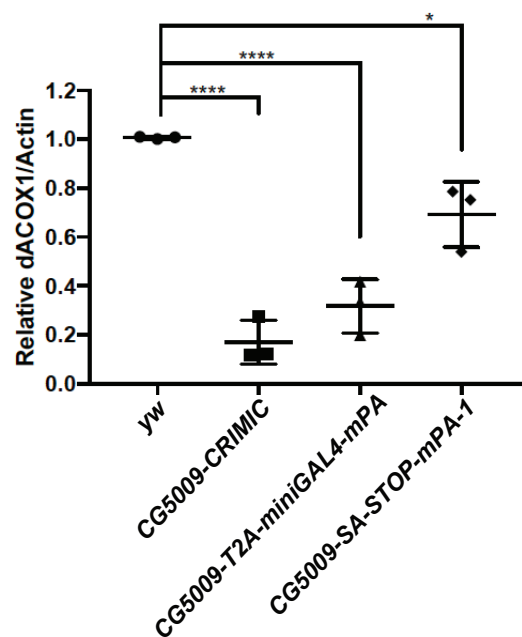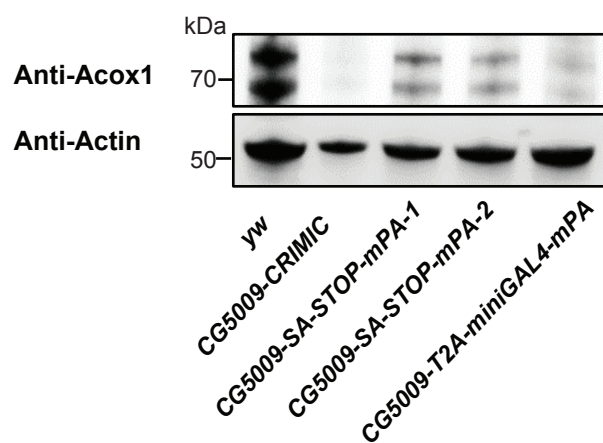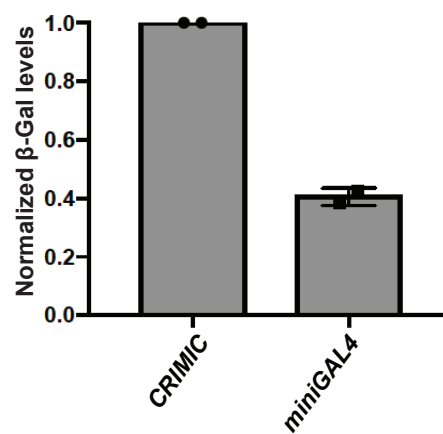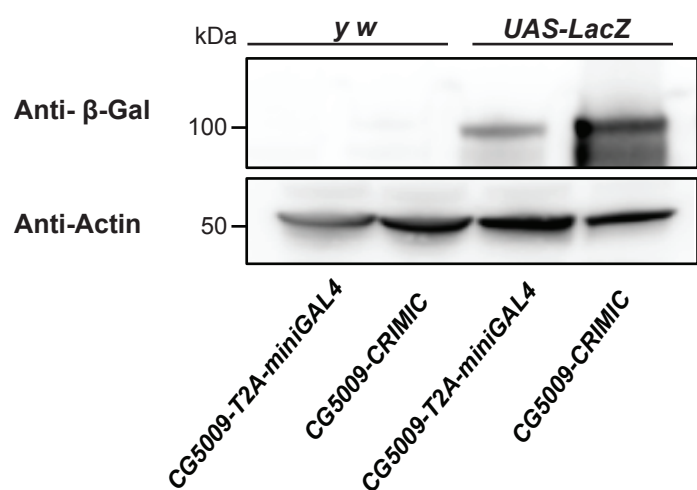
Kanca et al. Figure 6. Single step cloning method allows cheaper and efficient insertion of CRIMIC cassette in coding introns.

A
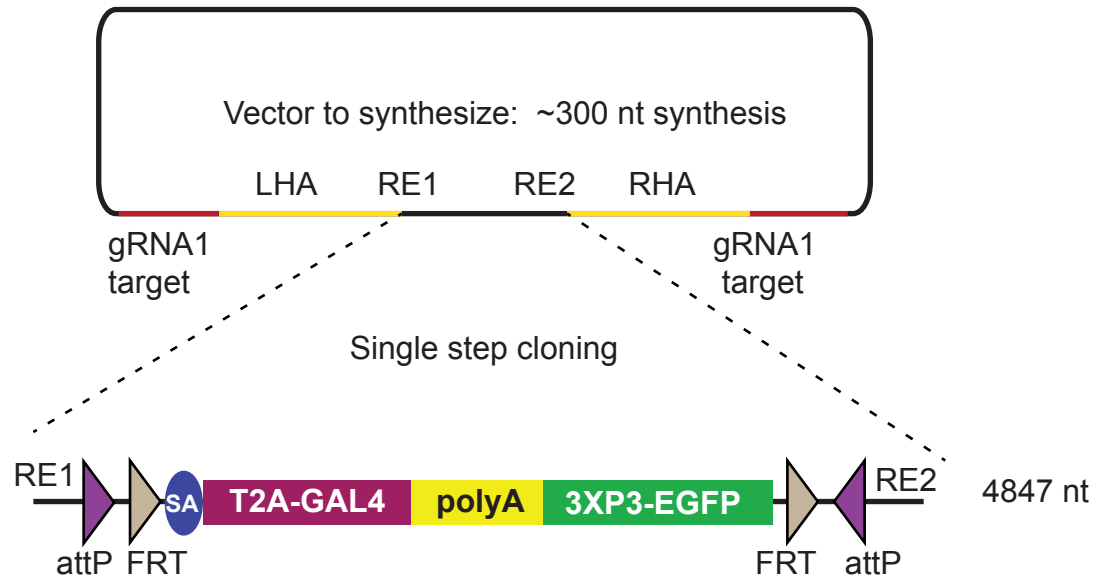
B

Gene	CRIMIC	Drop-in int100-CRIMIC
CG5009	10 / 160 crosses	22 / 140 crosses
CG9527	10 / 225 crosses	0 / 55 crosses
Cp1	0 / 181 crosses	3 / 65 crosses
endoB	Failed at cloning	6 / 135 crosses
Khc	10 / 229 crosses	3 / 110 crosses
Lst	Failed at cloning	0 / 90 crosses
NLaz	7 / 162 crosses	3 / 66 crosses

C
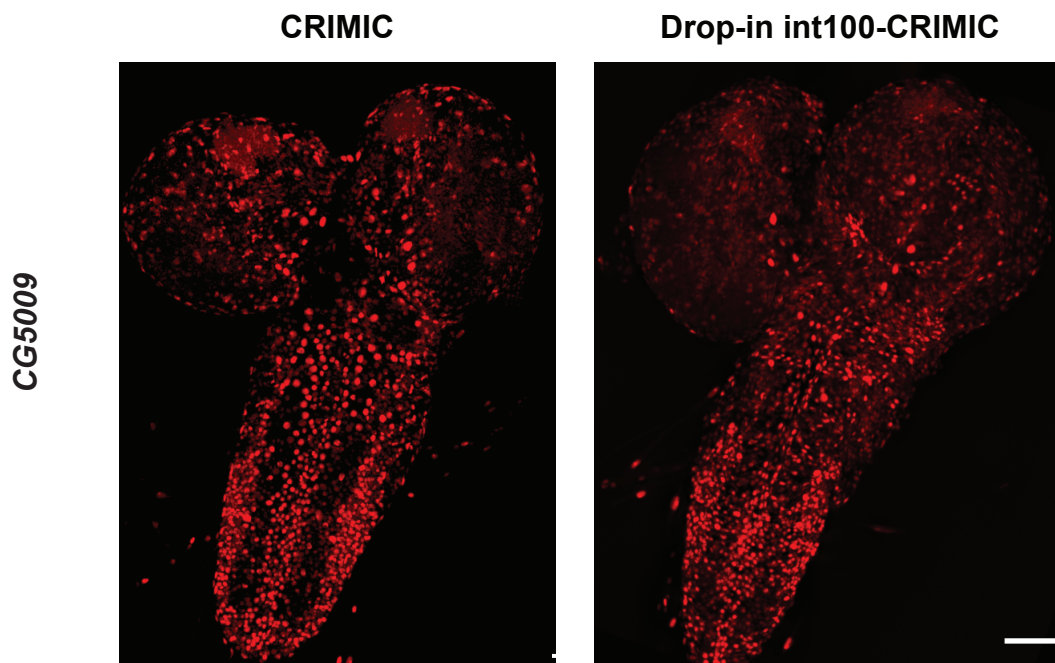
Organelle	Fly protein	clones obtained	population imaged	clones imaged	insertion sequence verified	Immunostained	Correct GFP localization	DGRC stock#	S2R+ expression (modENCODE RPKM)
Autophagosomes	Atg8a	14	✔	✔	N	N	N/A	N/A	153
Autophagosomes/ aggregates	Ref(2)P	9	✔	✔	Y	α-Ref2P, α-FK2	N	N/A	138
Endoplasmic reticulum (ER)	Calnexin99A	16	✔	✔	Y	α-Cnx99a	Y	273	235
Endoplasmic reticulum (ER), transitional	Sec23	30	✔	✔	Y	N	*	294	101
Endosomes, early	Rab5	9	✔	✔	N	N	N/A	N/A	98
Endosomes, recycling	Rab11	23	✔	✔	Y	N	*	274	302
G-Bodies (cytoplasmic puncta)	Pfk	14	✔	✔	N	N	N/A	N/A	23
Golgi (cis-Golgi)	Gmap	10	✔	✔	Y	α-GMAP	Y	276, 277	15
Golgi (trans-Golgi)	Sec71	10	✔	✔	N	N	N/A	N/A	8
Golgi (trans-Golgi)	Golgin245	1	✔	✔	Y	α-Golgin245	Y	280	27
Kinetochore	Polo	2	✔	✔	Y	N	Y	275	50
Lipid droplets	Seipin	12	✔	✔	N	N	N/A	N/A	9
Lysosomes	spin	2	✔	✔	Y	α-Arl8	Y	293	112
Lysosomes	Arl8	9	✔	✔	Y	α-Arl8	Y	291	78
Mitochondria	Tim17b	3	✔	✔	N/A	α-ATP5A	N/A	N/A	266
Mitochondria	Tom20	17	✔	✔	Y	α-ATP5A	Y	302	117
Nuclear membrane, inner	dLBR	0	✔	N/A	N/A	N/A	N/A	N/A	42
Nuclear membrane, inner	Lamin	53	✔	✔	Y	α-Lamin	Y	292	249
Nucleolus	Fibrillarin	14	✔	✔	Y	α-Fib	Y	278, 279	53
Peroxisomes	Pmp70	6	✔	✔	N	N	N/A	N/A	26

* = distribution was as expected, but no antibody available to test by co-stain

Sequence of the attP-SA-Linker-sfGFP-Linker-SD-attP (phase 0)

GAATTCTGTAAAACGACGGCCAGTGGAGTAGTGCCCCAACTGGGGTAACCTTT
GAGTTCTCTCAGTTGGGGGCGTAGGCCTGCAGGAGTCGATCCAACATGGCGAC
TTGTCCCATCCCCGGCATGTTTAAATATACTAATTATTCTTGAACTAATTTTAATC
AACCGATTTATCTCTCTTCCGCAGGTGGGAGGTTCCGGTGGAAGCGGAGGTAG
CGGCGGATCCATGGTGTCCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCC
ATCCTGGTGGAGCTGGATGGCGACGTGAACGGCCACAAGTTCAGCGTGCGCG
GCGAGGGCGAGGGCGACGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACC
TACGGCGTGCAGTGCTTCAGCCGCTACCCCGATCACATGAAGCAGCACGATTT
CTTCAAGAGCGCCATGCCCGAGGGCTACGTGCAGGAGCGCACCATCAGCTTCA
AGGATGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGATAC
CCTGGTGAACCGCATCGAGCTGAAGGGCATCGATTTCAAGGAGGATGGCAACA
TCCTGGGCCACAAGCTGGAGTACAACTTCAACAGCCACAACGTGTACATCACC
GCCGATAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAATGT
GGAGGATGGCTCCGTGCAGCTGGCCGATCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCAGTGCTGCTGCCCGATAACCACTACCTGAGCACCCAGAGCG
TGCTGTCCAAGGACCCCAACGAGAAGCGCGATCACATGGTGCTGCTGGAGTTC
GTGACCGCCGCCGGCATCACCCTGGGCATGGATGAGCTGTACAAGGGTACCG
GATCCGGAGGTAGCGGTGGAAGCGGAGGTTCCGGCGAGGTAAGTTATTGAAC
AATGGCATCAAATGCCTTCATCATCACTACCCTTTAGCCCTTAAGACCCCACAAT
GACCTTACCCACTCAGAGAAAAAGTAAATATGAAAGCCCATTTGAACTTCTGC
GGCCGCcctacgcccccaactgagagaactcaaaggttaccccagttggggcactactccgtcatagctgtttcc
tgggggc

NNNN primer binding site; NNNN phiC31 attP; NNNN *mhc* splice acceptor; NNNN
Linker sequence; NNNN sfGFP; NNNN *mhc* splice donor; nnnn reverse strand

Linker sequence for phase 1:
GTGGCGGAGGTTCCGGTGGAAGCGGAGGTAGCGGC //
GGTACCGGATCCGGAGGTAGCGGTGGAAGCGGAGGTTCCGGAG

Linker sequence for phase 2:
GTCGGGAGGTTCCGGTGGAAGCGGAGGTAGCGGC //
GGTACCGGATCCGGAGGTAGCGGTGGAAGCGGAGGTTCCGGCAG

Sequence of the attP-3XP3-EGFP-attP

GAATTCTGTAAAACGACGGCCAGTGGAGTAGTGCCCCAACTGGGGTAACCTTT
GAGTTCTCTCAGTTGGGGGCGTAGGGGATCTAATTCAATTAGAGACTAATTCAA
TTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCC
GGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGC
AAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAA
CAAGCTAAACAATCGGACTAGAGCCGGTCGCCGGCCGGCCACCATGGTGTCC
AAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCAATTCTGGTGGAGCTGGATG
GCGACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGACG
CCACCTATGGAAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC
GTGCCATGGCCAACCCTCGTGACCACCCTGACCTATGGCGTGCAGTGCTTCAG

CCGCTACCCCGATCACATGAAGCAGCACGATTTCTTCAAGAGCGCCATGCCCG
AGGGCTACGTGCAGGAGCGCACCATCTTTTTCAAGGATGACGGCAACTACAAG
ACCCGCGCCGAAGTGAAGTTCGAGGGCGATACCCTCGTGAACCGCATCGAGC
TGAAGGGCATCGATTTCAAGGAGGATGGAAACATCCTGGGCCACAAGCTGGAG
TACAACTACAACAGCCACAACGTGTACATCATGGCCGACAAGCAGAAGAACGG
CATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGATGGCGGCGTGCAGC
TGGCCGATCACTACCAGCAGAACACCCCAATCGGCGACGGCCCAGTGCTGCT
GCCCGATAACCATTACCTGAGCACCCAGAGCGCCCTGAGCAAGGATCCCAACG
AGAAGCGCGACCACATGGTGCTGCTGGAGTTTGTGACCGCCGCCGGCATTACC
CTGGGCATGGATGAGCTGTACAAGTAGGATCCAGACATGATAAGATACATTGAT
GAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAA
TTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTccta
cgcccccaactgagagaactcaaaggttaccccagttggggcactactccgtcatagctgtttcctggcggc

NNNN primer binding site; NNNN phiC31 attP; NNNN 3XP3GFP-polyA;
nnnn reverse strand

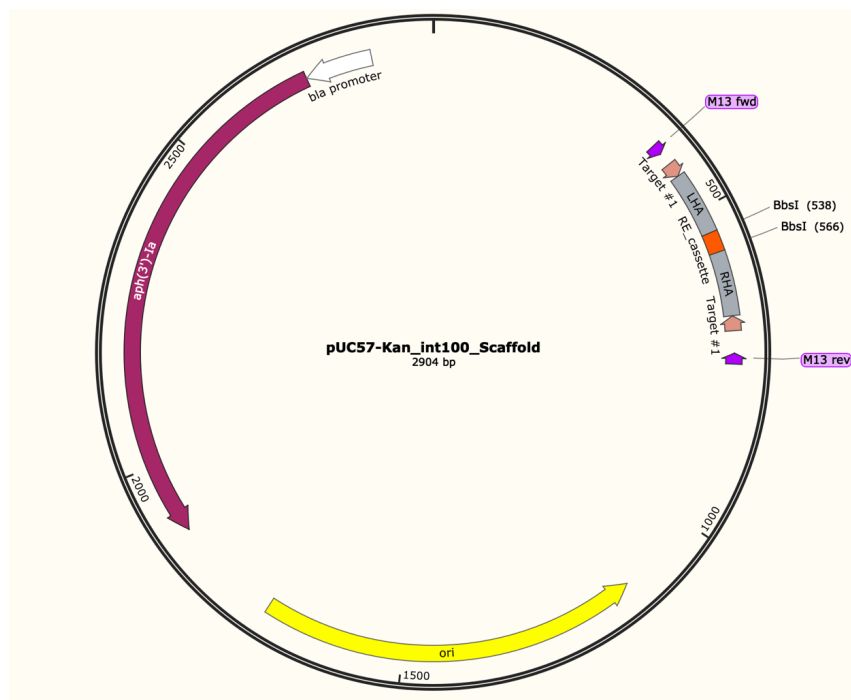**Sequence of the gRNA1 target-Homology arm-attP-SA-3XSTOP-minipolyA-U6gRNA1-attP-Homology arm-gRNA1 target**

GTAGTACGATCATAACAACGCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNGGAGTAGTGCCCCAACTGGGGTAACCTTTGAGTT
CTCTCAGTTGGGGGCGTAGGCCTGCAGGAGTCGATCCAACATGGCGACTTGTC
CCATCCCCGGCATGTTTAAATATACTAATTATTCTTGAACTAATTTTAATCAACCG
ATTTATCTCTCTTCCGCAGGTTAACGTAACCTAGGAAATAAAATACGAAATGAAT
TCTACCCAAAAGCAGAGAGGGCGCCAGTGCTCACTACTTTTTATAATTCTCAAC
TTCTTTTTCCAGACTCAGTTCGTATATATAGACCTATTTTCAATTTAACGTCGTAG
TACGATCATAACAACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT
CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGCGGCCGCcct
acgcccccaactgagagaactcaaaggttaccccagttggggcactactccNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNccgcgttgttatgatcgtactac

NNNN gRNA1-target; NNNN Left homology arm; NNNN phiC31 attP; NNNN *mhc*
splice acceptor; NNNN Stop codons; NNNN short polyA; NNNN U6 promoter-gRNA1
Dominant marker ; NNNN Right homology arm; nnnn reverse strand

Sequence of the gRNA1 target-Homology arm-attP-SA-T2A-miniGAL4-miniPA-U6gRNA1-attP-Homology arm-gRNA1 target (phase 0)

GTAGTACGATCATAACAACGCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNGAATTCTGTAAAACGACGGCCAGTGGAGTAGTGCC
CCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGCCTGCAGGAG
TCGATCCAACATGGCGACTTGTCCCATCCCCGGCATGTTTAAATATACTAATTAT
TCTTGAACTAATTTTAATCAACCGATTTATCTCTCTTCCGCAGGTGGGAGGTTCC
GGTGGAAGCGGAGGTAGCGGCGGATCCGAGGGCCGCGGCAGCCTGCTGACC
TGCGGCGATGTGGAGGAGAACCCCGGGCCCATGAAGCTGTTGTCCTCCATCG
AGCAAGCTTGTGACATCTGCCGTTTAAAGAAGCTGAAGTGCAGCAAGGAGAAG

CCCAAATGCGCTAAGTGTTTAAAGAACAATTGGGAATGCCGCTACAGCCCCAAG
ACCAAGCGCAGCCCCTTGACCCGCGCTCATTTAACCGAAGTCGAGAGCCGTTT
AGAGCGCTTGGAGCAACTGTTTTTACTGATCTTTCCCCGCGAGGATTTAGACAT
GATTTTAAAGATGGACTCTTTACAAGATATCAAGGCTTTACTGACCGGCTTGTTC
GTGCAAGATAACGTGAATAAGGATGCCGTGACCGACCGTTTAGCTTCCGTGGA
AACTGATATGCCTTTAACTTTACGTCAGCATCGTATCTCCGCCACCAGCTCCTC
CGAGGAAAGCAGCAACAAGGGCCAGCGCCAGTTGACCGTGTCCATCGATAGC
GCCGCCCACCACGACAACTCCACCATTCCGCTGGACTTCATGCCCCGCGATGC
TTTACACGGATTCGACTGGTCCGAGGAGGATGATATGTCCGACGGTTTACCCTT
TTTAAAAACCGACCCCAACAACAACGGCTTCTTTGGCGATGGCTCTTTATTGTG
CATTTTACGCTCCATCGTCAAGCTGCTGTCCAATCGCCCCCCGTCCCGCAACA
GCCCCGTGACCATTCCGCGCAGCACCCCCAGCCATCGCTCCGTCACCCCGTTC
TTGGGCCAGCAGCAGCAGCTGCAATCTTTAGTGCCGTTGACGCCCAGCGCTTT
ATTTGGCGGAGCCAATTTCAATCAGAGCGGCAACATCGCCGACAGCTCTTTATC
CTTCACCTTCACCAACTCCAGCAACGGCCCGAATTTAATCACCACGCAGACCAA
CAGCCAAGCTTTAAGCCAGCCGATTGCCTCCTCCAACGTCCACGACAACTTCAT
GAACAACGAGATTACCGCCTCCAAGATCGATGACGGCAATAATTCCAAGCCTTT
ATCCCCGGGATGGACGGACCAAACCGCCTACAATGCCTTCGGCATCACCACGG
GCATGTTCAACACCACCACCATGGACGACGTGTACAACTATTTATTCGATGACG
AGGATACCCCCCCGAACCCCAAAAAGGAGTAACCTAGGAAATAAAATACGAAAT
GAATTCTACCCAAAAGCAGAGAGGGCGCCAGTGCTCACTACTTTTTATAATTCT
CAACTTCTTTTTCCAGACTCAGTTCGTATATATAGACCTATTTTCAATTTAACGTC
GTAGTACGATCATAACAACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGCGGC
CGCcctacgcccccaactgagagaactcaaaggttaccccagttggggcactactccgtcatagctgtttcctggggg
gcNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNccgc
gttgttatgatcgtactac

NNNN gRNA1-target; NNNN Left homology arm; NNNN primer binding site; NNNN phiC31 attP; NNNN *mhc* splice acceptor; NNNN Linker sequence; NNNN T2A-miniGAL4; NNNN short polyA; NNNN U6 promoter-gRNA1 Dominant marker NNNN Right homology arm; nnnn reverse strand

Linker sequence for phase 1: TGGGAGGTTCCGGTGGAAGCGGAGGTAGCGGC
Linker sequence for phase 2: GTCGGGAGGTTCCGGTGGAAGCGGAGGTAGCGGC

Sequence of the Int100-Scaffold

GTAGTACGATCATAACAACGCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNAAACTTGTCTTCATTATATAGAAGACTTCGCGNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNccgcgttgtta
tgatcgtactac

NNNN gRNA1_-arget; NNNN Left homology arm; NNNN Right homology arm; NNNN RE Cassette ; nnnn reverse strand

Design of Int-100 CRIMIC constructs:

Design region:

- 100 nucleotides on both sides of the cut site are used for synthesis using the template DNA file (int100_scaffold file, see above)

After synthesis the construct looks like below:

Cloning of ds drop-in constructs:

- Cut the vector with BbsI-HF (NEB R3539L). Cut ~2ug of plasmid for 2hrs at 37℃ in 30 ul reaction volume with 1 ul of enzyme (~20Units).

-Cut pM37 of proper phase with BbsI-HF and ScaI-HF (ScaI-HF cuts the pM37 backbone to help resolve the fragments easier (below is how BbsI alone vs BbsI+ScaI looks). If you do not wish to cut with ScaI-HF resolving the gel longer would work too). 4848 bp fragment is the one to isolate.

MW: 1 kb Plus DNA Ladder

1: pM37_p2
   BbsI
       1. 4848 bp
       2. 2982 bp
2: pM37_p2
   BbsI + ScaI
       1. 4848 bp
       2. 1848 bp
       3. 1134 bp
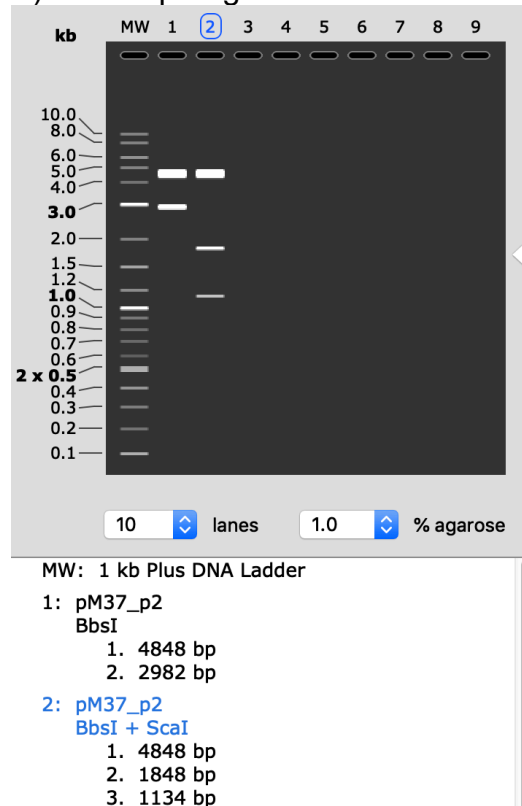
- Gel isolate fragments and set up ligation reaction:

    100ng vector backbone
    150-200 ng pM37 fragment
    1ul T4 ligase buffer
    1ul T4 ligase (NEB M0202L, not quick ligase)
    dH$_2$O to 10ul

Incubate 1-2 hrs at RT. Transform 1ul to 50 ul chemocompetent DH5-alpha. Selection antibiotic is Kanamycin, hence 1-hour recovery is necessary after heatshock. Plate on LB plates with Kanamycin. Incubate at 37℃ overnight.

-Next day do colony PCR with primers M13F-Long for CRIMIC-ch-rev

M13F-long for gacgttgtaaaacgacggccag
CRIMIC-ch-rev gcggaagagagataaatcggttg

Use autoclaved micropipette tip to pick a colony, touch it on a gridded plate to copy the colony and dip the same pipette tip to PCR mix.

    PCR conditions
    0.2ul forward primer
    0.2ul reverse primer

12.1 ul dH$_2$O
12.5 ul OneTaq Quick-Load 2X Master Mix (M0486L)

Rxn: 94℃ 30 sec
　　　94℃ 30 sec |
　　　58℃ 30 sec | 34 cycles
　　　68℃ 30 sec |
　　　68℃ 5 minutes
　　　8℃ Hold

-The positive colonies will show ~535 nt amplicon. (Example below)
-Positive colonies can be grown for Mini/Midi prep. Final constructs are sequenced using M13 forward/reverse primers.
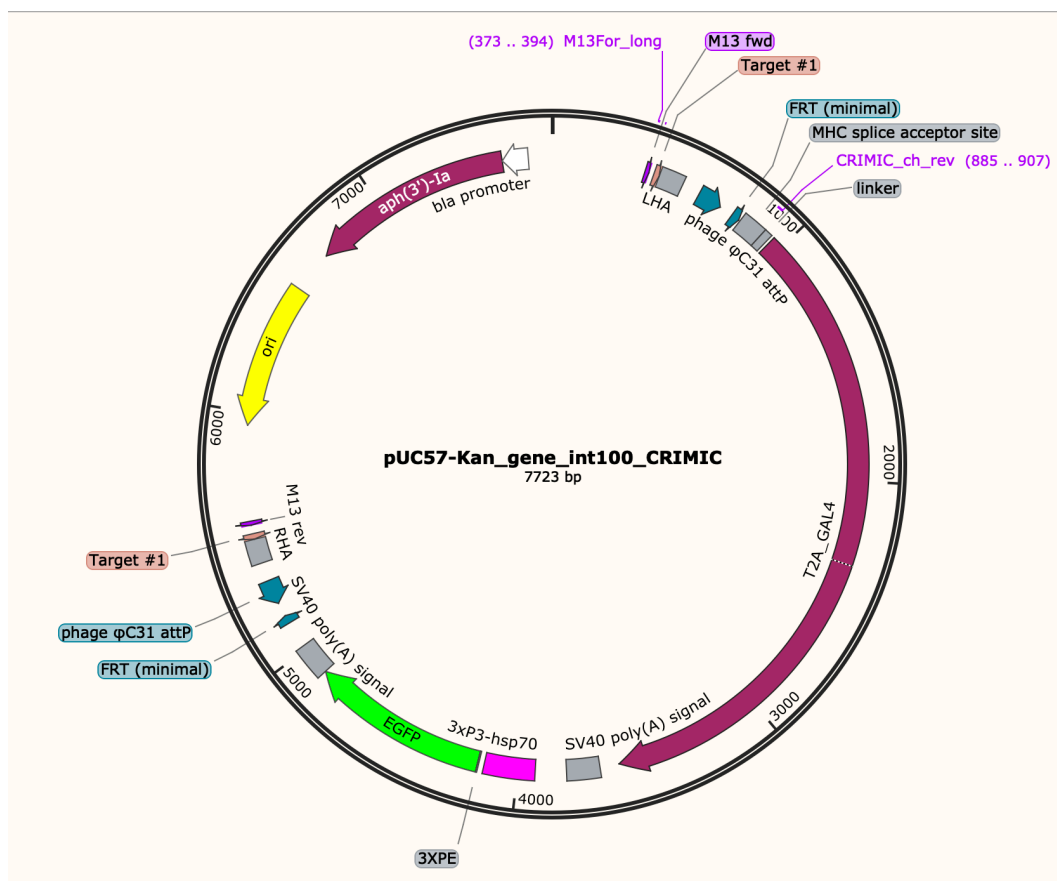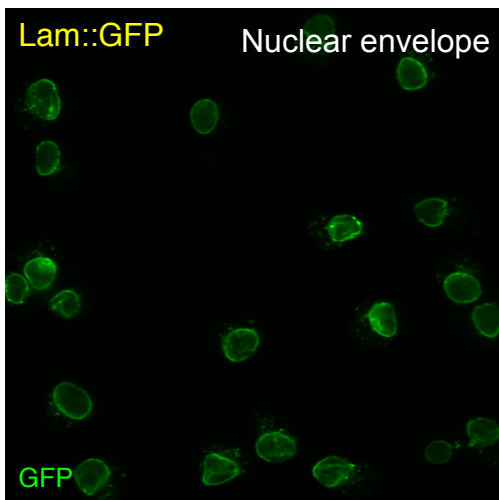

Injection:

-We inject in ~400 embryos of isogenized nos-Cas9 fly lines (iso6 for X, iso18 for 2nd and iso5 for 3rd chromosome). The WGS for these lines should be used for the design.
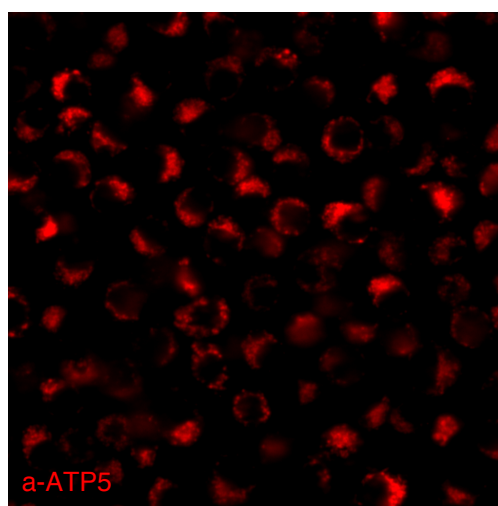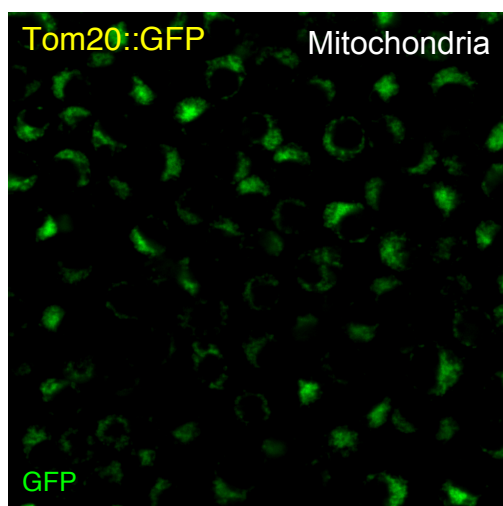-Inject 200-400 ng/ul of donor DNA together with 25ng/ul pCDF3 gene specific gRNA plasmid.
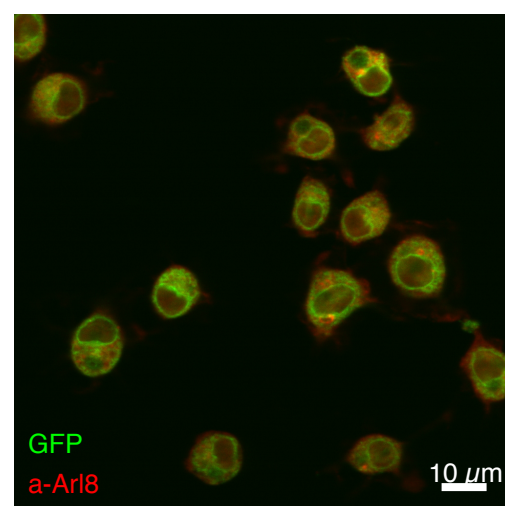
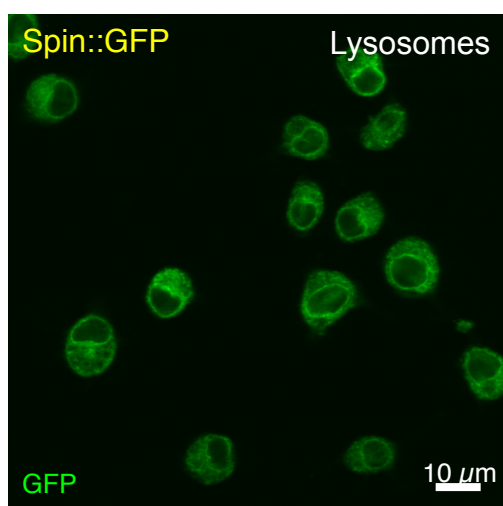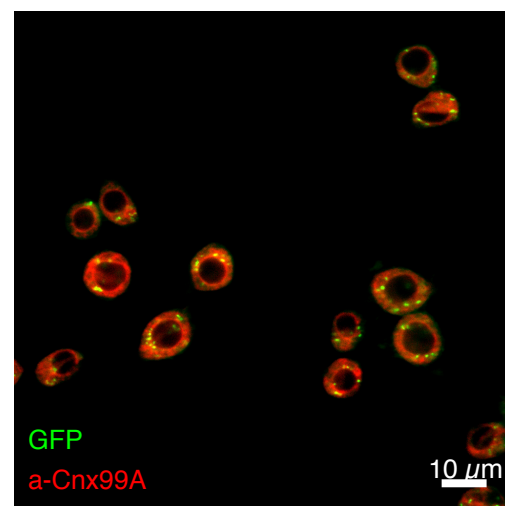**Kanca et al. Figure S1. FACS data of control cells (left) and ssDNA knock-in cells (right).**
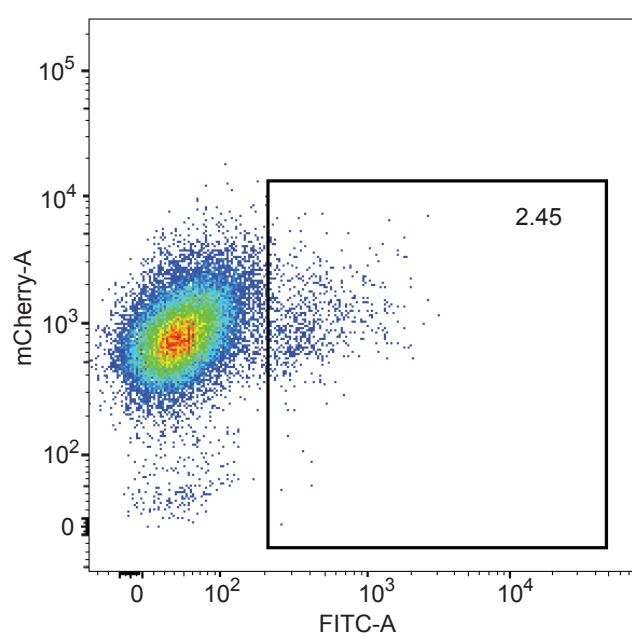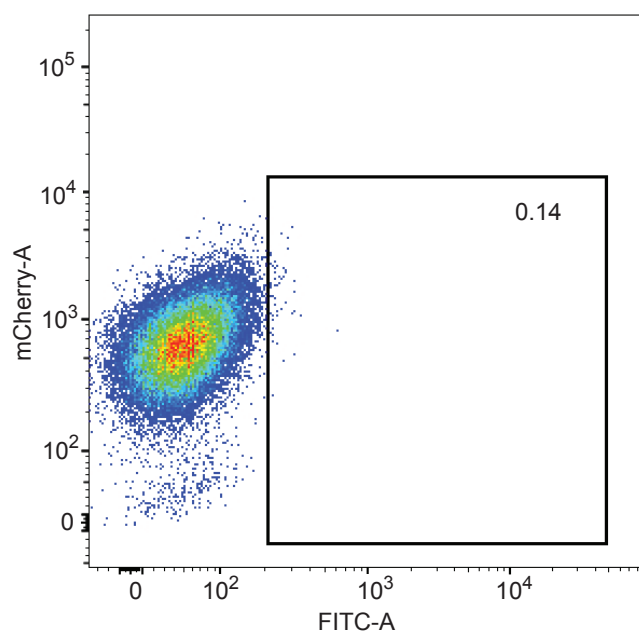
Golgin245::GFP — trans-Golgi
GFP | a-Golgin245 | GFP / a-Golgin245 | 10 μm

Lam::GFP — Nuclear envelope
GFP | a-Lam | GFP / a-Lam | 10 μm

Polo::GFP — Kinetochore
GFP | DAPI | GFP / DAPI | 10 μm

Rab11::GFP — Recycling endosomes
GFP | 10 μm
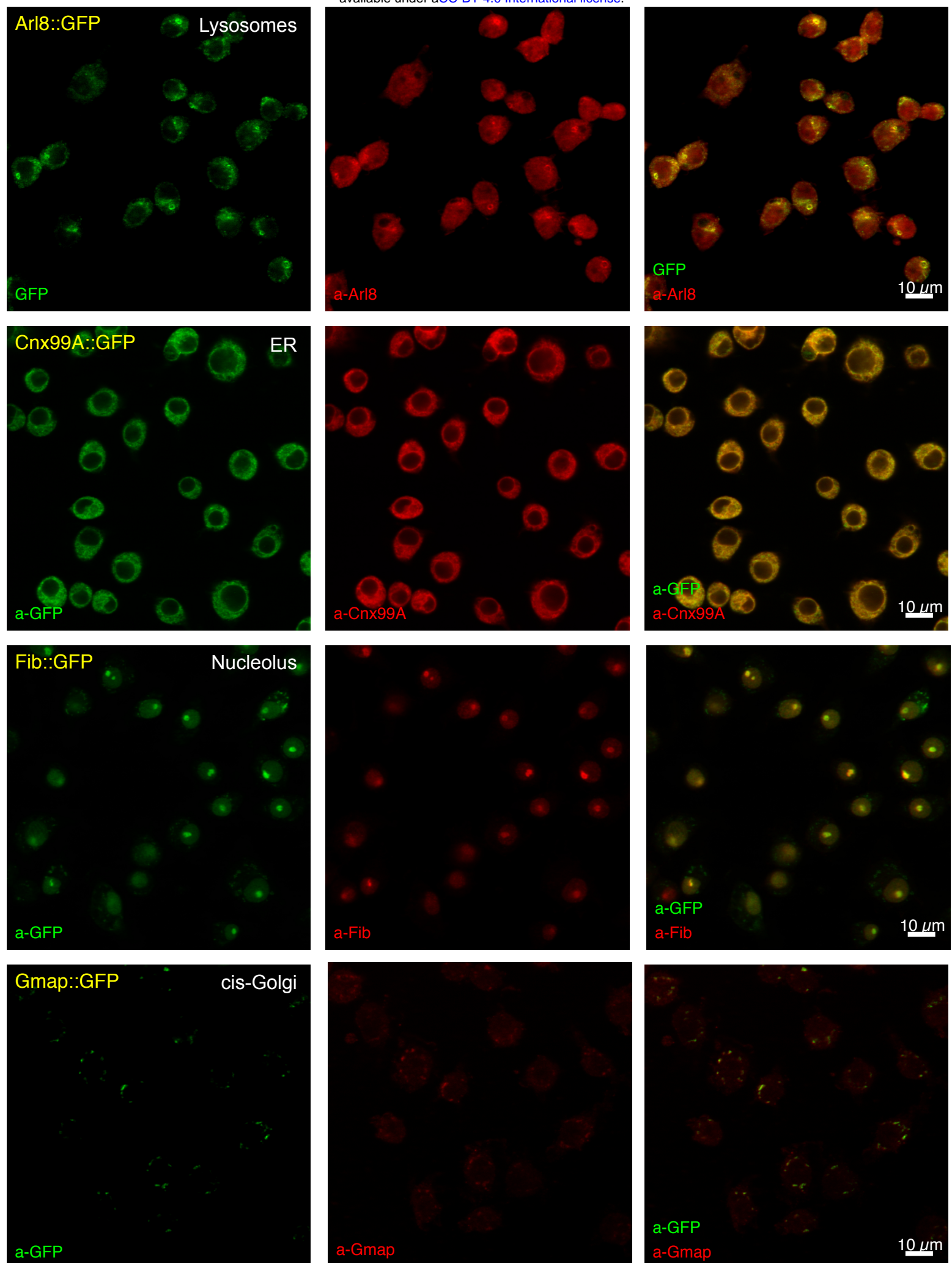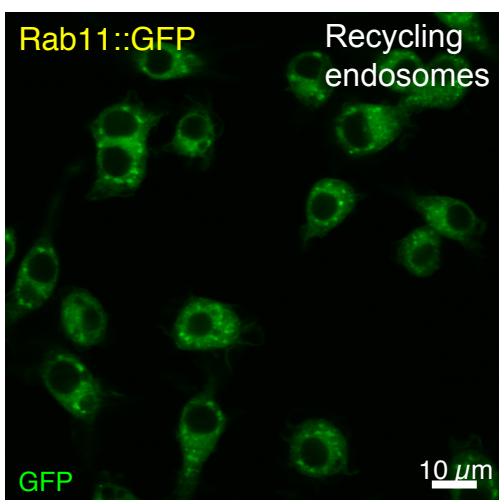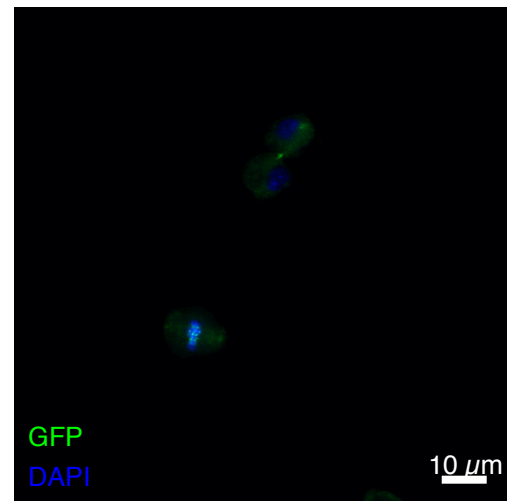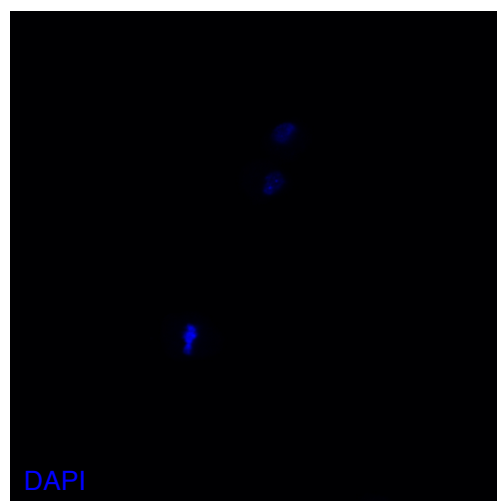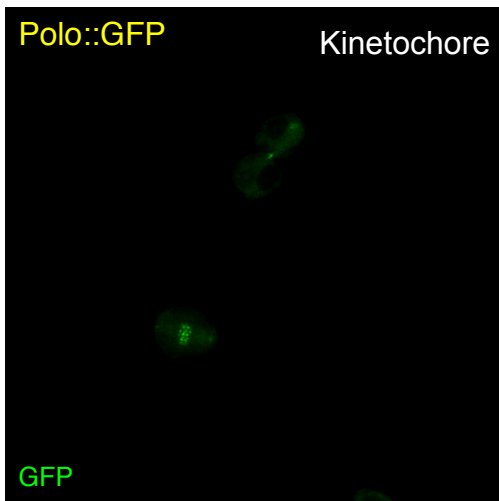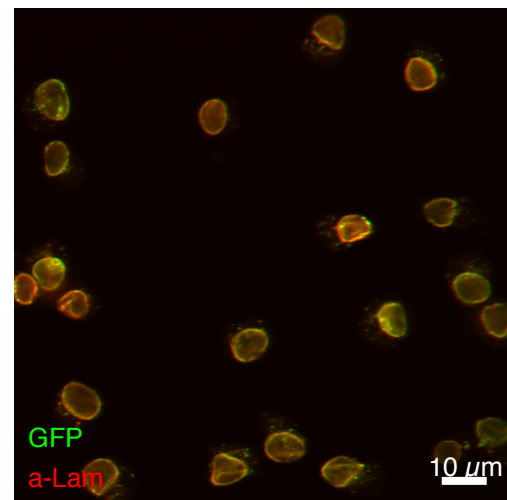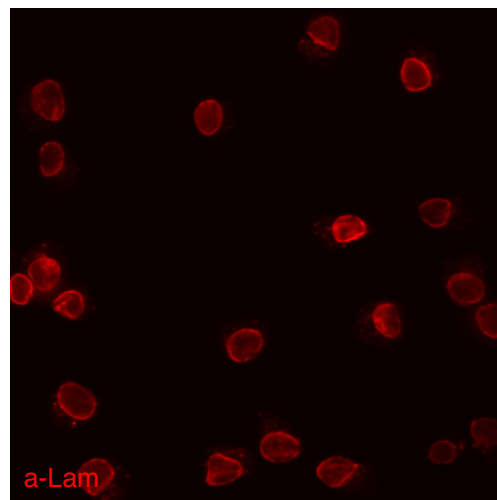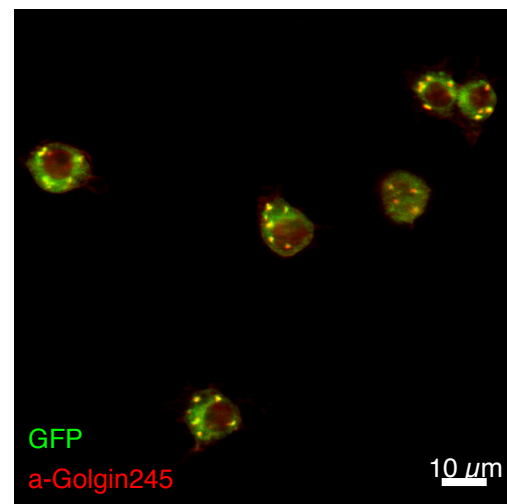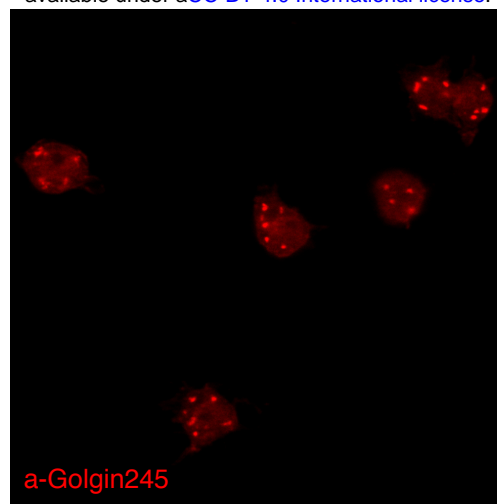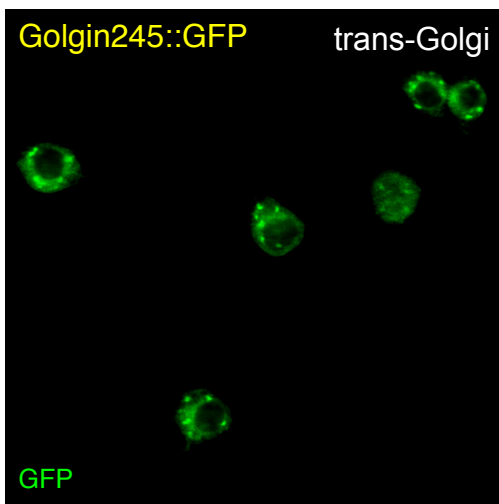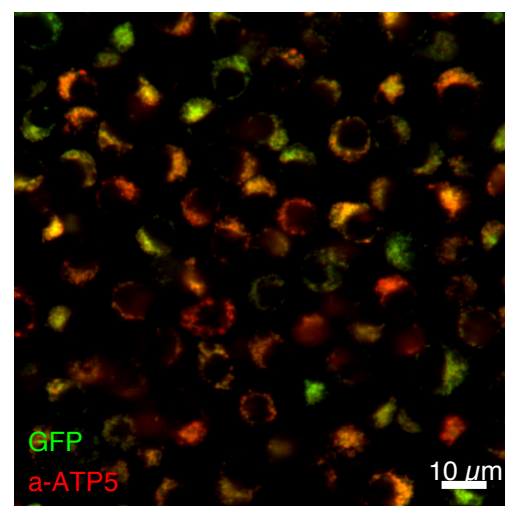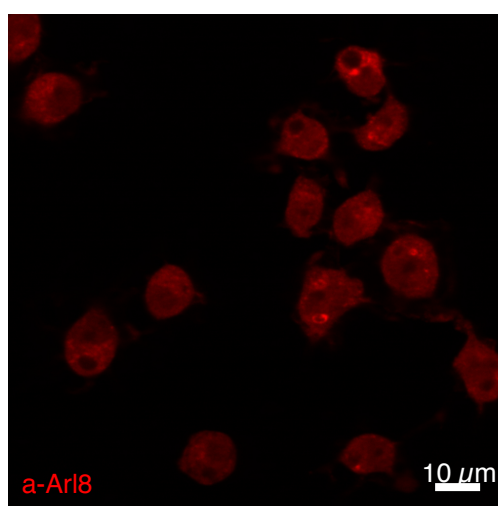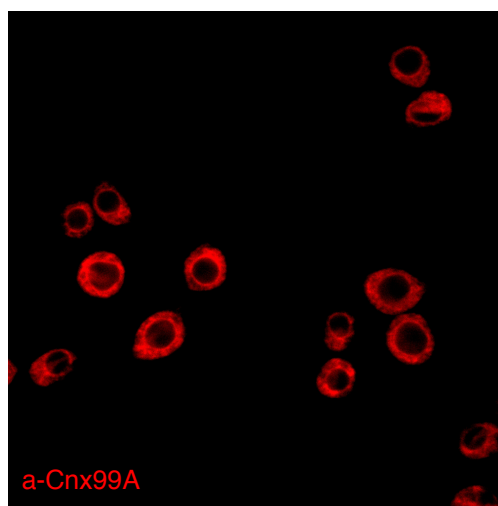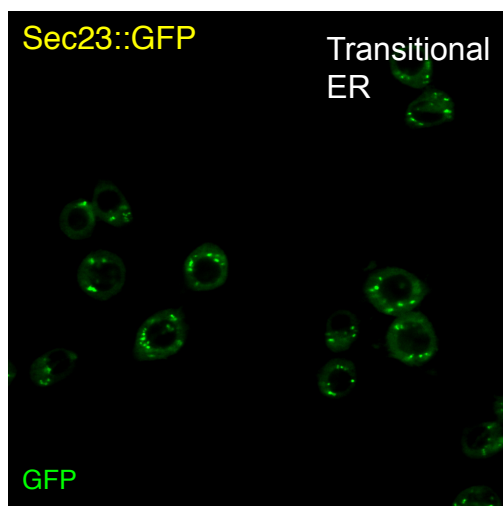
Kanca et al. Figure S3. Western blot analysis of tagged proteins observed in S2R+ cells.
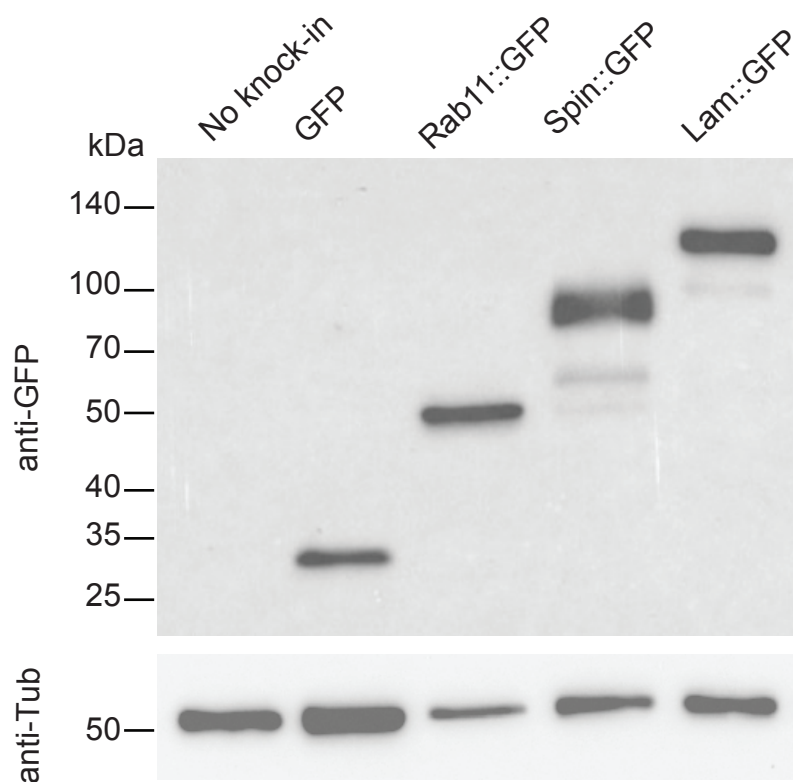
Kanca et al. Figure S4. Schematic and crossing scheme for CRISPR gRNA-based dominant marker strategy.

A

gRNA^1 target

Actin | mCher r | r y

Without gRNA and Cas9   mCher r | r y   → **No signal**

Cas9 | gRNA^1

With gRNA and Cas9   mCher r | r y   →   mCher r y   **Ubiquitous mCherry signal**

B

$$\frac{y\,w}{\quad} ; iso ; \frac{attP2(y+)\{nos\text{-}Cas9(v+)\}}{attP2(y+)\{nos\text{-}Cas9(v+)\}} \quad X \quad \frac{Actin5C\text{-}Cas9\,(w+)\,(ZH\text{-}2A)}{Actin5C\text{-}Cas9\,(w+)\,(ZH\text{-}2A)} ; \f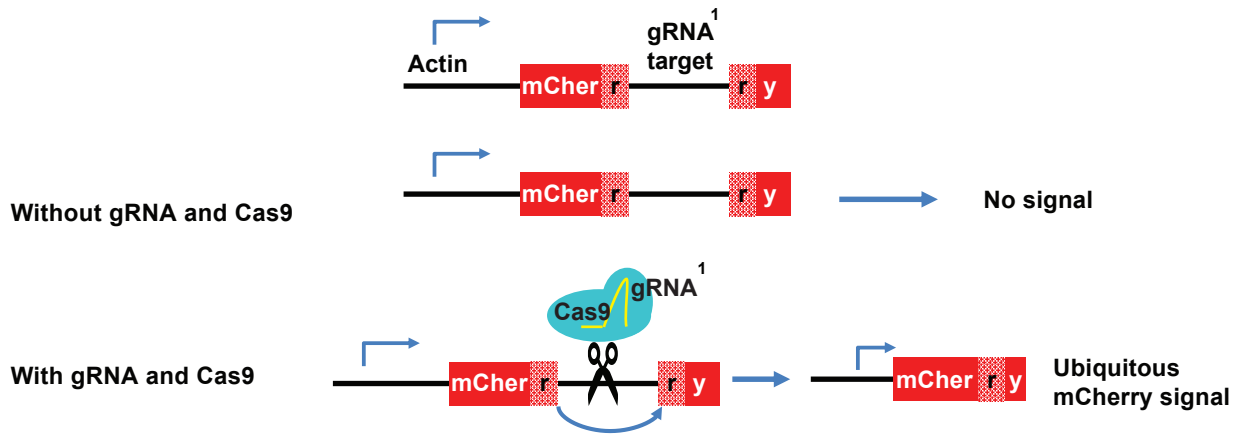rac{Actin5C\text{-}GF\#2FP\,(w+)}{Actin5C\text{-}GF\#2FP\,(w+)} ; \frac{Actin5C\text{-}mCherr\#1ry\,(w+)}{Actin5C\text{-}mCherr\#1ry\,(w+)}$$

Select for mCherry expression (reconstituted ubiquitous signal)

$$Actin5C\text{-}Cas9\,(w+)\,(ZH\text{-}2A) ; \frac{iso\,GOI(Drop\text{-}in\,U6gRNA1)}{Actin5C\text{-}GF\#2FP\,(w+)} ; \frac{Actin5C\text{-}mCherry\,(w+)}{attP2(y+)\{nos\text{-}Cas9(v+)\}} \quad X \quad \frac{y\,w}{y\,w} ; \frac{wg[Sp],Kr[If]}{CyO}$$

Select w-, CyO+

$$y\,w \quad ; \frac{iso\,GOI(Drop\text{-}in\,U6gRNA1)}{CyO} ; \frac{attP2(y+)\{nos\text{-}Cas9(v+)\}}{+} \quad X \quad y\,w \quad ; \frac{wg[Sp],Kr[If]}{CyO}$$

Select w-, CyO+, Sp-, If-

$$y\,w \quad ; \frac{iso\,GOI(drop\text{-}in\,U6gRNA1)}{CyO}$$

← sibling crosses to establish the line

Kanca et al. FigureS5. Comparison of expression domain obtained by T2A-miniGAL4 and drop-in int100-CRIMIC.

Drop-in int100-CRIMIC

T2A-miniGAL4

Khc