An expanded toolkit for Drosophila gene tagging using synthesized homology donor constructs for CRISPR mediated homologous recombination

Oguz Kanca[1,2], Jonathan Zirin[3], Yanhui Hu[3], Burak Tepe[1,2], Debdeep Dutta[1,2], Wen-Wen Lin[1,2], Liwen Ma[1,2], Ming Ge[1,2], Zhongyuan Zuo[1,2], Lu-Ping Liu[3], Robert W. Levis[4], Norbert Perrimon[3,6], Hugo J. Bellen[1,5]

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
2) Duncan Neurological Research Institute, Texas Children Hospital, Houston, TX
3) Department of Genetics Harvard Medical School, Boston, MA
4) Department of Embryology and Howard Hughes Medical Institute, Carnegie Institution for Science, Baltimore, MD
5) Department of Neuroscience, Baylor College of Medicine, Houston, TX
6) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA

Abstract

Previously, we described a large collection of Drosophila strains that each carry an artificial exon containing a *T2AGAL4* cassette inserted in an intron of a target gene based on CRISPR-mediated homologous recombination (Lee *et al.*, 2018). These alleles permit numerous applications and have proven to be very useful. Initially, the homologous recombination-based donor constructs had long homology arms (>500 bps) to promote precise integration of large constructs (>5kb). Recently, we showed that *in vivo* linearization of the donor constructs enables insertion of large artificial exons in introns using short homology arms (100-200 bps) (Kanca *et al.*, 2019a). Shorter homology arms make it feasible to commercially synthesize homology donors and minimize the cloning steps for donor construct generation. Unfortunately, about 50% of Drosophila genes lack suitable coding introns for integration of artificial exons. Here, we report the development of new set of constructs that allow the replacement of the coding region of genes that lack suitable introns with a *KozakGAL4* cassette, generating a knock-out/knock-in allele that expresses GAL4 similarly as the targeted gene. We also developed custom vector backbones to

further facilitate and improve transgenesis. Synthesis of homology donor constructs in custom plasmid backbones that contain the target gene sgRNA obviates the need to inject a separate sgRNA plasmid and significantly increases the transgenesis efficiency. These upgrades will enable the targeting of nearly every fly gene, regardless of exon-intron structure, with a 70-80% success rate.

Introduction

The Drosophila Gene Disruption Project (GDP) aims to generate versatile genetic tools for most genes to facilitate the study of gene function *in vivo* and to create fly stocks for the community. The CRISPR mediated integration cassette (CRIMIC) approach is a recent addition to the GDP to target fly genes. The CRIMIC strategy is based on integrating a Swappable Integration Cassette (SIC) containing an artificial exon encoding *attP-FRT-Splice Acceptor (SA)-T2AGAL4-polyA-3XP3EGFP-polyA-FRT-attP* (T2AGAL4). The SIC is integrated in an intron between two coding exons (coding intron) by CRISPR mediated homologous recombination (Lee *et al.*, 2018; Gnerer *et al.*, 2015; Diao *et al.*, 2015). The viral T2A sequence leads to the truncation of the nascent target gene polypeptide and re-initiation of translation of the downstream GAL4 as an independent protein. This cassette typically creates a strong loss of function allele of the targeted gene and expresses the yeast GAL4 transcription factor in a similar spatial and temporal pattern as the protein encoded by the targeted gene (Lee et al., 2018). These alleles can be used to: 1) determine the gene expression pattern; 2) study the effect of loss of function of the gene product; 3) replace the SIC through Recombinase Mediated Cassette Exchange (RMCE) (Bateman *et al.*, 2006; Venken *et al.*, 2011) with an artificial coding exon that encodes a fluorescent protein to assess protein subcellular localization (Venken et al., 2011) and identify interacting proteins ; 4) express UAS-cDNAs of the targeted gene and its variants to assess rescue of the mutant phenotype and conduct structure/function studies (Wangler *et al.*, 2017); 5) excise the insert with UAS-Flippase to revert the phenotype (Lee *et al.*, 2018).

The introduction of an artificial exon is only feasible for genes that contain a suitable large coding intron, typically 100 nt or more. This requirement makes nearly half of the genes inaccessible to strategies based on the use of artificial exons (Supplementary table 1; Figure 1A). In addition, the genes that do not have a suitable intron are typically smaller in size than genes that contain a suitable intron, and usually have fewer previously isolated, publicly available alleles than larger genes with a suitable intron.

Here, we describe the development of a knock-out/knock-in strategy to replace the coding sequence of genes with a *Kozak sequence-GAL4 polyA-FRT-3XP3GFP-polyA-FRT* (*KozakGAL4*) cassette to target genes that lack introns that are suitable for artificial exon knock-ins. The targeted gene is cut by Cas9 using two sgRNAs, one targeting the 5' untranslated region (UTR) and the other the 3' UTR. We targeted ~100 genes with this strategy and show that about 80% of the integrated KozakGAL4 cassettes lead to *UAS-mCherry* expression in the 3rd instar larval brain, a ratio that is similar to what was observed for the T2AGAL4 strategy (Lee *et al.*, 2018).

We also improved the design of the homology donor constructs that can be used for integration of either *KozakGAL4* or *T2AGAL4* cassettes. The use of short homology arms allows commercial DNA synthesis of the entire gene-specific portion of the donor plasmid, a cheaper and more efficient option than PCR-amplification and cloning of each homology arm (Kanca *et al.*, 2019a). To further extend this approach, we developed a method in which the DNA sequences directing the transcription of the target gene-specific sgRNA(s) are synthesized together in the same segment as the homology arms. The design allows gene specific sgRNA to be synthesized together with the homology arms, eliminating the need to clone and inject a separate vector for the sgRNAs. We tested our new designs on ~200 genes and show that the upgrades result in a transgenesis efficacy of ~80%. The strategies that we introduce here allow targeting of nearly every gene in the fly genome, further streamline the generation of homology donor DNAs, increase efficiency as compared to previous strategies, and improve the rate of precise genome editing.

Results and Discussion

The KozakGAL4 cassette

Integration of an artificial exon is only feasible for genes that have coding introns large enough to identify an sgRNA target site that is located at a sufficient distance from the preceding splice donor and the following splice acceptor site. Based on our experience, an intron should be larger than 100 nt to be suitable for integration of an artificial exon. An analysis of the Drosophila genome shows that 5787 out of 13,931 protein-coding genes have a sufficiently large coding intron that is shared among all the annotated splicing isoforms of the gene. However, 8144 genes lack such introns, making them inaccessible for the T2AGAL4 and other artificial exon-based strategies. Genes with a suitable intron typically have larger coding sequences (2051 nt vs 1321 nt) and have a larger number of previously isolated mutant alleles based on FlyBase data (10.4 vs 3.4) than genes without a suitable intron(s) (Figure 1A, Supplementary table 1; Larkin *et al.*, 2020). To integrate a GAL4 cassette that can be used in a similar manner as the *T2AGAL4* insertions, we developed the *KozakGAL4* knock-in/knock-out strategy. Kozak sequence is an optimal translation initiation site in eukaryotic mRNAs, and it is identified as (C/A) AA (C/A) AUG in Drosophila (Cavener, 1987; Kozak, 1986). We use CAAA as a Kozak sequence upstream of the start codon of GAL4. To replace the coding region of genes, we typically identify sgRNA target sites in the 5' UTR and 3' UTR (Figure 1B). To retain possible gene expression regulation by the 5'UTR, we select the upstream sgRNA target site that is closest to the start codon and that is not predicted to have off-target activity based on CRISPR Optimal Target Finder (Gratz *et al.*, 2014). The location of the downstream sgRNA target site in the 3'UTR is less stringent since the endogenous 3'UTR is not included in the final transcript due to the *polyA* signal in the *KozakGAL4* cassette. The median 5' and 3' UTR lengths for *Drosophila* genes are 214 and 224 bps respectively which are typically large enough to identify putative sgRNA targets (Chen *et al.*, 2011; Jan *et al.*, 2011). In our experience the 5'UTR typically contains multiple sgRNA targets whereas 3'UTRs contain few candidate sgRNA target sites due to their A/T rich nature. In cases where a suitable sgRNA target site cannot be found in the 3'UTR, we target a site within the coding region, close to the stop codon, to minimize the coding region of the gene that remains. In cases where a suitable sgRNA site cannot be found within the 5'UTR region, the

search is expanded to the promoter region and the sequence between the gRNA cut site and transcription start site is added to the homology region. In such cases, a single nucleotide substitution to eliminate the PAM sequence is introduced in the homology donor construct, preventing cutting of the homology donor.

We also developed alternative strategies to target genes without suitable introns and for which no proper sgRNA site could be identified within the 5'UTR or promoter. We generated the *SA-KozakGAL4-polyA-3XP3EGFP-polyA* cassette that can be introduced in an intron within the 5'UTR (Figure 1 Supplementary Figure1A). Alternatively, for genes with small coding introns, we make two cuts: one within a coding intron just upstream of the SA of an exon; the second in the 3'UTR. The excised sequence is then replaced with a T2AGAL4 cassette (Figure 1 Supplementary Figure 1B).

KozakGAL4 alleles drive expression of UAS-transgenes

There are two main approaches to generate alleles that express GAL4 that are not based on the T2AGAL4-based strategies. The first one is based on an insertion of minimal promoter-GAL4 coding sequences in a transposon backbone. The strategy is called enhancer trapping and was based originally on *GFP* and *LacZ* rather than *GAL4* (O'Kane and Gehring, 1987; Bellen *et al.*, 1989). Upon mobilization of the transposon, lines are established where the GAL4 expression pattern is of interest (Brand and Perrimon, 1993; Lukacsovich *et al.*, 2001; Hacker *et al.*, 2003; Gohl *et al.*, 2011). Given that they are inserted in the genome by transposable elements they are not always optimally placed to report the full expression pattern of the gene (Spradling *et al.*, 1995; Mayer *et al.*, 2013) but they have been used extensively as many reflect the expression pattern of a nearby gene (Wilson et al., 1989). Many however are not mutagenic (Spradling *et al.*, 1999). The second strategy to generate alleles that may express GAL4 in the expression pattern of a gene is to clone a 500bps-5kb region upstream of the promoter of the gene upstream of the GAL4 coding sequences and inserting the transgene in the genome. There are large collections of these enhancer-GAL4 alleles and most aim to report the expression of enhancer fragments rather than reporting the expression pattern of the gene from which they are derived (Jenett *et al.*, 2012; Manning *et al.*, 2012; Pfeiffer *et al.*,

2008). Hence, the available approaches can now be complemented with the KozakGAL4 approach that should incorporate all or most upstream regulatory information and generate a null allele of the targeted gene. The latter greatly facilitates rescue experiments using UAS-cDNA transgenes.

To assess the KozakGAL4 strategy, we targeted 109 genes to date and successfully replaced the coding region of 82 genes with the *KozakGAL4* cassette (Supplemental Table 2). We crossed 57 of these with *UAS-CD8mCherry* transgenic flies to determine the GAL4 expression of the targeted gene in the brain of wandering third instar larvae. Our previous findings, using *T2AGAL4* alleles have shown that ~80% of all *T2AGAL4* alleles lead to specific expression in third instar larval brains (Lee *et al.*, 2018). Similarly, with *KozakGAL4* alleles, we detected specific GAL4 expression for about 80% (46/57) of the genes (Figure 2A). Although *KozakGAL4* targeted genes are typically small, limiting the possible regulatory information in the coding region, it is possible that excision of coding and some UTR sequences may remove part of the regulatory input. We therefore tested whether a few *KozakGAL4* alleles drive the expression of the *UAS-CD8mCherry* in a similar pattern as the targeted gene. We selected a *KozakGAL4* allele that drives expression of the reporter in a restricted group of cells in the 3rd instar larval brain and analyzed the single cell RNA sequencing (scRNAseq) data for the 3rd instar larval brain to determine cell clusters that express the targeted gene (Ravenscroft *et al.,* 2020). We then used the same scRNAseq dataset to determine other genes expressed in overlapping clusters and that we previously targeted with *T2AGAL4*. *KozakGAL4* driven *UAS-CD8mCherry* reporter is expressed in a very similar expression pattern compared to *T2AGAL4* driven reporter expression of the genes that we identified through scRNAseq. The other genes expressed in the overlapping cluster according to scRNAseq, such as *serp*, *verm* and *emp* suggest that this cluster corresponds to tracheal cells, which is in line with the observed expression pattern through imaging (Figure 2B, Figure 2 Supplementary Figure 1B; Luschnig *et al.,* 2006; Lee *et al.*, 2018). Comparison of the expression patterns of the other tested *KozakGAL4* targeted genes and *T2AGAL4* targeted genes that are expressed in overlapping cell groups showed overlapping expression patterns based on imaging as well (Figure 2 Supplementary Figure 1A). Hence, the use of scRNAseq data can provide an independent means of verification of accuracy of the observed reporter expression patterns.

Next, we determined if UAS-human or fly cDNAs could rescue the phenotype associated with 11 gene deletions caused by the *KozakGAL4* knock-ins. For *pngl*, *Wdr37, Tom70, CG8320, CG16787* and *IntS11* a UAS-fly or human cDNA rescued the *KozakGAL4* induced phenotypes, suggesting that the *KozakGAL4* is expressed where the targeted gene product is required for the gene function. For *pex2, pex16, fitm, PIG-A* and *CG34293* the expression of orthologous human cDNA did not rescue the associated phenotypes.

In summary, *KozakGAL4* offers a new means to disrupt gene function while expressing the GAL4 in the expression domain of the targeted genes. This approach allows us to tag the remainder of the genes that do not contain a suitable coding intron for the T2AGAL4 strategy which corresponds to 58% of all the genes.

New vector backbones for synthesis of homology donor constructs that are also templates for sgRNA expression

We previously showed that linearizing the homology donor constructs *in vivo* allows for integration of large constructs in the genome through CRISPR-mediated homologous recombination even using short homology arms (Kanca *et al.,* 2019a). This approach makes inexpensive commercial synthesis of homology donor intermediates feasible. The intermediate vectors can be used for a single step directional cloning of the SIC in the homology donor intermediate vector. This greatly facilitates the generation of homology donor vectors which previously required four-way ligations with large homology arms. Moreover, this eliminates cloning failures (~20-30%) and troubleshooting associated constructs with large homology arms (Kanca *et al.*, 2019a). The resulting new homology donor vectors were previously injected together with two vectors that express two sgRNAs (pCFD3, Port *et al.*, 2014) in embryos that express Cas9 in their germline. The first sgRNA targets the homology donor vector backbone to linearize the homology donor and does not have a target in the Drosophila genome (sgRNA1, Garcia-Marques *et al.*, 2019). The second sgRNA vector expresses the sgRNA to target the gene and introduce the double strand DNA break that serves as a substrate for homologous recombination. In Kanca *et al.* (2019a) we demonstrated that injection of these constructs resulted in transgenesis efficiencies of about 60%.

We developed new approaches to increase the transgenesis efficiency of the custom DNA backbones, decrease the workload, and to simplify the generation of homology donor constructs. The first custom vector backbone that we tested has the U6-3::sgRNA1 sequence in the vector backbone and sgRNA1 targets, on either side of the EcoRV site where the synthesized fragments are directionally integrated (vector backbone named pUC57_Kan_gw_OK, design named int200, Figure 3A). With this design, the homology donor vector intermediates that are commercially synthesized contain the sgRNA1 coding sequence, obviating the need to co-inject one of the sgRNA vectors. Having the sgRNA1 coding region in the backbone also helps with *in vivo* linearization of the homology donor since the homology donor construct and the sgRNA1 are delivered together in a single vector. The int200 design also removes the sgRNA1 target sites from the synthesized region as they are present in the vector. This allows increasing the homology arm length to 200 bps without increasing the cost of synthesis.

We tested this int200 design for 397 genes with T2AGAL4 cassettes and 36 genes with KozakGAL4 cassette. For each construct, we injected 400-600 embryos that express Cas9 in the germline. For inserting the KozakGAL4 cassette, the two gene specific sgRNAs were cloned into pCFD5 (Port and Bullock, 2016). For both T2AGAL4 and KozakGAL4 insertions the int200 homology donor plasmid was co-injected with the plasmid that encodes the target specific sgRNA (pCFD3 for the former and pCFD5 for the latter case)**.** We successfully integrated T2AGAL4 cassette in 252 genes (~65% success rate) and replaced the coding region of 22 genes with KozakGAL4 (~61% success rate) (Figure 3B). PCR verification of the inserts was performed by using gene specific PCR primers outside the homology region pointing towards the insert and a construct specific PCR primer. For 88% of the T2AGAL4 inserts we obtained PCR verification on both sides of the insert and for the remaining 12% we obtained PCR products on one side of the construct. For 91% of KozakGAL4 inserts we obtained amplicons on both sides of the insert. For the inserts with a single PCR verification, we sequenced the amplicon to ensure the insert is in the proper locus. Hence, the overall transgenesis success rate of the int200 method is about 65% (Figure 3B). This is very similar to the injection success rate of homologous recombination using large (0.5-1kb) homology arms (1165

insertions in 1784 targeted genes, Lee *et al.*, 2018) but leads to very significant reductions in labor and cost. Additionally, int200 facilitates the cloning of homology donor constructs and eliminates cloning failures which reduce the overall successful targeting rate using large homology arms to ~50% (successful cloning of 80% constructs that are injected with 65% transgenesis success rate). In summary, the int200 method provides a ~30% gain in overall efficiency (from 50% to 65%).

To further optimize the custom vector backbone we repositioned the U6-3::sgRNA1and added a partial tRNA construct directly upstream of the EcoRV site that is used to insert the synthesized fragments (Figure 4A). The partial tRNA allows adding the gene-specific sgRNA sequence (vector named pUC57_Kan_gw_OK2 and design named gRNA_int200 for T2AGAL4 constructs and named 2XgRNA_int200 for KozakGAL4 constructs). Hence, two or three sgRNA can be produced from the single injected plasmid. One of the sgRNA1 target sites is added to the synthesized fragment before the start of the homology arm and the other sgRNA1 target site is added to the backbone just downstream of the EcoRV site where the synthesized fragment is directionally inserted. This design obviates the need to clone a separate sgRNA vector to target the genomic locus. It also ensures simultaneous delivery of all the components of the homologous recombination reaction as they are delivered on a single plasmid. We have targeted 127 genes with gRNA_int200_T2AGAL4 donor plasmids (Supplemental Table 2) and successfully inserted the T2AGAL4 cassette in 95 genes (~75% success rate, Figure 4B). We also tested whether genes for which the tagging failed using the int200 strategy (Figure 3) could be targeted with the gRNA_int200_T2AGAL4 using the same gene-specific sgRNA and homology arms. For 3 out of 4 genes tested, use of the gRNA_int200 strategy resulted in successful integration of T2AGAL4 cassette. These data show that incorporating all the sgRNAs in the donor vector improves the transgenesis efficiency.

For the KozakGAL4 constructs, we inserted a second tRNA sequence after the first synthesized sgRNA and added the second gene specific sgRNA sequence to the synthesis reaction. We targeted 72 genes with 2XgRNA_int200_KozakGAL4 cassette and successfully inserted KozakGAL4 in 59 genes (~82% success rate, Figure 4B; Supplemental Table 2). We tested whether genes for which the tagging

failed using the int200_KozakGAL4 strategy (Figure 3) could be properly targeted with the 2XgRNA_int200_KozakGAL4 using the same gene specific sgRNAs and homology arms sequences and again observed that for 3 out of 4 tested genes, the 2XgRNA_int200_KozakGAL4 strategy was successful. In summary, the gRNA_int200 design increases transgenesis rate and streamlines the creation of T2AGAL4 and KozakGAL4 constructs by obviating the need to generate a separate sgRNA expression plasmid and ensuring co-delivery of all components for homologous recombination. In summary, the gRNA_int200 allows a 78% transgenesis success rate, or an additional 20% when compared to the int200 approach (increase from 65% to 78%).

Use of 2XgRNA_int200 intermediate vectors for GFP tagging

We have previously shown that integrating a *SA-linker-EGFP-FlAsH-StrepII-3xTEVcs-3xFLAG-linker-SD* (*SA-GFP-SD*) in coding introns of genes is an efficient approach to tag proteins with GFP (Venken *et al.*, 2011; Nagarkar-Jaiswal *et al.,* 2015a; 2015b; 2017; Li-Kroeger *et al.*, 2018). We typically generate these alleles through Recombinase Mediated Cassette Exchange of existing MiMIC SICs and we have shown that they are functional in 72% of tested genes (Nagarkar-Jaiswal *et al.*, 2015a). A major factor that affects the functionality of the GFP protein trap is the insertion position. In cases where the artificial exon encoding for protein trap is inserted in a coding intron that bisects a predicted functional protein domain, the resulting protein trap is often not functional. Hence, another efficient approach to tag proteins encoded by genes that have no intron, small introns or no suitable MiMICS in any preselected position in the protein structure is highly desirable.

We tested the use of synthesized homology donor intermediate vectors to replace the coding sequence of genes without suitable coding introns with the gene coding sequence fused to GFP at different locations. We selected the *Wdr37* gene as it has a small intron (Kanca *et al.*, 2019b). We amplified *Wdr37* sequences from the genome by PCR and used NEB HiFi DNA assembly to generate homology donor constructs where a sfGFP tag is integrated at the N terminus, C terminus or internally (Figure 5, Figure 5 Supplementary figure 1,2,3 for schematics of HiFi assembly). The 3XP3 DsRed flanked by PiggyBac transposase inverted repeats is integrated after the 3'UTR and serves as the transformation marker that can be

excised precisely using the PiggyBac transposase (flyCRISPR.molbio.wisc.edu; Bier *et al.*, 2018) (Figure 5A). The assembled sequences are subcloned in the synthesized homology donor intermediate. Injection of the homology donor plasmids in embryos expressing Cas9 in their germline resulted in positive transgenics in each case. Western blot of the resulting protein trap alleles using anti-GFP antibody detected bands at the expected length for the tagged protein in each case. However, the internally tagged allele is less abundant, underlining that the placement of sfGFP tag can affect protein stability (Figure 5B). Hence, the strategy to replace the whole coding region with a GFP tagged coding region allows tagging almost any gene in any position in the coding sequence.

In summary, we developed a *KozakGAL4* strategy to target the genes that do not have a suitable intron and a set of novel custom vector backbones to facilitate homology donor construct production and increase transgenesis rate. The methods we developed are versatile and can be modified to generate GAL4 gene traps or GFP protein fusions of the targeted genes. Finally, the methodology we describe should be easy to implement in any other model organisms to facilitate generation of gene trap and protein trap alleles.

Acknowledgements

We thank the Bloomington *Drosophila* Stock Center and FlyBase for fly stocks acquisition and distribution and the Developmental Studies Hybridoma Bank for antibodies. We thank Stephanie Mohr for critical reading of the manuscript. We thank Yuchun He, Ying Fang, Minhua Huang, Zhihua Wang, Yaping Yu, Junyan Fang, Ruifang Zhang, and Lily Wang for generating and maintaining MiMIC/CRIMIC *T2A-GAL4* fly stocks. Confocal microscopy was performed in the BCM IDDRC Neurovisualization Core, supported by the NICHD (U54HD083092). HJB received support from NIH R01GM067858, R24 OD031447, R24OD022005, U54NS093793 HJB and the Huffington Foundation. OK receive support from NIH R24 OD031447. JZ, YHu, and NP receive support from NIGMS (GM067761 and GM084947). NP is an investigator of the Howard Hughes Medical Institute. RWL was supported by the Carnegie Institution for Science.

Materials and Methods

Generation of homology donor constructs

Templates for ordering the int200 and gRNA_int200 constructs can be found in Supplementary Materials and Methods. Homology donor intermediate vectors were ordered for production from Genewiz ("ValueGene" option) in pUC57 Kan_gw_OK (for int200 strategy) or pUC57 Kan_gw_OK2 (for gRNA_int200 strategy) vector backbone at 4 μg production scale. The lyophilized vectors were resuspended in 53 μl of ddH_2O. 1 ul was used for Golden Gate assembly with 290 ng of pM37 vector of reading frame phase corresponding to the targeted intron (for T2GAL4, Lee *et al.*, 2018) or 265 ng of pM37_KozakGAL4 vector (for KozakGAL4). The Golden Gate Assembly reaction was set in 200μl PCR tubes (ThermoScientific AB2000) with 2.5 μl 10X T4 DNA ligase buffer (NEB B0202S), 0.5 μl T4 DNA ligase (NEB M0202L), 1 μl restriction enzyme (BbsI_HF or BsaI_HFv2 NEB R3559L and R3733L respectively), 1 μl of SIC (pM37_T2AGAL4 or pM37_KozakGAL4 at 290 ng/ μl or 265 ng/ μl respectively), 19 μl of dH_2O and 1 μl of homology donor construct. For cloning multiple constructs in parallel, master mixes were prepared including all the components except for the homology donor intermediate vector. The reactions were incubated in a Thermocycler (cycle 30 times between 37°C 5 minutes, 16°C 5 minutes, then 65°C 20 minutes, 8°C hold). An additional digestion step was done to remove self ligating plasmid backbones by adding 19.5 μl dH_2O, 5 μl 10X CutSmart buffer (NEB B7204S) and 0.5 μl BbsI or BsaI_HFv2 (the enzyme used for the cloning reaction). The reaction product was transformed in DH5α competent cells and plated on Kanamycin+ LB plates.

Fly injections

Int200-T2AGAL4 and int200-KozakGAL4 constructs were injected at 250 ng/μl along with 100ng/μl gene specific gRNA(s) cloned in pCFD3 or pCFD5 respectively (Port *et al.* 2014; Port and Bullock, 2016). Injections were performed as described in (Lee *et al.*, 2018). 400-600 embryos from *y^1w^*; iso18; attP2(y+){nos-Cas9(v+)}* for genes on the 2nd or 4th chromosome and *y^1w^* iso6*;; *attP2(y+){nos-Cas9(v+)}* for genes on the X chromosome and *y^1w^*; attP40(y+){nos-Cas9(v+)}; iso5* (Kondo and Ueda, 2013) for genes on the 3rd chromosome per genotype were injected. Whole genome

sequencing BAM files of isogenized lines can be found at: https://zenodo.org/record/1341241. Resulting G0 males and females were crossed individually to *y*^{1} *w** flies as single fly crosses for *3XP3-EGFP* detection. Positive lines were balanced, and stocks were established. Up to 5 independent lines were generated per construct per gene. The list of generated alleles can be found on Supplementary table 2. The sequences of homology arms and sgRNA(s) as well as the results of PCR validation and imaging on third instar larval brain are available at http://flypush.imgen.bcm.tmc.edu/pscreen/crimic/crimic.php. The stocks are deposited in the Bloomington Drosophila Stock Center (BDSC) on a regular basis. The stocks are available from the Bellen lab until they are deposited and established in the BDSC.

PCR validation

PCR primers that flank the integration site were designed for each targeted gene. 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, Primer sequences can be found in the supplementary material). 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 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.

Confocal imaging of transgenic larval brains

Dissection and imaging were performed following the protocols in (Lee *et al.*, 2018). 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.

Analysis of single cell sequencing data

To identify genes with expression profiles that overlap with expression of genes replaced with *KozakGAL4* sequences, we queried the data from third instar larval CNS scRNAseq data described in Ravenscroft *et al.* (2020). The data (http://scope.aertslab.org/#/Larval_Brain/*/welcome) were imported into Seurat (version 4.0.1). Cells expressing the selected genes, for which the KozakGAL4 allele was generated (e.g *CG3770*, *CG10939*, *CG10947* and *CG15093*), were identified using WhichCells function and genes enriched in these cells were identified using FindMarkers with default parameters. A list of the top 10 genes that were minimally expressed outside the expression domain of genes with KozakGAL4 alleles was generated. We then selected genes from the list for which T2AGAL4 were generated and compared the expression profiles using available images.

Western blots

Flies were homogenized using Cell Lysis Buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1% Triton-X 100, 1X liquid protease inhibitor (Gen DEPOT), 0.1 M DTT). The supernatant was collected after centrifugation at 13, 000 rpm for 10 min at 4°C (Eppendorf 5424R with rotor Eppendorf FA-45-24-11). The supernatant was mixed with Laemmli Buffer containing β-mercaptoethanol and heated at 95°C for 10 min. Subsequently, the samples were loaded in 4–20% gradient polyacrylamide gels (Bio-Rad Mini-PROTEAN® TGX™). Following electrophoresis, proteins were transferred onto a polyvinylidene difluoride membrane (Immobilon, Sigma). The membrane was blocked using skimmed milk and treated with the primary antibody for overnight. The following antibodies were used in the present study: rabbit anti-GFP (1:1000) (Thermo Fisher Scientific, #A-11122), mouse anti-Actin (1:5000) (EMD Millipore, #MAB1501). Horseradish peroxidase-conjugated secondary antibody was used to detect the respective primary antibody. Blots were imaged on a Bio-Rad ChemiDocMP.

Cloning of *Wdr37-KIGFP* constructs

The fragments that position the GFP tag to the selected sites were PCR amplified from genomic DNA, sfGFP was amplified from pBS_SA_sfGFP_SD (Kanca *et al.*, 2019a) and scarless DSred from pScarlessHD-DsRed (pScarlessHD-DsRed was a gift from Kate O'Connor-Giles (Addgene plasmid # 64703 ;

http://n2t.net/addgene:64703 ; RRID:Addgene_64703). The fragments were used together with homology donor intermediate for *Wdr37* gene used for generating *KozakGAL4* allele (CR70111) cut with BbsI-HF to assemble NEB-HiFi DNA assembly following manufacturer's instructions. Schematics of HiFi assembly can be found in Figure 5 Supplementary figures 1, 2 and 3.

Figure Legends

Figure 1. KozakGAL4 strategy can be used to generate GAL4 gene trap alleles for approximately 50% of Drosophila genes. (A) Analysis of Drosophila genome indicates that about half of the Drosophila genes do not have a suitably large coding intron for insertion of a T2AGAL4 cassette. These genes are on average shorter and have fewer genetic reagents compared to the genes that have a suitably large coding intron for inserting T2AGAL4 cassette. (B) Schematics of the KozakGAL4 targeting. Gray boxes, UTRs; orange box, gene coding region.

Figure 2. KozakGAL4 alleles document intricate gene expression patterns in third instar larval brains. (A)Examples of third instar larval brain gene expression patterns obtained by crossing KozakGAL4 allele of indicated genes with *UAS-CD8mCherry* flies. (B) The imaging results of reporter expression generated with *KozakGAL4* allele were compared to the expression pattern of genes that are expressed in similar cells by analysis of single cell sequencing data imaged using *T2AGAL4* alleles. Images are taken by crossing the *GAL4* alleles with *UAS-CD8mCherry.* Arrowheads point to the shared expression pattern.

Figure 3. int200 strategy results in similar transgenesis success rates as the long homology arms CRIMICs. (A) Schematics of the int200 strategy. (B) Transgenesis data using int200_T2AGAL4 or int200_KozakGAL4 strategies.

Figure 4. gRNA_int200 strategy increases the transgenesis success rates. (A) Schematics of the gRNA_int200 strategy. (B) Transgenesis data using gRNA_int200_T2AGAL4 or 2XgRNA_int200_KozakGAL4 strategies.

Figure 5. 2XgRNA_int200 strategy can be used to tag any gene at any coding region to generate protein trap alleles. (A) Schematics of the targeting constructs

to integrate sfGFP protein tag at an N-terminal, internal or C-terminal location in *Wdr37* gene locus. (B) Western blot analysis from adult flies show full length protein in all protein trap alleles with the arrow indicating the 81 kDa band that is the length predicted for the Wdr37 protein fused to sfGFP.

Supplementary Figure 1. Alternative strategies to generate gene trap alleles in genes without suitable introns. For genes that cannot be targeted by artificial exon strategies and where suitable sgRNAs could not be found in the 5'UTR an artificial exon with SA_KozakGAL4 can be inserted in an intron in the 5'UTR (A) or in a short intron by deleting the exons following the intron (B).

Supplementary Figure 2. Identification of genes expressed in similar cells to the *KozakGAL4* alleles expressed in restricted patterns. (A) The single cell sequencing data from Ravenscroft *et al.* 2020 with the cells expressing the gene targeted by KozakGAL4 marked by red circles and cells expressing the gene targeted with T2AGAL4 allele marked by green circles. The imaging results of reporter expression generated with *KozakGAL4* alleles were compared to the expression patterns of genes that are expressed in similar cells by analysis of single cell sequencing data imaged using *T2AGAL4* alleles. Images are taken by crossing the *GAL4* alleles with *UAS-CD8mCherry.* Arrowheads show the regions with the most overlap. (B) Cluster of trachea markers in the scRNA data from Ravenscroft *et al.* 2020.

Supplementary Table 1. Analysis of Drosophila genome for the presence of suitable introns.

Supplementary Table 2. List of the 428 alleles generated in this study
The alleles are indicated in tabs corresponding to the strategy used to generate the allele.

References

Bateman, J. R., Lee, A. M. and Wu, C. T. (2006) 'Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange', *Genetics,* 173(2): pp. 769-77.

Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, R. K. and Gehring, W. J. (1989) 'P-element-mediated enhancer detection: a versatile method to study development in Drosophila', *Genes Dev,* 3(9): pp. 1288-300.

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

Brand, A. H., and Perrimon N. (1993) 'Targeted gene expression as a means of altering cell fates and generating dominant phenotypes', *Development*, 118: pp. 401-15.

Cavener, D. R. (1987) 'Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates', *Nucleic Acids Res,* 15(4): pp. 1353-61.

Chen, C. H., Lin, H. Y., Pan, C. L. and Chen, F. C. (2011) 'The plausible reason why the length of 5' untranslated region is unrelated to organismal complexity', *BMC Res Notes,* 4: pp. 312.

Diao, F., Ironfield, H., Luan, H., Shropshire, W. C., Ewer, J., Marr, E., Potter, C. J., Landgraf, M. and White, B. H. (2015) 'Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes', *Cell Rep,* 10(8): pp. 1410-21.

Garcia-Marques, J., Yang, C. P., Espinosa-Medina, I., Mok, K., Koyama, M. and Lee, T. (2019) 'Unlimited Genetic Switches for Cell-Type-Specific Manipulation', *Neuron,* 104(2): pp. 227-238.e7.

Gloor, G. B., Preston, C. R., Johnson-Schlitz, D. M., Nassif, N. A., Phillis, R. W., Benz, W. K., Robertson, H. M. and Engels, W. R. (1993) 'Type I repressors of P element mobility', *Genetics,* 135(1): pp. 81-95.

Gnerer, J. P., Venken, K. J. and Dierick, H. A. (2015) 'Gene-specific cell labeling using MiMIC transposons', *Nucleic Acids Res,* 43(8): pp. e56.

Gohl, D. M., Silies, M. A., Gao, X. J., Bhalerao, S., Luongo, F. J., Lin. C-C., Potter, C. J.. Clandinin, T. R. (2011) 'A versatile *in vivo* system for directed dissection of gene expression patterns', *Nat Methods,* 8: 231–7.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., O'Connor-Giles, K. M. (2014) 'Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila', *Genetics*, 196(4): pp. 961-71.

Hacker, U., Nystedt S., Barmchi, M. P., Horn, C., Wimmer, E. A. (2003) 'piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in Drosophila', *Proc. Natl. Acad. Sci.*, 100(13): pp. 7720–5.

Jan, C. H., Friedman, R. C., Ruby, J. G. and Bartel, D. P. (2011) 'Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs', *Nature,* 469(7328): pp. 97-101.

Jenett, A., Rubin, G. M., Ngo, T. T., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, J. H., Peng, H., Trautman, E. T., Svirskas, R. R., Myers, E. W., Iwinski, Z. R., Aso, Y., DePasquale, G. M., Enos, A., Hulamm, P., Lam, S. C., Li, H. H., Laverty, T. R., Long, F., Qu, L., Murphy, S. D., Rokicki, K., Safford, T., Shaw, K., Simpson, J. H., Sowell, A., Tae, S., Yu, Y. and Zugates, C. T. (2012) 'A GAL4-driver line resource for Drosophila neurobiology', *Cell Rep,* 2(4): pp. 991-1001.

Kanca, O., Zirin, J., Garcia-Marques, J., Knight, S. M., Yang-Zhou, D., Amador, G., Chung, H., Zuo, Z., Ma, L., He, Y., Lin, W. W., Fang, Y., Ge, M., Yamamoto, S., Schulze, K. L., Hu, Y., Spradling, A. C., Mohr, S. E., Perrimon, N. and Bellen, H. J. (2019a) 'An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms', *Elife,* 8: e51539.

Kanca, O., Andrews, J.C., Lee, P.T., Patel, C., Braddock, S.R., Slavotinek, A.M., Cohen, J.S., Gubbels, C.S., Aldinger, K.A., Williams, J., Indaram, M., Fatemi, A., Yu, T.W., Agrawal, P.B., Vezina, G., Simons, C., Crawford, J., Lau, C.C.; Undiagnosed Diseases Network, Chung, W.K., Markello, T.C., Dobyns, W.B., Adams, D.R., Gahl, W.A., Wangler, M.F., Yamamoto, S., Bellen, H.J., Malicdan, M.C.V. (2019b) '*De novo* variants in *WDR37* are associated with epilepsy, colobomas, dysmorphism, developmental delay, intellectual disability, and cerebellar hypoplasia', *Am J of Hum Genet,* 105: pp. 413-24.

Kondo, S. and Ueda, R. (2013) 'Highly improved gene targeting by germline-specific Cas9 expression in Drosophila', *Genetics,* 195(3): pp. 715-21.

Kozak, M. (1986) 'Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes', *Cell,* 44(2): pp. 283-92.

Larkin, A., Marygold, S.J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P.V., Goodman, J.L., Gramates, L.S., Millburn, G., Strelets, V.B., Tabone, C.J., Thurmond, J. and the FlyBase Consortium (2021) 'FlyBase: updates to the Drosophila melanogaster knowledge base', Nucleic Acids Res. 49(D1): D899–D907.

Lee, P. T., Zirin, J., Kanca, O., Lin, W. W., Schulze, K. L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., Fang, Y., He, Y., Pan, H., Ge, M., Zuo, Z., Housden, B. E., Mohr, S. E., Yamamoto, S., Levis, R. W., Spradling, A. C., Perrimon, N. and Bellen, H. J. (2018) 'A gene-specific T2A-GAL4 library for Drosophila', *Elife,* 7: e35574.

Li-Kroeger D., Kanca O., Lee P-T, Cowan S., Lee M. T., Jaiswal M., Salazar, J. L., He Y., Zuo Z., Bellen H. J. (2018) 'An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in Drosophila', *Elife*, 7: e38709.

Lukacsovich, T., Asztalos, Z., Awano, W., Baba, K., Kondo, S., Niwa, S., Yamamoto, D. (2001) 'Dual-tagging gene trap of novel genes in Drosophila melanogaster', *Genetics*, 157: pp. 727–42.

Luschnig, S., Bätz, T., Armbruster, K. and Krasnow, M. A. (2006) 'serpentine and vermiform encode matrix proteins with chitin binding and deacetylation domains that limit tracheal tube length in Drosophila', *Curr Biol,* 16(2): pp. 186-94.

Manning, L., Heckscher, E. S., Purice, M. D., Roberts, J., Bennett, A. L., Kroll, J. R., Pollard, J. L., Strader, M. E., Lupton, J. R., Dyukareva, A. V., Doan, P. N., Bauer, D. M., Wilbur, A. N., Tanner, S., Kelly, J. J., Lai, S. L., Tran, K. D., Kohwi, M., Laverty, T. R., Pearson, J. C., Crews, S. T., Rubin, G. M. and Doe, C. Q. (2012) 'A resource for manipulating gene expression and analyzing cis-regulatory modules in the Drosophila CNS', *Cell Rep,* 2(4): pp. 1002-13.

Mayer, L. R., Diegelmann, S., Abassi, Y., Eichinger, F. and Pflugfelder, G. O. (2013) 'Enhancer trap infidelity in Drosophila optomotor-blind', *Fly (Austin),* 7(2): pp. 118-28.

Nagarkar-Jaiswal S., Lee P-T, Campbell M. E., Chen K., Anguiano-Zarate S.,Gutierrez M. C., Busby T., Lin W. W., He Y., Schulze K. L., Booth B. W., Evans-Holm M., Venken K. J., Levis R. W., Spradling A. C., Hoskins R. A., Bellen H. J. (2015a) 'A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila', *Elife,* 4: e05338.

Nagarkar-Jaiswal S., DeLuca S. Z., Lee P-T, Lin W. W., Pan H., Zuo Z., Lv J., Spradling A. C., Bellen H.J. (2015b) 'A genetic toolkit for tagging intronic MiMIC containing genes', *Elife,* 4: e08469.

Nagarkar-Jaiswal S., Manivannan S. N., Zuo Z., Bellen H. J. (2017) 'A cell cycle-independent, conditional gene inactivation strategy for differentially tagging wild-type and mutant cells', *Elife,* 31: e26420.

O'Kane, C. J. and Gehring, W. J. (1987) 'Detection in situ of genomic regulatory elements in Drosophila', *Proc Natl Acad Sci U S A,* 84(24): pp. 9123-7.

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T. T., Misra, S., Murphy, C., Scully, A., Carlson, J. W., Wan, K. H., Laverty, T. R., Mungall, C., Svirskas, R., Kadonaga, J. T., Doe, C. Q., Eisen, M. B., Celniker, S. E. and Rubin, G. M. (2008) 'Tools for neuroanatomy and neurogenetics in Drosophila', *Proc Natl Acad Sci U S A,* 105(28): pp. 9715-20.

Port, F. and Bullock, S. L. (2016) 'Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs', *Nat Methods,* 13(10): pp. 852-4.

Port, F., Chen, H. M., Lee, T. and Bullock, S. L. (2014) 'Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila', *Proc Natl Acad Sci U S A,* 111(29): pp. E2967-76.

Ravenscroft, T.A., Janssens, J., Lee, P.T., Tepe, B., Marcogliese, P.C., Makhzami, S., Holmes, T., Aerts, S., Bellen, H.J. (2020) '*Drosophila* voltage-gated sodium

channels are only expressed in active neurons and are localized to distal axonal initial segment-like domains', *J Neurosci,* 40: pp. 7999-8024.

Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T. and Rubin, G. M. (1995) 'Gene disruptions using P transposable elements: an integral component of the Drosophila genome project', *Proc Natl Acad Sci U S A,* 92(24): pp. 10824-30.

Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverty, T., Mozden, N., Misra, S. and Rubin, G. M. (1999) 'The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes', *Genetics,* 153(1): pp. 135-77.

Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A. and Bellen, H. J. (2011) 'MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes', *Nat Methods,* 8(9): pp. 737-43.

Wangler, M. F., Yamamoto, S., Chao, H. T., Posey, J. E., Westerfield, M., Postlethwait, J., Hieter, P., Boycott, K. M., Campeau, P. M., Bellen, H. J. and (UDN), M. o. t. U. D. N. (2017) 'Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research', *Genetics,* 207(1): pp. 9-27.

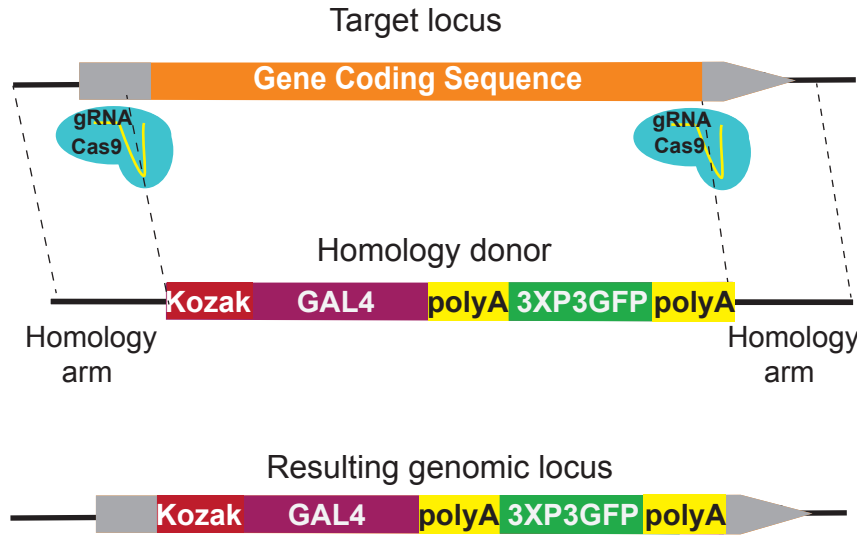
Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C. J., Grossniklaus, U. and Gehring, W. J. (1989) 'P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila', *Genes Dev,* 3(9): pp. 1301-13.

Kanca et al. 2021 Figure 1. KozakGAL4 strategy

A

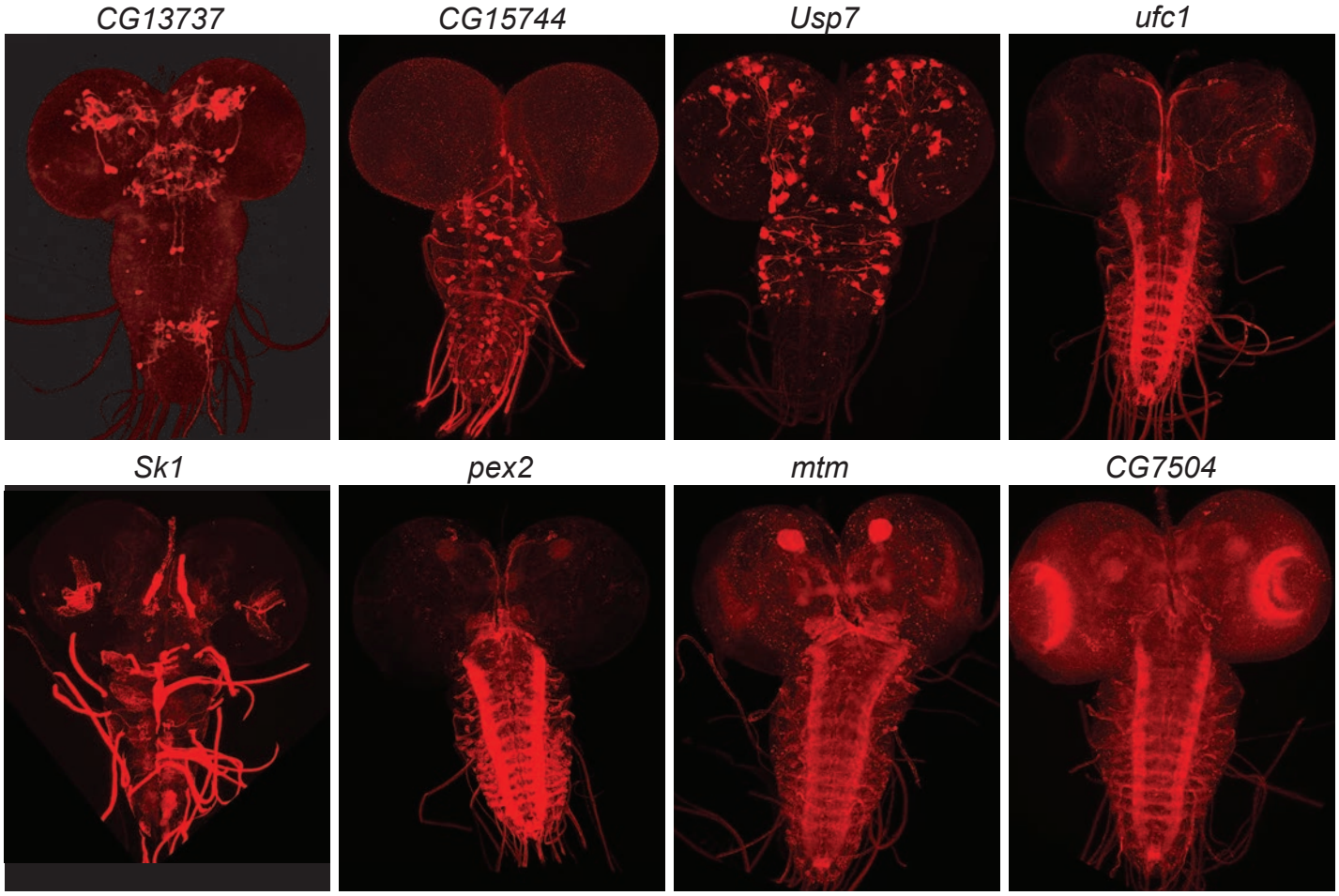
CDS length	with suitable intron	without suitable intron
<500	412	759
500-1000	1118	1530
1000-2000	2262	1887
2000-3000	965	554
3000-4000	479	151
4000-5000	241	54
>=5000	310	62
avg	2051	1321
# of genes	5787	4997

B

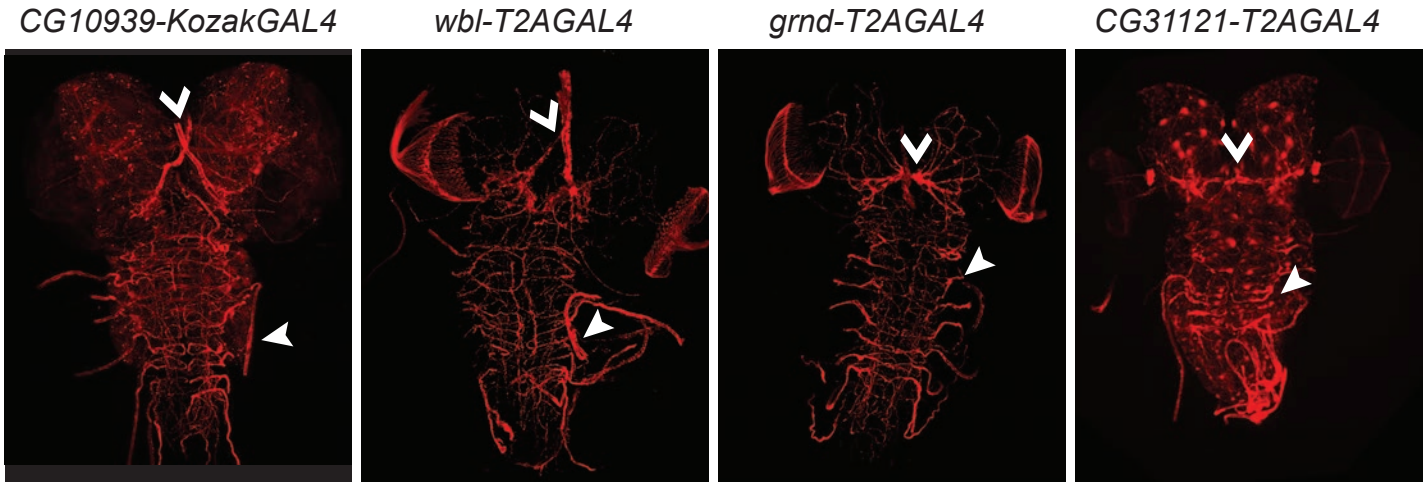


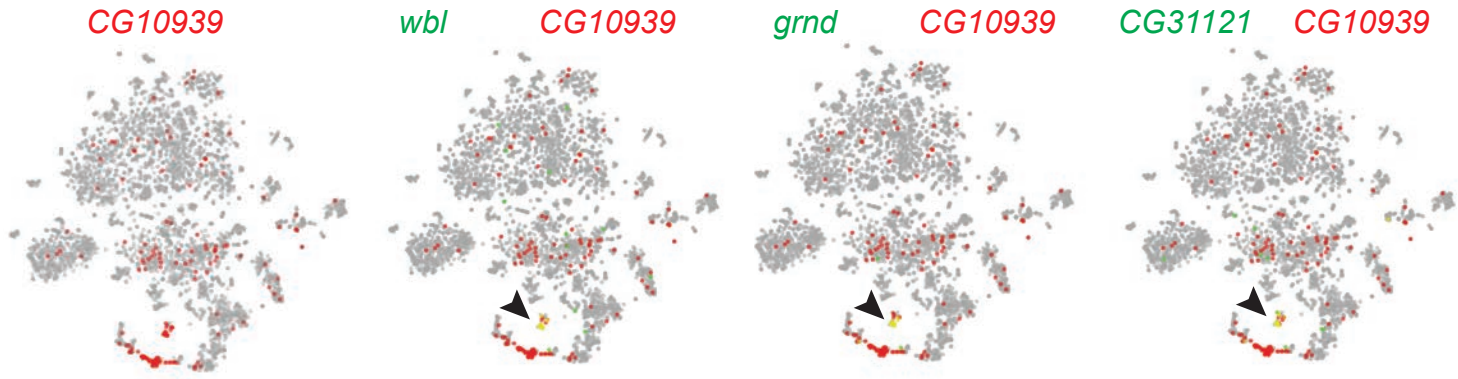
Kanca et al. 2021 Figure 2. KozakGAL4 expression

A *KozakGAL4* alleles document intricate gene expression patterns in 3rd instar larval brain



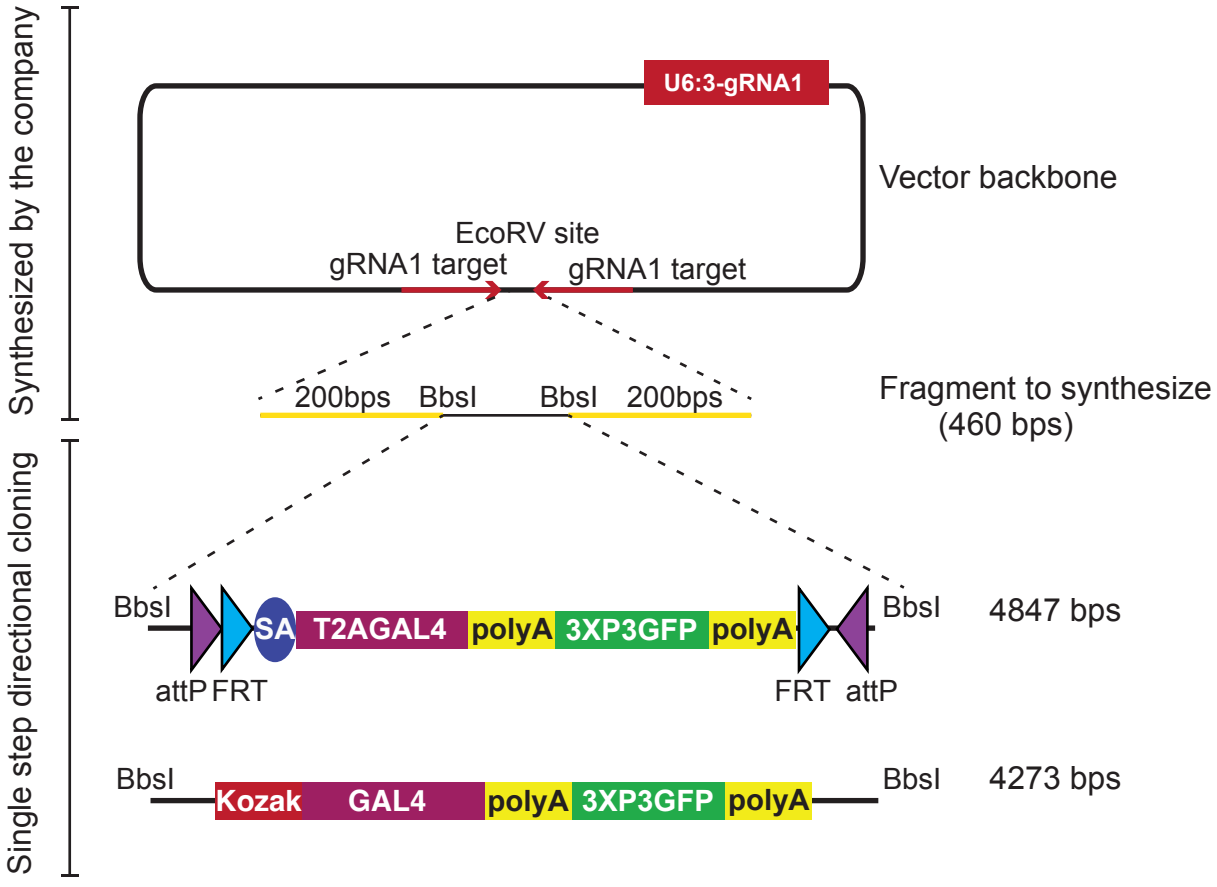
B *KozakGAL4* patterns overlap with *T2AGAL4* patterns of genes in the overlapping single cell sequencing clusters





Kanca et al. 2021 Figure 3. int200 strategy

A Schematics of int200 strategy

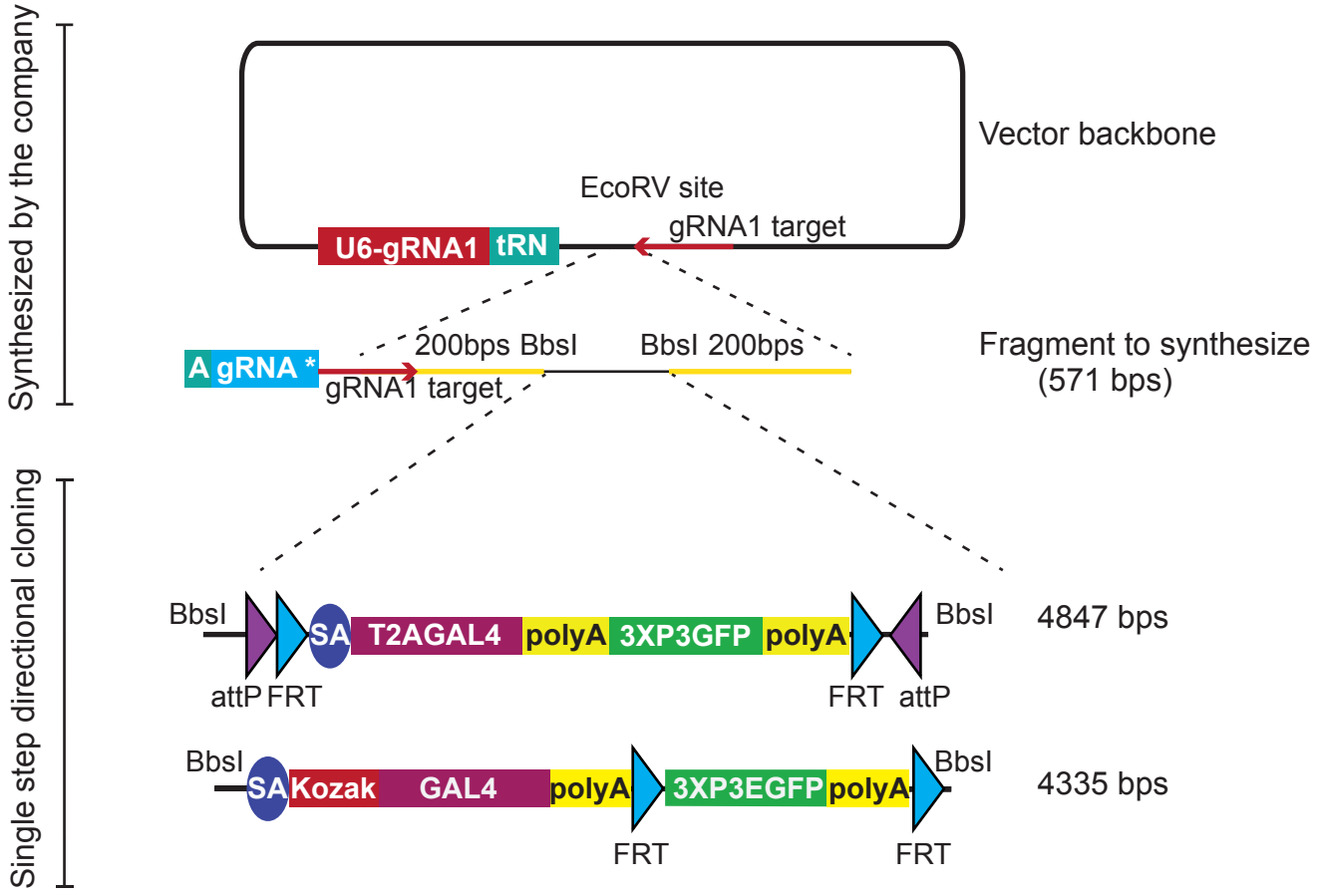


B Transgenesis rates of int200 strategy

int200 strategy	Genes targeted	Successful integration	Transgenesis Rate
T2AGAL4	397	252	63.48
KozakGAL4	36	22	61.11

Kanca et al. 2021 Figure 4. gRNA_int200 strategy

A Schematics of gRNA_int200 strategy

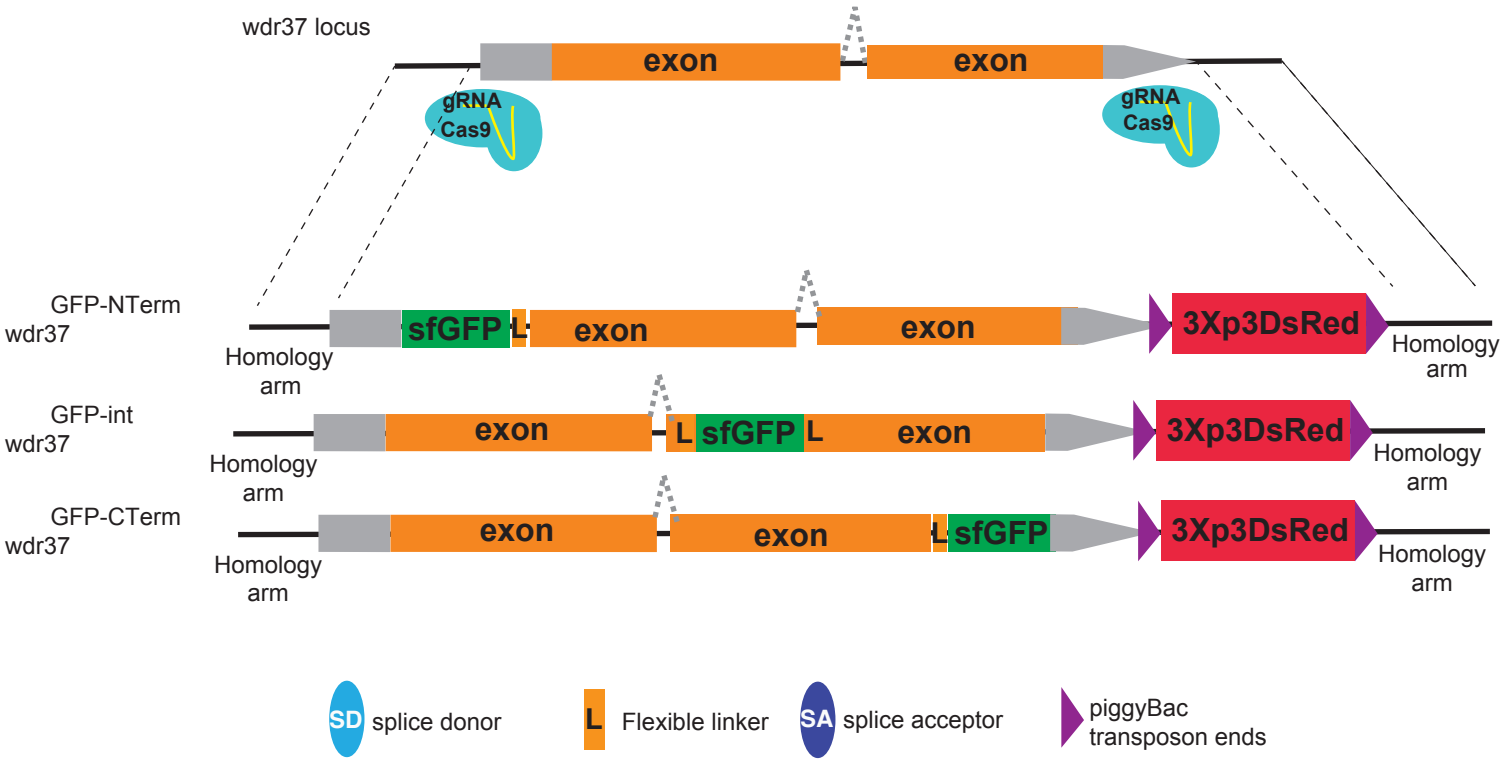


B Transgenesis rates of gRNA_int200 strategy

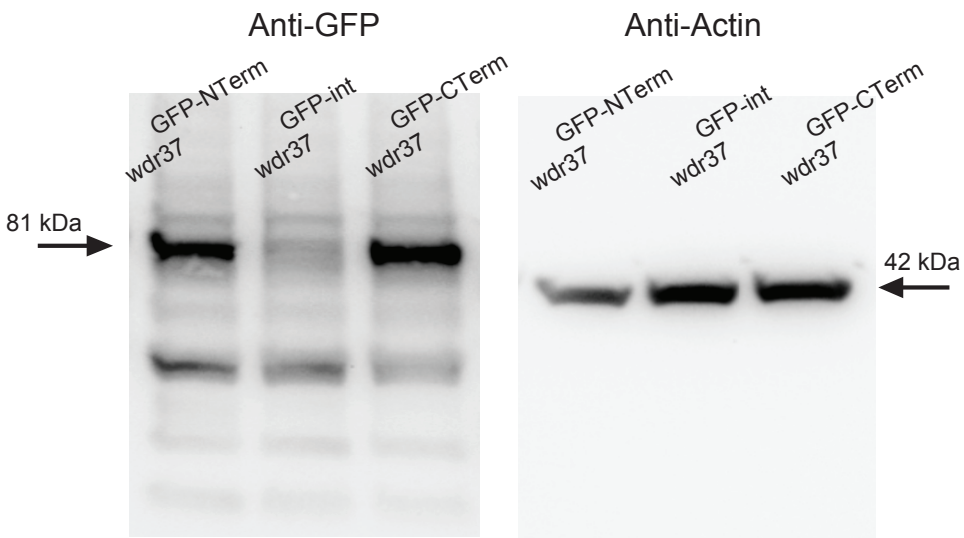
gRNA_int200 strategy	Genes targeted	Successful integration	Transgenesis Rate
T2AGAL4	127	95	74.80
KozakGAL4	72	59	81.94

Kanca et al. 2021 Figure 5. GFP protein trap allele generation

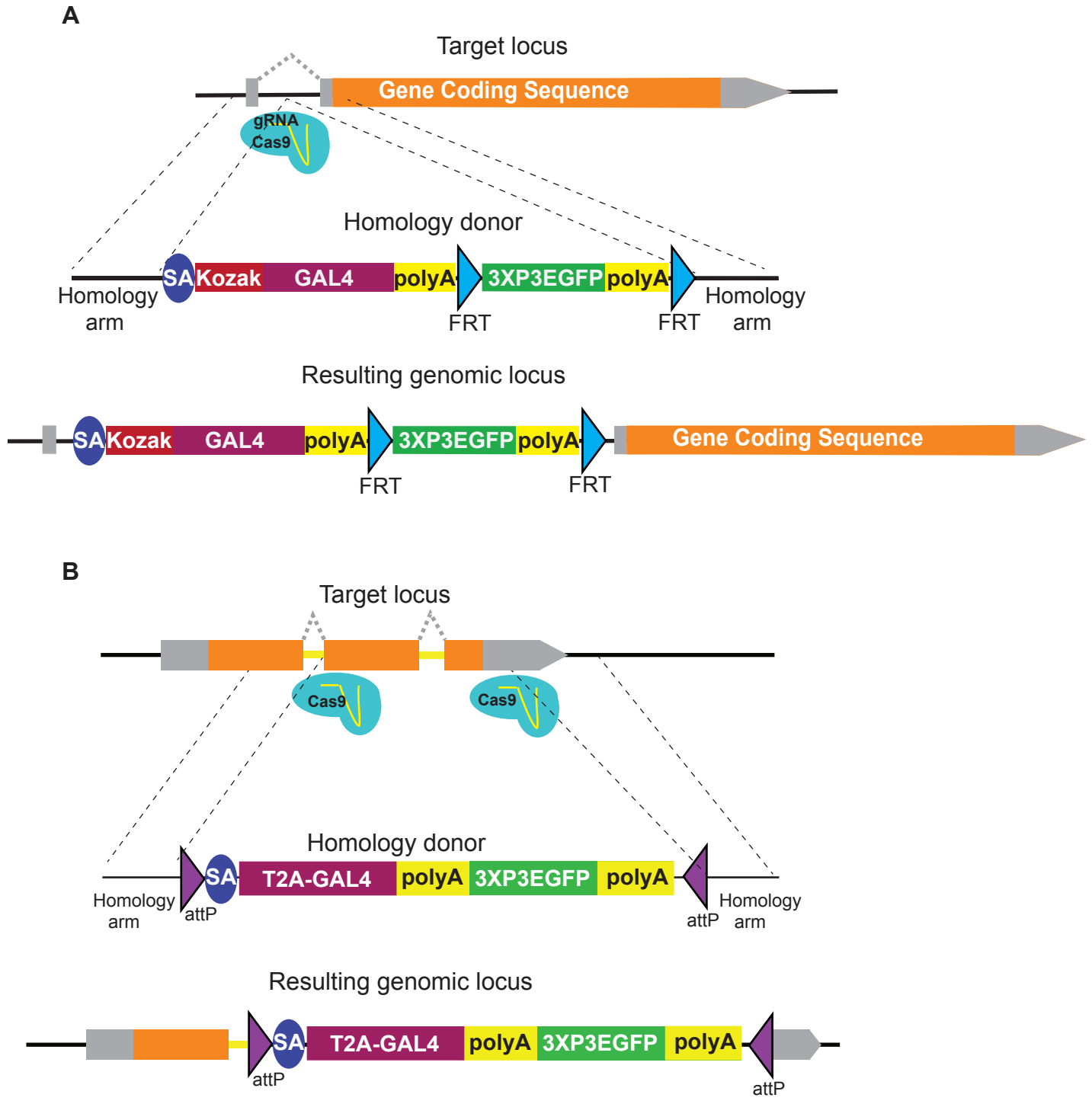
A Schematics of GFP tagging using coding region replacement



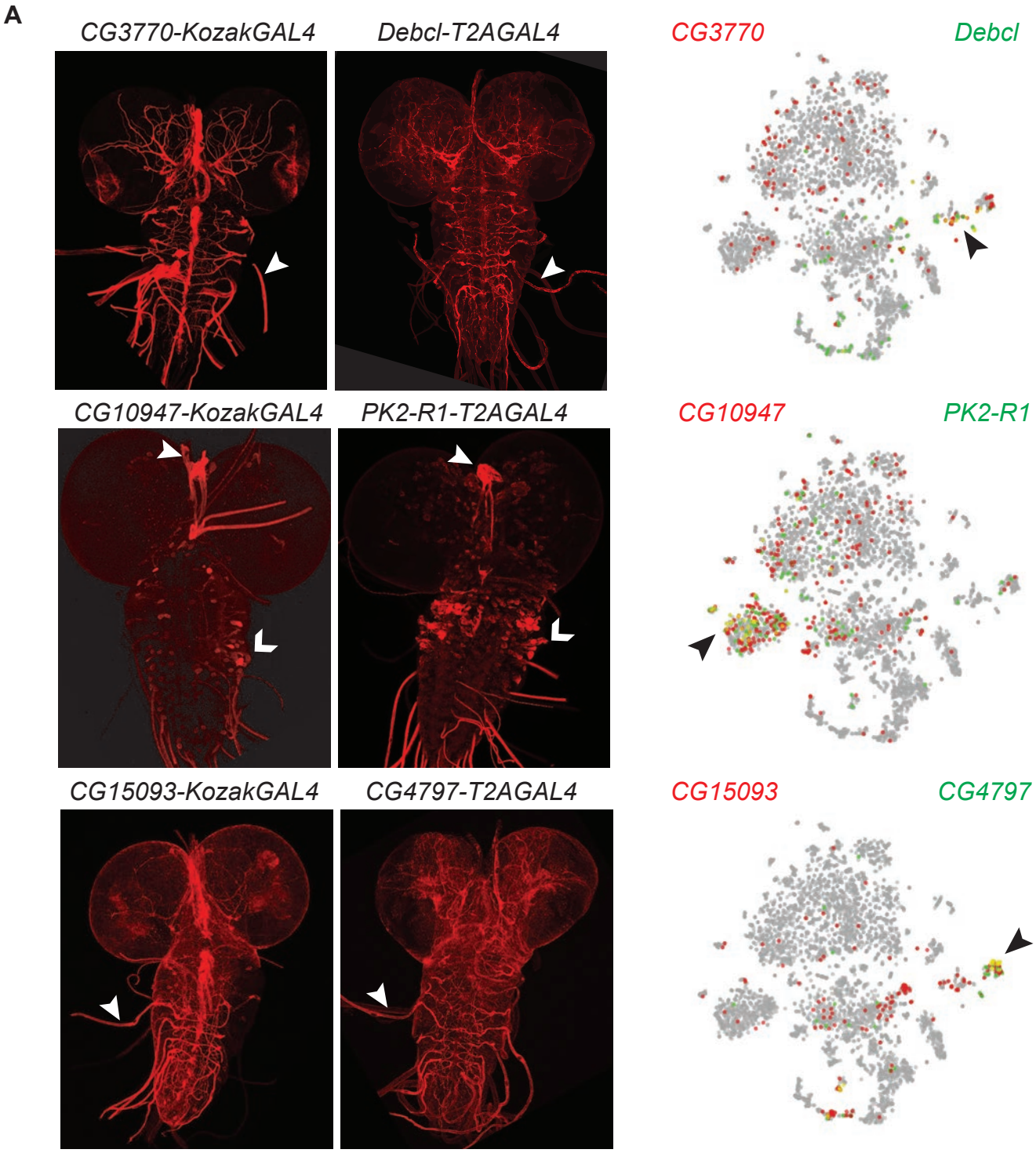
B Western Blot of samples from indicated alleles show a band of expected size for full length Wdr37-GFP



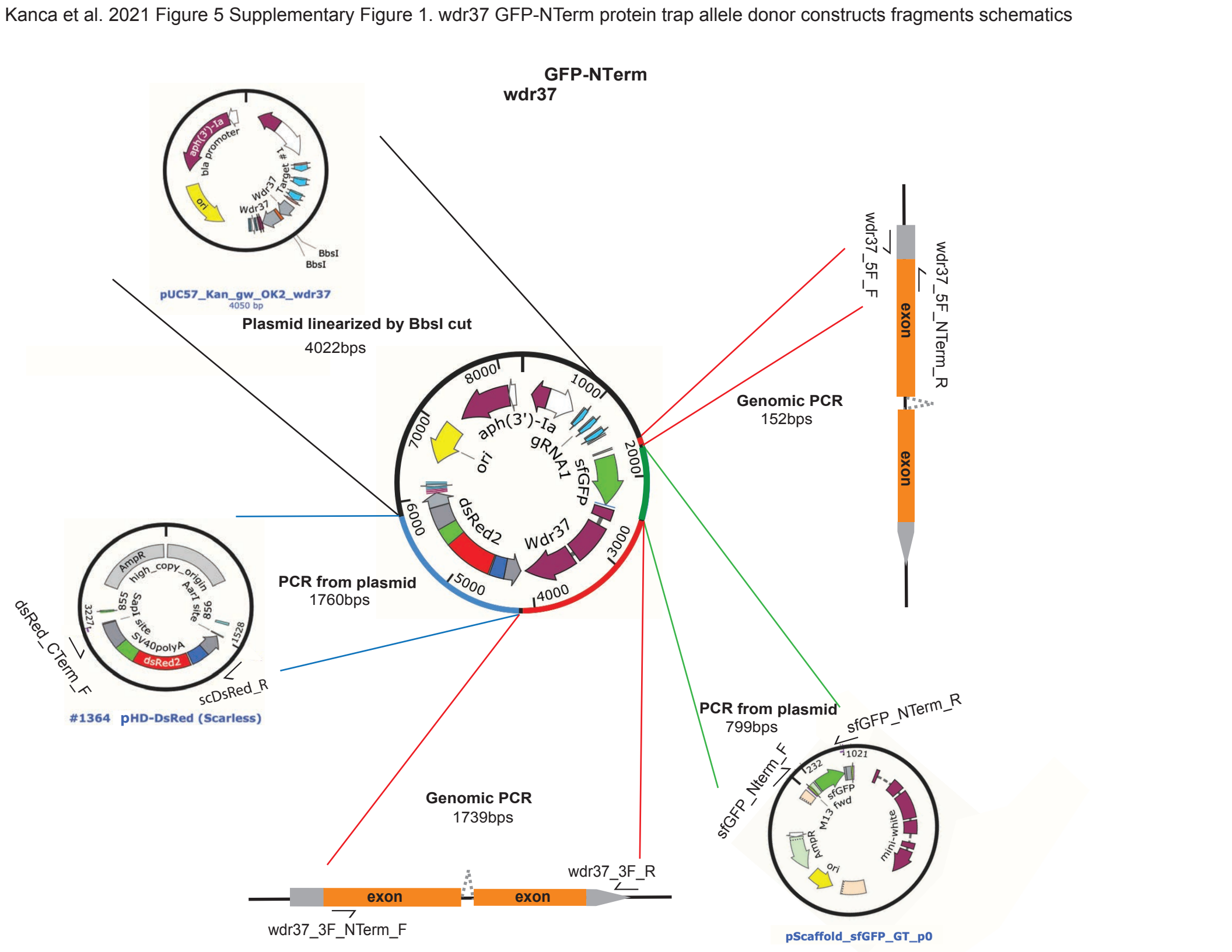
Kanca et al. 2021 Supplemental Figure 1. Alternative strategies to generate GAL4 gene traps



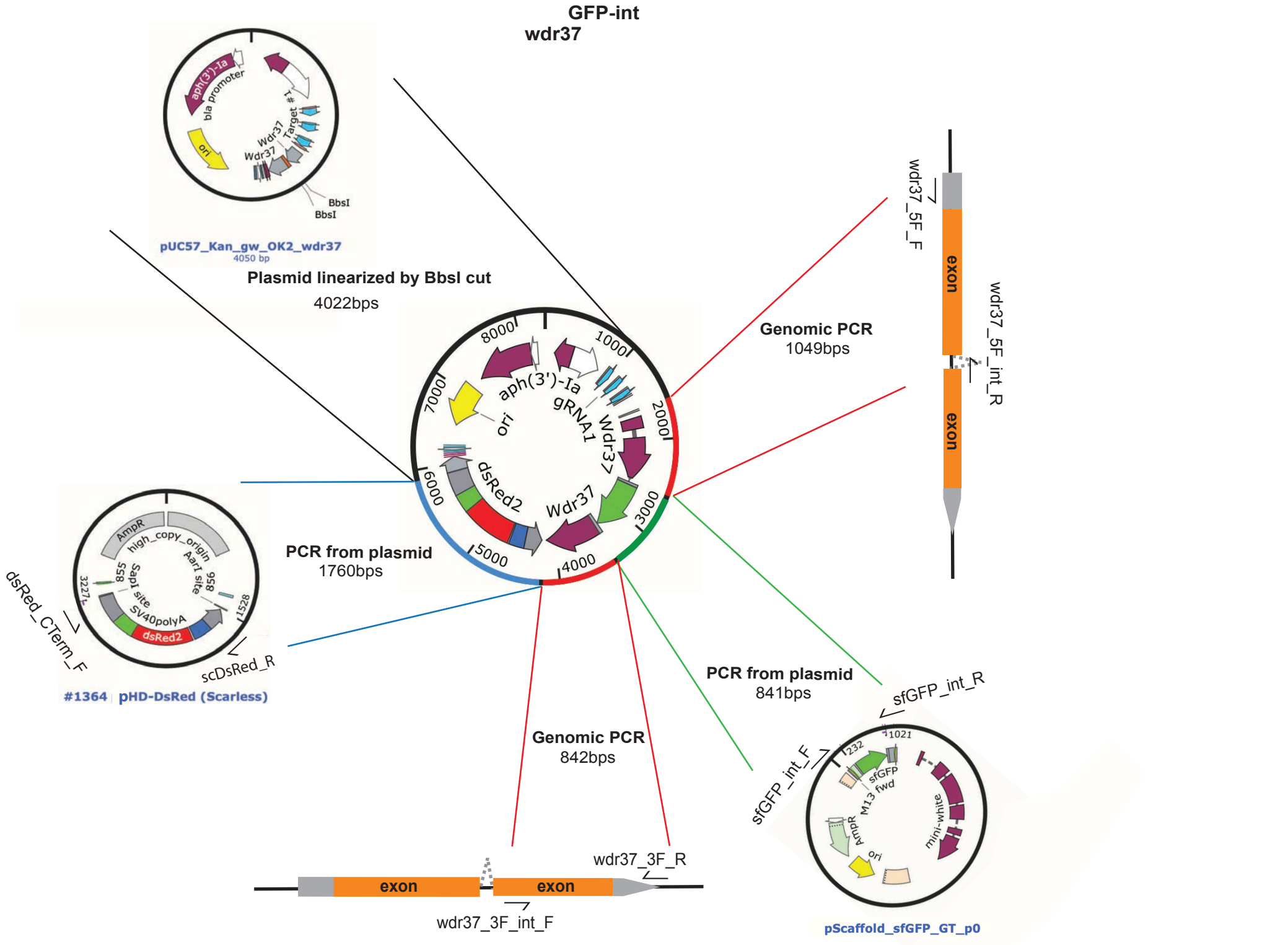
Kanca et al. 2021 Supplementary Figure 2. Single cell sequencing based overlaps in expression of indicated genes in third instar larval brain



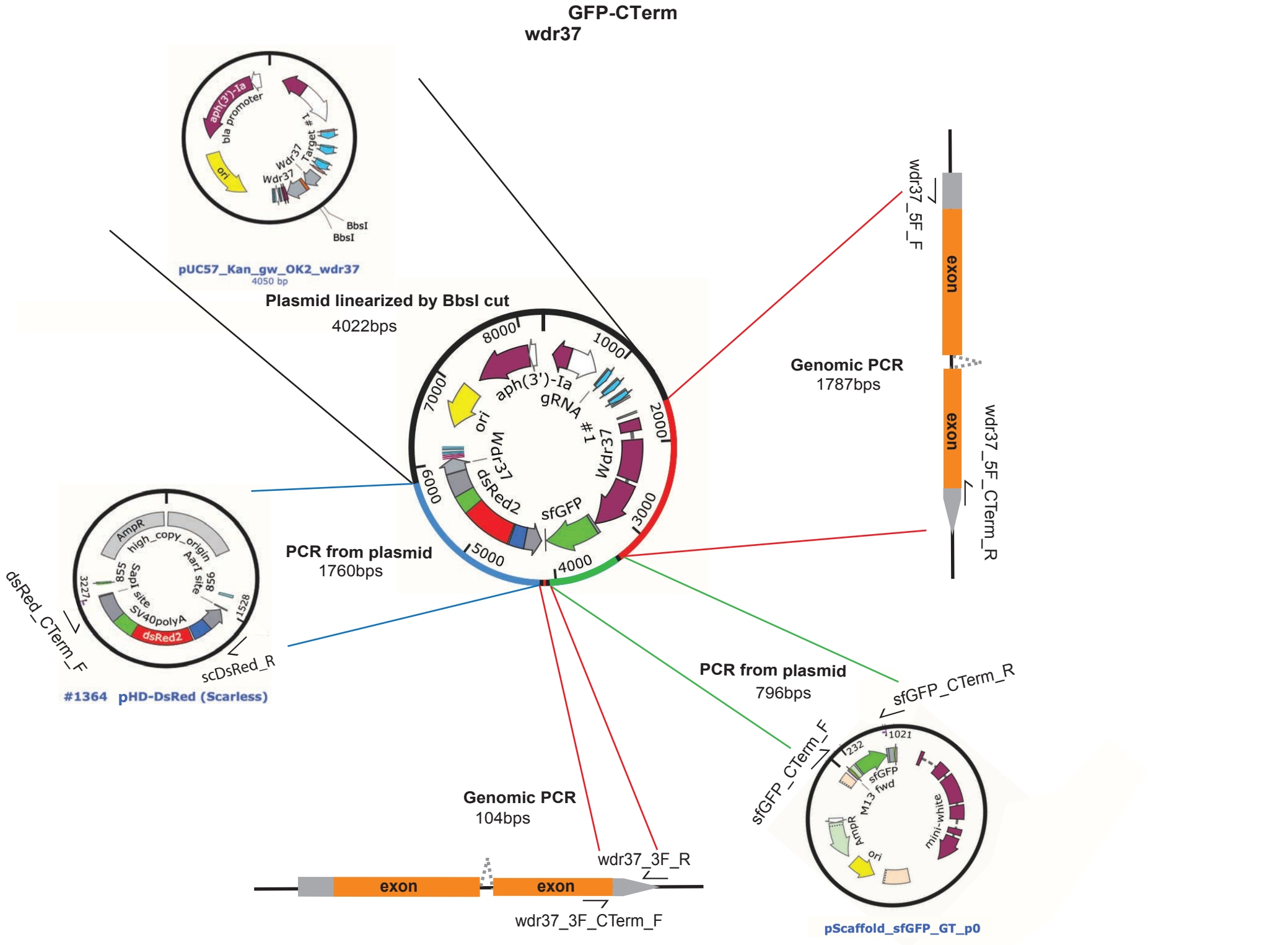
Kanca et al. 2021 Figure 5 Supplementary Figure 1. wdr37 GFP-NTerm protein trap allele donor constructs fragments schematics



Kanca et al. 2021 Figure 5 Supplementary Figure 2. wdr37 GFP-int protein trap allele donor constructs fragments schematics



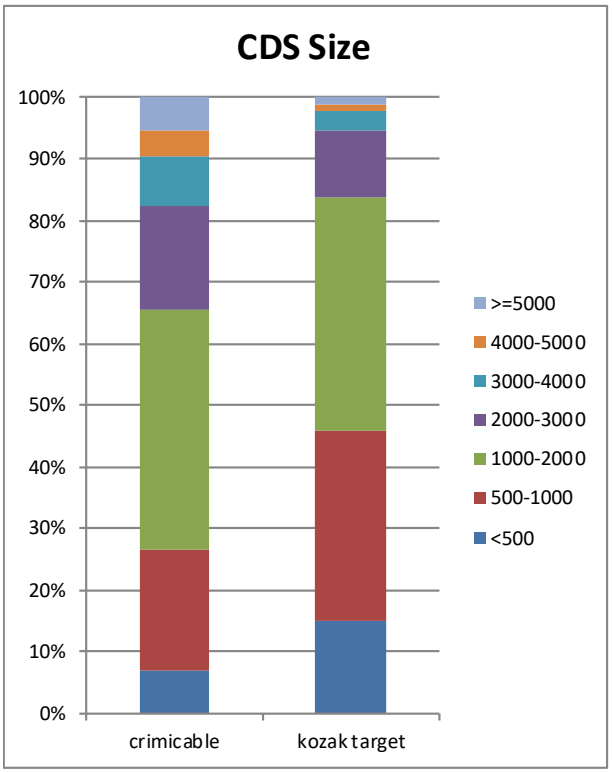
Kanca et al. 2021 Figure 5 Supplementary Figure 3. wdr37 GFP-CTerm protein trap allele donor constructs fragments schematics

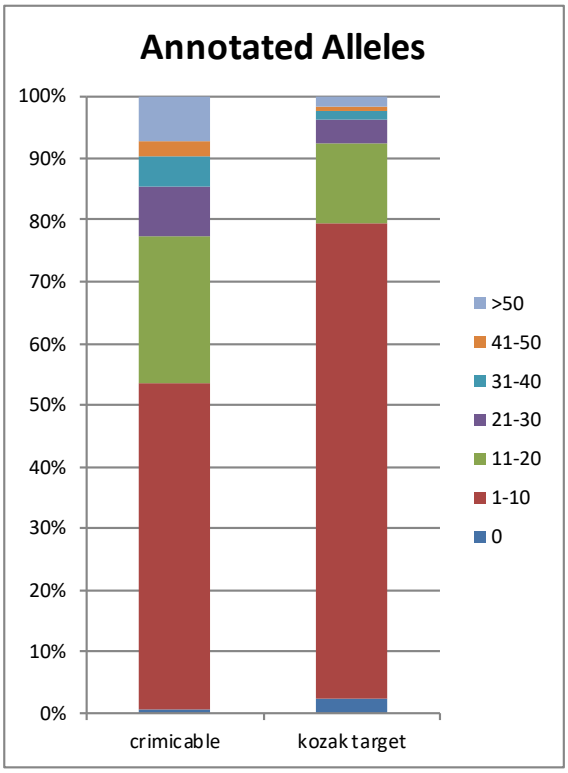


Supplementary Table 1. Analysis of Drosophila genome for the presence of suitable introns.

CDS length	crimicable	kozak target
<500	412	759
500-1000	1118	1530
1000-2000	2262	1887
2000-3000	965	554
3000-4000	479	151
4000-5000	241	54
>=5000	310	62
avg	2051	1321
all genes	5787	4997

Annotated alleles	crimicable	kozak target
0	27	110
1-10	3069	3867
11-20	1380	648
21-30	476	185
31-40	265	67
41-50	156	42
>50	414	78
avg	18.5	9
all genes	5787	4997





Supplementary Table 2. List of the 428 alleles generated in this study

int200_T2AGAL4 strategy

CR	Gene	Chr.	Donor DNA	FlyBase FBgn
CR02214	Ge-1	2	int200_pM37_p1	FBgn0283682
CR02215	trp	3	int200_pM37_p2	FBgn0003861
CR02222	REPTOR-BP	2	int200_pM37_p0	FBgn0032202
CR02226	hang	X	int200_pM37_p0	FBgn0026575
CR02229	Spt6	X	int200_pM37_p1	FBgn0028982
CR02234	CG4115	3	int200_pM37_p0	FBgn0038017
CR02237	l(2)k05911	2	int200_pM37_p1	FBgn0284244
CR02239	mmd	X	int200_pM37_p1	FBgn0259110
CR02242	CG1371	2	int200_pM37_p1	FBgn0033482
CR02244	Cirl	2	int200_pM37_p2	FBgn0033313
CR02253	yellow-h	4	int200_pM37_p0	FBgn0039896
CR02254	Fatp3	2	int200_pM37_p0	FBgn0034999
CR02257	ey	4	int200_pM37_p1	FBgn0005558
CR02258	nAChRalpha3	X	int200_pM37_p1	FBgn0015519
CR02259	nAChRbeta2	3	int200_pM37_p1	FBgn0004118
CR02260	side-II	2	int200_pM37_p1	FBgn0259213
CR02261	ATPsynbeta	4	int200_pM37_p1	FBgn0010217
CR02262	myo	4	int200_pM37_p1	FBgn0026199
CR02267	Tace	3	int200_pM37_p1	FBgn0039734
CR02270	poe	2	int200_pM37_p2	FBgn0011230
CR02279	lectin-46Ca	2	int200_pM37_p0	FBgn0040093
CR02280	He	2	int200_pM37_p0	FBgn0028430
CR02283	TbCMF46	2	int200_pM37_p0	FBgn0032163
CR02284	CG30383	2	int200_pM37_p0	FBgn0050383
CR02285	Adgf-E	2	int200_pM37_p0	FBgn0033952
CR02287	Ns2	2	int200_pM37_p0	FBgn0034243
CR02289	CG44434	2	int200_pM37_p0	FBgn0265626
CR02292	Cyp9h1	2	int200_pM37_p0	FBgn0033775
CR02293	Usp14	2	int200_pM37_p0	FBgn0032216
CR02296	Gr39b	2	int200_pM37_p0	FBgn0041245
CR02298	Cul3	2	int200_pM37_p0	FBgn0261268
CR02303	CG14339	2	int200_pM37_p0	FBgn0031301
CR02304	Ugt36F1	2	int200_pM37_p0	FBgn0027074
CR02306	CG2772	2	int200_pM37_p0	FBgn0031533
CR02308	CG33509	2	int200_pM37_p0	FBgn0053509
CR02309	CG31816	2	int200_pM37_p0	FBgn0051816
CR02310	CG3907	2	int200_pM37_p0	FBgn0034958
CR02311	Wdr59	2	int200_pM37_p0	FBgn0032339
CR02315	rhi	2	int200_pM37_p0	FBgn0004400
CR02316	CG17974	2	int200_pM37_p0	FBgn0034624
CR02319	CG34430	2	int200_pM37_p0	FBgn0085459
CR02327	CG34205	2	int200_pM37_p0	FBgn0085234
CR02329	CG17768	2	int200_pM37_p0	FBgn0032240
CR02336	CG4294	2	int200_pM37_p0	FBgn0034742
CR02341	lok	2	int200_pM37_p0	FBgn0019686

CR02342	CG6220	2	int200_pM37_p0	FBgn0033865
CR02343	ACXE	2	int200_pM37_p0	FBgn0040506
CR02344	CG18480	2	int200_pM37_p0	FBgn0028518
CR02345	cbs	2	int200_pM37_p0	FBgn0086757
CR02346	Obp58b	2	int200_pM37_p0	FBgn0034768
CR02348	Tsp42Eq	2	int200_pM37_p0	FBgn0033138
CR02349	CG15483	2	int200_pM37_p0	FBgn0032457
CR02351	CG18301	2	int200_pM37_p0	FBgn0032265
CR02352	Tsp39D	2	int200_pM37_p0	FBgn0032943
CR02354	CG5096	2	int200_pM37_p0	FBgn0032235
CR02355	rho-5	2	int200_pM37_p0	FBgn0041723
CR02357	CG8407	2	int200_pM37_p0	FBgn0033687
CR02358	CG16974	2	int200_pM37_p0	FBgn0032479
CR02359	Tsp42Ei	2	int200_pM37_p0	FBgn0033130
CR02360	Or49b	2	int200_pM37_p0	FBgn0028963
CR02361	lid	2	int200_pM37_p0	FBgn0031759
CR02362	CG30043	2	int200_pM37_p0	FBgn0050043
CR02363	CG13160	2	int200_pM37_p0	FBgn0033720
CR02365	stil	2	int200_pM37_p0	FBgn0003527
CR02367	l(2)05714	2	int200_pM37_p0	FBgn0010607
CR02368	CG30499	2	int200_pM37_p0	FBgn0050499
CR02369	CG1360	2	int200_pM37_p0	FBgn0033195
CR02370	Incenp	2	int200_pM37_p0	FBgn0260991
CR02371	CG9426	2	int200_pM37_p0	FBgn0032485
CR02374	Gr28a	2	int200_pM37_p0	FBgn0041247
CR02375	Odc2	2	int200_pM37_p0	FBgn0013308
CR02376	dila	2	int200_pM37_p0	FBgn0033447
CR02378	CG8738	2	int200_pM37_p0	FBgn0033321
CR02380	Alp2	2	int200_pM37_p0	FBgn0283480
CR02383	CG4670	2	int200_pM37_p0	FBgn0033814
CR02385	BicC	2	int200_pM37_p0	FBgn0000182
CR02388	CG4712	2	int200_pM37_p0	FBgn0033818
CR02393	Rab3	2	int200_pM37_p0	FBgn0005586
CR02394	mahj	2	int200_pM37_p0	FBgn0034641
CR02397	DnaJ-H	2	int200_pM37_p0	FBgn0032474
CR02398	Tsp42Er	2	int200_pM37_p0	FBgn0033139
CR02399	Tsp42Eo	2	int200_pM37_p1	FBgn0033136
CR02400	CG7222	2	int200_pM37_p1	FBgn0033551
CR02407	Orcokinin	2	int200_pM37_p1	FBgn0034935
CR02416	CG30087	2	int200_pM37_p1	FBgn0050087
CR02419	Syx6	2	int200_pM37_p1	FBgn0037084
CR02420	l(2)03659	2	int200_pM37_p1	FBgn0010549
CR02424	ldgf6	2	int200_pM37_p1	FBgn0013763
CR02425	CG31706	2	int200_pM37_p1	FBgn0051706
CR02427	emb	2	int200_pM37_p1	FBgn0020497
CR02428	MED15	2	int200_pM37_p1	FBgn0027592
CR02429	CG14401	2	int200_pM37_p1	FBgn0032900
CR02431	CG10764	2	int200_pM37_p1	FBgn0034221

CR02432	cnk	2	int200_pM37_p1	FBgn0286070
CR02433	RhoGAP54D	2	int200_pM37_p1	FBgn0034249
CR02437	CG10621	2	int200_pM37_p1	FBgn0032726
CR02438	CG6967	2	int200_pM37_p1	FBgn0034187
CR02442	CG42697	2	int200_pM37_p1	FBgn0261587
CR02443	CG33296	2	int200_pM37_p1	FBgn0053296
CR02444	spict	2	int200_pM37_p1	FBgn0032451
CR02445	ldgf3	2	int200_pM37_p1	FBgn0020414
CR02446	Sec71	2	int200_pM37_p1	FBgn0028538
CR02448	hoe2	2	int200_pM37_p1	FBgn0031649
CR02450	CG13793	2	int200_pM37_p1	FBgn0031935
CR02451	CG12910	2	int200_pM37_p1	FBgn0033502
CR02453	CG13437	2	int200_pM37_p1	FBgn0034541
CR02454	CG42319	2	int200_pM37_p1	FBgn0259219
CR02458	CG31784	2	int200_pM37_p1	FBgn0051784
CR02459	CG30371	2	int200_pM37_p1	FBgn0050371
CR02462	CG8353	2	int200_pM37_p1	FBgn0032002
CR02464	CG43244	2	int200_pM37_p1	FBgn0262889
CR02467	CG5421	2	int200_pM37_p1	FBgn0032434
CR02468	CG17760	2	int200_pM37_p1	FBgn0033756
CR02470	Ugt36A1	2	int200_pM37_p1	FBgn0015663
CR02475	CG14352	2	int200_pM37_p1	FBgn0031351
CR02476	CG30172	2	int200_pM37_p2	FBgn0050172
CR02477	CG17047	2	int200_pM37_p2	FBgn0033827
CR02479	Myd88	2	int200_pM37_p2	FBgn0033402
CR02480	CG16799	2	int200_pM37_p2	FBgn0034538
CR02481	pgant3	2	int200_pM37_p2	FBgn0027558
CR02482	CG5867	2	int200_pM37_p2	FBgn0027586
CR02484	sha	2	int200_pM37_p2	FBgn0003382
CR02486	CG3294	2	int200_pM37_p2	FBgn0031628
CR02488	stc	2	int200_pM37_p2	FBgn0001978
CR02490	Zir	2	int200_pM37_p2	FBgn0031216
CR02491	CG9270	2	int200_pM37_p2	FBgn0032908
CR02492	DUBAI	2	int200_pM37_p2	FBgn0033738
CR02493	GalNAc-T1	2	int200_pM37_p2	FBgn0034025
CR02494	Or33b	2	int200_pM37_p2	FBgn0026391
CR02496	Or42a	2	int200_pM37_p2	FBgn0033041
CR02497	Tep1	2	int200_pM37_p2	FBgn0041183
CR02499	CG15153	2	int200_pM37_p2	FBgn0032663
CR02500	CG31741	2	int200_pM37_p2	FBgn0051741
CR02501	Gr33a	2	int200_pM37_p2	FBgn0032416
CR02502	CG18063	2	int200_pM37_p2	FBgn0028856
CR02504	CG15118	2	int200_pM37_p0	FBgn0034418
CR02505	CG9515	2	int200_pM37_p0	FBgn0032077
CR02506	List	2	int200_pM37_p0	FBgn0034381
CR02509	crq	2	int200_pM37_p0	FBgn0015924
CR02510	CG7115	2	int200_pM37_p0	FBgn0027515
CR02511	scat	2	int200_pM37_p0	FBgn0011232

CR02514	NimB1	2	int200_pM37_p0	FBgn0027929
CR02517	SmydA-3	2	int200_pM37_p0	FBgn0262599
CR02520	CG30183	2	int200_pM37_p0	FBgn0050183
CR02521	Vajk2	2	int200_pM37_p0	FBgn0032538
CR02522	CG44836	3	int200_pM37_p0	FBgn0266099
CR02523	Madm	3	int200_pM37_p0	FBgn0027497
CR02529	Dop1R2	3	int200_pM37_p0	FBgn0266137
CR02532	CG12880	3	int200_pM37_p0	FBgn0046258
CR02533	Ppcs	3	int200_pM37_p0	FBgn0261285
CR02534	fau	3	int200_pM37_p0	FBgn0266451
CR02535	Glg1	3	int200_pM37_p0	FBgn0264561
CR02536	CG5768	3	int200_pM37_p0	FBgn0039198
CR02537	CG34377	3	int200_pM37_p0	FBgn0263117
CR02542	dmrt93B	3	int200_pM37_p0	FBgn0038851
CR02543	Arl8	3	int200_pM37_p0	FBgn0037551
CR02544	Ttc26	3	int200_pM37_p0	FBgn0038358
CR02546	CG42327	3	int200_pM37_p0	FBgn0259227
CR02550	CG11791	3	int200_pM37_p0	FBgn0039266
CR02554	CG5196	3	int200_pM37_p0	FBgn0038039
CR02555	CG2182	3	int200_pM37_p0	FBgn0037360
CR02556	Karybeta3	3	int200_pM37_p0	FBgn0087013
CR02559	CG6695	3	int200_pM37_p0	FBgn0039215
CR02560	Ada2b	3	int200_pM37_p0	FBgn0037555
CR02562	Nelf-A	3	int200_pM37_p0	FBgn0038872
CR02565	Spn85F	3	int200_pM37_p0	FBgn0037772
CR02566	Root	3	int200_pM37_p0	FBgn0039152
CR02569	CG15715	3	int200_pM37_p0	FBgn0036538
CR02573	unc-13	4	int200_pM37_p0	FBgn0025726
CR02574	Arl4	4	int200_pM37_p0	FBgn0039889
CR02575	su(f)	X	int200_pM37_p0	FBgn0003559
CR02576	ogre	X	int200_pM37_p0	FBgn0004646
CR02579	CG2556	X	int200_pM37_p0	FBgn0030396
CR02580	CG15765	X	int200_pM37_p0	FBgn0029814
CR02581	CG32547	X	int200_pM37_p0	FBgn0052547
CR02582	br	X	int200_pM37_p0	FBgn0283451
CR02583	CG15211	X	int200_pM37_p0	FBgn0030234
CR02584	CG4928	X	int200_pM37_p0	FBgn0027556
CR02585	IP3K2	X	int200_pM37_p0	FBgn0283680
CR02586	rho-4	X	int200_pM37_p0	FBgn0030318
CR02587	CG9411	X	int200_pM37_p0	FBgn0030569
CR02588	CG1504	X	int200_pM37_p0	FBgn0031100
CR02589	CG43901	X	int200_pM37_p0	FBgn0264502
CR02591	CG32815	X	int200_pM37_p0	FBgn0052815
CR02597	t	X	int200_pM37_p0	FBgn0086367
CR02598	CG42323	X	int200_pM37_p0	FBgn0259223
CR02599	CG14626	X	int200_pM37_p1	FBgn0040360
CR02602	Evi5	X	int200_pM37_p1	FBgn0262740
CR02604	CG14435	X	int200_pM37_p1	FBgn0029911

CR02605	Nup153	X	int200_pM37_p1	FBgn0061200
CR02607	OtopLa	X	int200_pM37_p1	FBgn0259994
CR02608	CG14625	X	int200_pM37_p1	FBgn0040358
CR02609	CG32536	X	int200_pM37_p1	FBgn0052536
CR02612	Ucp4A	X	int200_pM37_p1	FBgn0030872
CR02615	inc	X	int200_pM37_p1	FBgn0025394
CR02616	Kap3	X	int200_pM37_p1	FBgn0028421
CR02617	Ir8a	X	int200_pM37_p1	FBgn0052704
CR02621	par-6	X	int200_pM37_p1	FBgn0026192
CR02624	CG1636	X	int200_pM37_p1	FBgn0030030
CR02631	zf30C	2	int200_pM37_p1	FBgn0270924
CR02633	Ten-m	3	int200_pM37_p1	FBgn0004449
CR02634	nAChRbeta1	3	int200_pM37_p1	FBgn0000038
CR02670	CaMKI	4	int200_pM37_p1	FBgn0016126
CR02671	rdgA	X	int200_pM37_p1	FBgn0261549
CR02672	CG44422	X	int200_pM37_p1	FBgn0265595
CR02673	CG42340	X	int200_pM37_p1	FBgn0259242
CR02676	CG43740	X	int200_pM37_p1	FBgn0263997
CR02677	CG33181	X	int200_pM37_p1	FBgn0053181
CR02679	CG32647	X	int200_pM37_p1	FBgn0052647
CR02680	Hers	X	int200_pM37_p1	FBgn0052529
CR02681	CG1695	X	int200_pM37_p2	FBgn0031116
CR02684	CG42541	X	int200_pM37_p2	FBgn0260658
CR02687	CG2256	X	int200_pM37_p2	FBgn0029995
CR02693	inaF-D	X	int200_pM37_p1	FBgn0260812
CR02694	CG17162	X	int200_pM37_p2	FBgn0039944
CR02697	b6	X	int200_pM37_p0	FBgn0024897
CR02699	Hk	X	int200_pM37_p0	FBgn0263220
CR02701	CG8568	X	int200_pM37_p1	FBgn0030841
CR02702	CG12065	X	int200_pM37_p1	FBgn0030052
CR02705	CG1545	X	int200_pM37_p1	FBgn0030259
CR02707	dmrt11E	X	int200_pM37_p1	FBgn0030477
CR02708	CG1532	X	int200_pM37_p1	FBgn0031143
CR02709	CG7332	X	int200_pM37_p1	FBgn0030973
CR02711	CG13012	X	int200_pM37_p2	FBgn0030769
CR02713	CG4404	X	int200_pM37_p0	FBgn0030432
CR02721	CG8908	2	int200_pM37_p0	FBgn0034493
CR02722	snama	2	int200_pM37_p0	FBgn0086129
CR02723	TwdlE	2	int200_pM37_p0	FBgn0031957
CR02745	hiw	X	int200_pM37_p1	FBgn0030600
CR02779	Slip1	4	int200_pM37_p1	FBgn0024728
CR02793	CG11360	4	int200_pM37_p1	FBgn0039920
CR02798	mAChR-A	2	int200_pM37_p1	FBgn0000037
CR70009	lip3	3	int200_pM37_p0	FBgn0023495
CR70012	tau	3	int200_pM37_p0	FBgn0266579
CR70028	CtsB1	X	int200_pM37_p0	FBgn0030521
CR70032	ppt1	X	int200_pM37_p2	FBgn0030057
CR70034	spt6	X	int200_pM37_p1	FBgn0028982

CR70035	Tob	X	int200_pM37_p0	FBgn0028397
CR70041	CG16935	2	int200_pM37_p1	FBgn0033883
CR70044	iPLA2-VIA	3	int200_pM37_p1	FBgn0036053
CR70029	dsb	3	int200_pM37_p2	FBgn0035290
CR70033	prt	3	int200_pM37_p1	FBgn0043005
CR70037	aux	3	int200_pM37_p1	FBgn0037218
CR70036	anne	4	int200_pM37_p1	FBgn0052000
CR70070	lgs	4	int200_pM37_p1	FBgn0039907
CR70074	mim	2	int200_pM37_p0	FBgn0053558
CR70088	csp	3	int200_pM37_p2	FBgn0004179
CR70089	Cdase	3	int200_pM37_p0	FBgn0039774
CR70071	CG3226	X	int200_pM37_p2	FBgn0029882
CR70077	GLS	2	int200_pM37_p2	FBgn0261625
CR70087	CG7236	2	int200_pM37_p2	FBgn0031730
CR70048	Trp-Gamma	2	int200_TGEM_p2	FBgn0032593

int200_KozakGAL4 strategy

CR	Gene	Chr.	Donor DNA	FlyBase FBgn
CR70003	CG14044	2	int200_pM37_kG4	FBgn0031650
CR70002	CG15728	X	int200_pM37_kG4	FBgn0030409
CR70005	CG16787	2	int200_pM37_kG4	FBgn0034940
CR70008	CG8027	2	int200_pM37_kG4	FBgn0033392
CR70013	Ast-C	2	int200_pM37_kG4	FBgn0032336
CR70011	CG16979	3	int200_pM37_kG4	FBgn0036512
CR70014	CG13737	3	int200_pM37_kG4	FBgn0036382
CR70038	CG3376	2	int200_pM37_kG4	FBgn0034997
CR70045	oli	2	int200_pM37_kG4	FBgn0032651
CR70046	SK1	X	int200_pM37_kG4	FBgn0030300
CR70043	intS11	3	int200_pM37_kG4	FBgn0039691
CR70020	B-Glu	3	int200_pM37_kG4	FBgn0270927
CR70040	CG8320	2	int200_pM37_kG4	FBgn0034059
CR70022	CG14457	3	int200_pM37_kG4	FBgn0037174
CR70053	pngl	2	int200_pM37_kG4	FBgn0033050
CR70068	Bap111	X	int200_pM37_kG4	FBgn0030093
CR70069	Bap60	X	int200_pM37_kG4	FBgn0025463
CR70050	Vkor	2	int200_pM37_kG4	FBgn0053544
CR70079	Hip1	3	int200_pM37_kG4	FBgn0036309
CR70080	PIGA	2	int200_pM37_kG4	FBgn0288203
CR70082	elo68Beta	3	int200_pM37_kG4	FBgn0036128
CR70083	CG15744	X	int200_pM37_kG4	FBgn0030466

gRNA_int200_T2AGAL4 strategy

CR	Gene	Chr.	Donor DNA	FlyBase FBgn
CR70092	ara	3	gRNA_int200_pM37_p0	FBgn0015904
CR70094	lace	2	gRNA_int200_pM37_p0	FBgn0002524
CR70095	Pex23	3	gRNA_int200_pM37_p2	FBgn0052226
CR70096	DIP-λ	2	gRNA_int200_pM37_p0	FBgn0267428
CR70097	fwe	3	gRNA_int200_pM37_p1	FBgn0261722
CR70098	salr	2	gRNA_int200_pM37_p0	FBgn0000287
CR70099	CG15528	3	gRNA_int200_pM37_p2	FBgn0039742
CR70107	dmrt99B	3	gRNA_int200_pM37_p0	FBgn0039683
CR70108	Lsd2	X	gRNA_int200_pM37_p1	FBgn0030608
CR70110	SREBP	3	gRNA_int200_pM37_p0	FBgn0261283
CR70114	RhoGAP68F	3	gRNA_int200_pM37_p1	FBgn0036257
CR70115	ifc	2	gRNA_int200_pM37_p2	FBgn0001941
CR70118	KrT95D	3	gRNA_int200_pM37_p0	FBgn0020647
CR70121	Mfe2	X	gRNA_int200_pM37_p1	FBgn0030731
CR70122	Vps13	2	gRNA_int200_pM37_p1	FBgn0033194
CR70125	CG6967	2	gRNA_int200_pM37_p1	FBgn0034187
CR70127	CG33090	2	gRNA_int200_pM37_p2	FBgn0028916
CR70133	ec	X	gRNA_int200_pM37_p0	FBgn0000542
CR70135	srl	3	gRNA_int200_pM37_p2	FBgn0037248
CR70136	Root	3	gRNA_int200_pM37_p0	FBgn0039152
CR70137	Su(z)12	3	gRNA_int200_pM37_p0	FBgn0020887
CR70140	cnn	2	gRNA_int200_pM37_p1	FBgn0013765
CR70144	AdenoK	3	gRNA_int200_pM37_p0	FBgn0036337
CR70145	CG8199	3	gRNA_int200_pM37_p2	FBgn0037709
CR70149	CG5044	3	gRNA_int200_pM37_p0	FBgn0038326
CR70150	sub	2	gRNA_int200_pM37_p0	FBgn0003545
CR70151	Tusp	3	gRNA_int200_pM37_p0	FBgn0039530
CR70152	FarO	X	gRNA_int200_pM37_p2	FBgn0023550
CR70157	CG14736	3	gRNA_int200_pM37_p1	FBgn0037986
CR70162	CG4612	2	gRNA_int200_pM37_p0	FBgn0035016
CR70165	DNApolalpha	3	gRNA_int200_pM37_p1	FBgn0259113
CR70166	mav	4	gRNA_int200_pM37_p0	FBgn0039914
CR70167	rhea	3	gRNA_int200_pM37_p0	FBgn0260442
CR70168	PK2-R1	3	gRNA_int200_pM37_p2	FBgn0038140
CR70173	LKRSDH	2	gRNA_int200_pM37_p0	FBgn0286198
CR70174	CG7414	3	gRNA_int200_pM37_p1	FBgn0037135
CR70177	CG12268	3	gRNA_int200_pM37_p0	FBgn0039131
CR70178	UQCR-c1	3	gRNA_int200_pM37_p0	FBgn0038271
CR70179	wal	2	gRNA_int200_pM37_p0	FBgn0010516
CR70187	CG18155	X	gRNA_int200_pM37_p1	FBgn0029945
CR70188	CG33253	X	gRNA_int200_pM37_p1	FBgn0030992
CR70189	Grasp65	3	gRNA_int200_pM37_p1	FBgn0036919
CR70190	IntS11	3	gRNA_int200_pM37_p1	FBgn0039691
CR70191	Mrtf	3	gRNA_int200_pM37_p1	FBgn0052296
CR70192	pAbp	2	gRNA_int200_pM37_p1	FBgn0265297
CR70196	rin	3	gRNA_int200_pM37_p1	FBgn0015778

CR70197	wat	3	gRNA_int200_pM37_p1	FBgn0039620
CR70199	CG5599	X	gRNA_int200_pM37_p0	FBgn0030612
CR70200	CG17168	X	gRNA_int200_pM37_p2	FBgn0039943
CR70201	Edem2	2	gRNA_int200_pM37_p2	FBgn0032480
CR70202	GM130	2	gRNA_int200_pM37_p0	FBgn0034697
CR70205	nocte	X	gRNA_int200_pM37_p0	FBgn0261710
CR70207	pex1	3	gRNA_int200_pM37_p0	FBgn0013563
CR70208	Prx6005	2	gRNA_int200_pM37_p2	FBgn0031479
CR70209	Rlip	3	gRNA_int200_pM37_p2	FBgn0026056
CR70210	roe	2	gRNA_int200_pM37_p0	FBgn0267337
CR70212	vha68-2	2	gRNA_int200_pM37_p0	FBgn0263598
CR70213	sp1	X	gRNA_int200_pM37_p2	FBgn0020378
CR70216	br	X	gRNA_int200_pM37_p0	FBgn0283451
CR70218	Axn	3	gRNA_int200_pM37_p2	FBgn0026597
CR70219	Lpr1	3	gRNA_int200_pM37_p0	FBgn0066101
CR70220	Lpr2	3	gRNA_int200_pM37_p0	FBgn0051092
CR70225	Osi2	3	gRNA_int200_pM37_p2	FBgn0037410
CR70227	wap	X	gRNA_int200_pM37_p2	FBgn0266848
CR70229	LARP4B	3	gRNA_int200_pM37_p0	FBgn0283788
CR70231	DIP-β	X	gRNA_int200_pM37_p1	FBgn0259245
CR70233	cac	X	gRNA_int200_pM37_p1	FBgn0263111
CR70241	CG7956	3	gRNA_int200_pM37_p1	FBgn0038890
CR70242	lola	2	gRNA_int200_pM37_p1	FBgn0283521
CR70243	sli	2	gRNA_int200_pM37_p1	FBgn0264089
CR70244	Dl	3	gRNA_int200_pM37_p0	FBgn0000463
CR70246	Csp	3	gRNA_int200_pM37_p0	FBgn0004179
CR70249	MFS10	X	gRNA_int200_pM37_p1	FBgn0030452
CR70251	eIF4B	3	gRNA_int200_pM37_p1	FBgn0020660
CR70256	Pis	X	gRNA_int200_pM37_p1	FBgn0030670
CR70258	Rpt3	X	gRNA_int200_pM37_p0	FBgn0028686
CR70260	CG10907	3	gRNA_int200_pM37_p2	FBgn0036207
CR70261	tdc2	2	gRNA_int200_pM37_p2	FBgn0050446
CR70263	Gdap1	3	gRNA_int200_pM37_p0	FBgn0035587
CR70265	golgin245	2	gRNA_int200_pM37_p0	FBgn0034854
CR70266	ThrRS	2	gRNA_int200_pM37_p1	FBgn0027081
CR70268	CG10137	2	gRNA_int200_pM37_p2	FBgn0032800
CR70269	CG17739	2	gRNA_int200_pM37_p1	FBgn0033710
CR70281	CrebA	3	gRNA_int200_pM37_p0	FBgn0004396
CR70282	gl	3	gRNA_int200_pM37_p0	FBgn0004618
CR70284	ND-49	4	gRNA_int200_pM37_p1	FBgn0039909
CR70287	kis	2	gRNA_int200_pM37_p0	FBgn0266557
CR70289	erm	2	gRNA_int200_pM37_p0	FBgn0031375
CR70296	Yif1	3	gRNA_int200_pM37_p1	FBgn0039450
CR70298	RhoGAP100	3	gRNA_int200_pM37_p1	FBgn0039883
CR70299	gsc	2	gRNA_int200_pM37_p0	FBgn0010323
CR70312	CG7744	2	gRNA_int200_pM37_p2	FBgn0034447
CR70314	Itgbn	2	gRNA_int200_pM37_p1	FBgn0010395
CR70324	Tim10	2	gRNA_int200_pM37_p1	FBgn0027360

CR70329	stai	2	gRNA_int200_pM37_p1	FBgn0266521

2XgRNA_int200_KozakGAL4 strategy

CR	Gene	Chr.	Donor DNA	FlyBase FBgn
CR70091	CG7423	X	2XgRNA_int200_pM37_kG4	FBgn0030982
CR70093	CG7504	3	2XgRNA_int200_pM37_kG4	FBgn0035842
CR70104	CG14795	X	2XgRNA_int200_pM37_kG4	FBgn0025393
CR70106	CG33310	2	2XgRNA_int200_pM37_kG4	FBgn0053310
CR70109	PHGPx	3	2XgRNA_int200_pM37_kG4	FBgn0035438
CR70111	Wdr37	3	2XgRNA_int200_pM37_kG4	FBgn0038617
CR70112	ZnT49B	2	2XgRNA_int200_pM37_kG4	FBgn0033762
CR70116	CG14644	3	2XgRNA_int200_pM37_kG4	FBgn0250821
CR70119	CG10947	2	2XgRNA_int200_pM37_kG4	FBgn0032857
CR70120	Usp7	X	2XgRNA_int200_pM37_kG4	FBgn0030366
CR70124	Vps13B	3	2XgRNA_int200_pM37_kG4	FBgn0039727
CR70130	mAChR-C	X	2XgRNA_int200_pM37_kG4	FBgn0029909
CR70131	CG34293	3	2XgRNA_int200_pM37_kG4	FBgn0085322
CR70139	wfs1	3	2XgRNA_int200_pM37_kG4	FBgn0039003
CR70142	Ctl2	3	2XgRNA_int200_pM37_kG4	FBgn0039637
CR70143	CG8778	2	2XgRNA_int200_pM37_kG4	FBgn0033761
CR70146	GlcT	2	2XgRNA_int200_pM37_kG4	FBgn0067102
CR70147	CG11103	X	2XgRNA_int200_pM37_kG4	FBgn0030522
CR70148	CG30022	2	2XgRNA_int200_pM37_kG4	FBgn0050022
CR70153	Ufm1	2	2XgRNA_int200_pM37_kG4	FBgn0085220
CR70154	CG8646	2	2XgRNA_int200_pM37_kG4	FBgn0033763
CR70155	cln3	3	2XgRNA_int200_pM37_kG4	FBgn0036756
CR70156	CG10104	2	2XgRNA_int200_pM37_kG4	FBgn0033933
CR70160	CG3770	2	2XgRNA_int200_pM37_kG4	FBgn0035085
CR70169	rhea	3	2XgRNA_int200_pM37_kG4	FBgn0260442
CR70176	Loxl2	2	2XgRNA_int200_pM37_kG4	FBgn0034660
CR70181	CG2124	X	2XgRNA_int200_pM37_kG4	FBgn0030217
CR70182	CG4074	3	2XgRNA_int200_pM37_kG4	FBgn0037017
CR70183	CG6638	3	2XgRNA_int200_pM37_kG4	FBgn0035911
CR70184	CG8202	3	2XgRNA_int200_pM37_kG4	FBgn0037622
CR70185	CG9184	3	2XgRNA_int200_pM37_kG4	FBgn0035208
CR70186	CG15093	2	2XgRNA_int200_pM37_kG4	FBgn0034390
CR70193	pex2	3	2XgRNA_int200_pM37_kG4	FBgn0035876
CR70194	pex16	3	2XgRNA_int200_pM37_kG4	FBgn0037019
CR70215	ash1	3	2XgRNA_int200_pM37_kG4	FBgn0005386
CR70221	Snx17	2	2XgRNA_int200_pM37_kG4	FBgn0032191
CR70222	Vps29	2	2XgRNA_int200_pM37_kG4	FBgn0031310
CR70223	AP2-sigma	3	2XgRNA_int200_pM37_kG4	FBgn0043012
CR70226	CG7943	3	2XgRNA_int200_pM37_kG4	FBgn0039741
CR70237	Fitm	3	2XgRNA_int200_pM37_kG4	FBgn0035586
CR70238	armi	3	2XgRNA_int200_pM37_kG4	FBgn0041164
CR70245	CG5254	X	2XgRNA_int200_pM37_kG4	FBgn0040383
CR70247	Pka-C1	2	2XgRNA_int200_pM37_kG4	FBgn0000273
CR70248	Ufc1	2	2XgRNA_int200_pM37_kG4	FBgn0034061
CR70252	mtm	2	2XgRNA_int200_pM37_kG4	FBgn0025742
CR70262	CG7716	3	2XgRNA_int200_pM37_kG4	FBgn0035800

CR70264	mib1	3	2XgRNA_int200_pM37_kG4	FBgn0263601
CR70267	shop	X	2XgRNA_int200_pM37_kG4	FBgn0030966
CR70270	GLS	2	2XgRNA_int200_pM37_kG4	FBgn0261625
CR70271	CG13051	3	2XgRNA_int200_pM37_kG4	FBgn0040799
CR70272	CG32191	3	2XgRNA_int200_pM37_kG4	FBgn0052191
CR70273	Arc2	2	2XgRNA_int200_pM37_kG4	FBgn0033928
CR70274	SoxN	2	2XgRNA_int200_pM37_kG4	FBgn0029123
CR70276	CG10324	3	2XgRNA_int200_pM37_kG4	FBgn0038454
CR70278	CG34247	3	2XgRNA_int200_pM37_kG4	FBgn0085276
CR70279	whip	2	2XgRNA_int200_pM37_kG4	FBgn0250827
CR70280	CG7402	3	2XgRNA_int200_pM37_kG4	FBgn0036768
CR70283	Qsox2	3	2XgRNA_int200_pM37_kG4	FBgn0038919
CR70290	CG10073	2	2XgRNA_int200_pM37_kG4	FBgn0034440

Supplementary material Primer sequences

Primers for PCR verification of CRIMIC and KozakGAL4 alleles

KozakGAL4_L_R	ACTCCCAGTTGTTCTTCAGGCACTTGGC
KozakGAL4_R_F	GAAAGTATAGGAACTTCTCGCGCTCG

CRIMIC_ch_rev	GCGGAAGAGAGATAAATCGGTTG
CRIMIC_ch_for	GTGGTATGGCTGATTATGATCAGAAG

Primers to clone WDR37 Knock-in alleles

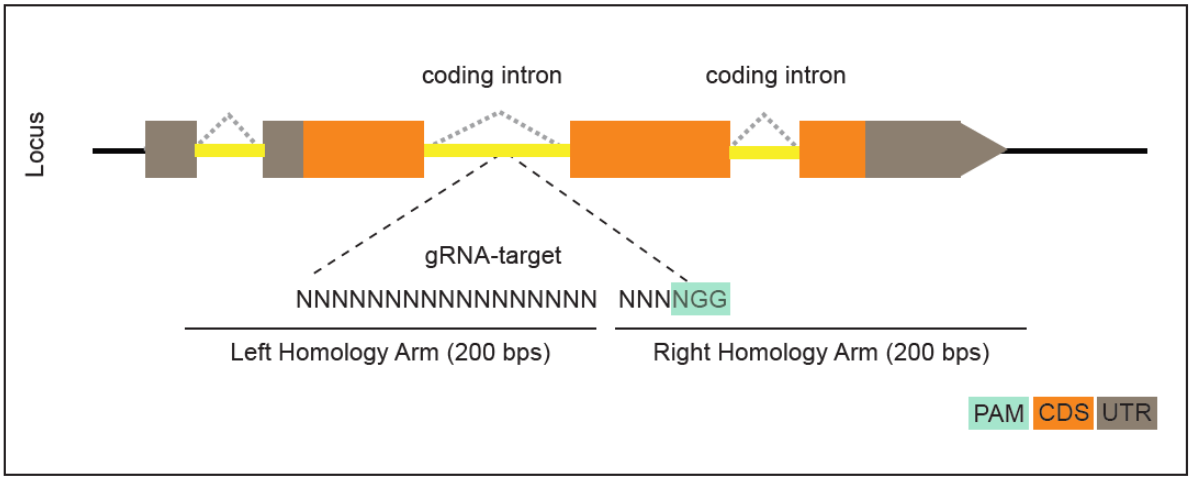
C-Term	
dsRed_CTerm_F	AGGCGCAGCTCTGCGAAATCTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATG
scDsRed_R	TGTGGGCATCGCTTTGGGGATGCTTCTTTGTTCGAGAGATTTAACCCTAGAAAGATAGTCTGCGTAAAATTGAC
sfGFP_Cterm_F	TCGTTCTTCCGAAGGACAATGTGGGAGGTTCCGGTGGAAG
sfGFP_CTerm_R	CTCCGCTTCAGTTGCCTTCACTTGTACAGCTCATCCATGCCCAGGGT
wdr37_3F_CTerm_F	GCATGGATGAGCTGTACAAGTGAAGGCAACTGAAGCGGAG
wdr37_3F_R	GATTATCTTTCTAGGGTTAAGATTTCGCAGAGCTGCGC
wdr37_5F_CTerm_R	CTTCCACCGGAACCTCCCACATTGTCCTTCGGAAGAACGATGG
wdr37_5F_F	GCGAGTCTGCAACCCCTGAACACTAATTGGATTCCCCCGGGCTCTGCCCACC
internal	
sfGFP_int_F	GCATGATTTTCAGGAAGGGCGTGGGAGGTTCCGGTGGAAG
sfGFP_int_R	AGTTCCTCCTCTGAAGAATGCTCGCCGGAACCTCCG
wdr37_3F_int_F	GAAGCGGAGGTTCCGGCGAGCATTCTTCAGAGGAGGAACTGGACG
wdr37_5F_int_R	CTTCCACCGGAACCTCCCACGCCCTTCCTGAAAATCATGCAACG
N-Term	
sfGFP_Nterm_F	GAGGCGAAAAGAAAAGCATGGTGTCCAAGGGCGAG
sfGFP_Nterm_R	GCCCGTGTCTTGCTAGCCTTCTCGCCGGAACCTCCG
wdr37_3F_NTerm_F	GAAGCGGAGGTTCCGGCGAGAAGGCTAGCAAGACACGGGCT
wdr37_5F_Nterm_R	AGCTCCTCGCCCTTGGACACCATGCTTTTCTTTTCGCCTCTTTGT

Supplementary methods Detailed cloning protocol and sequences

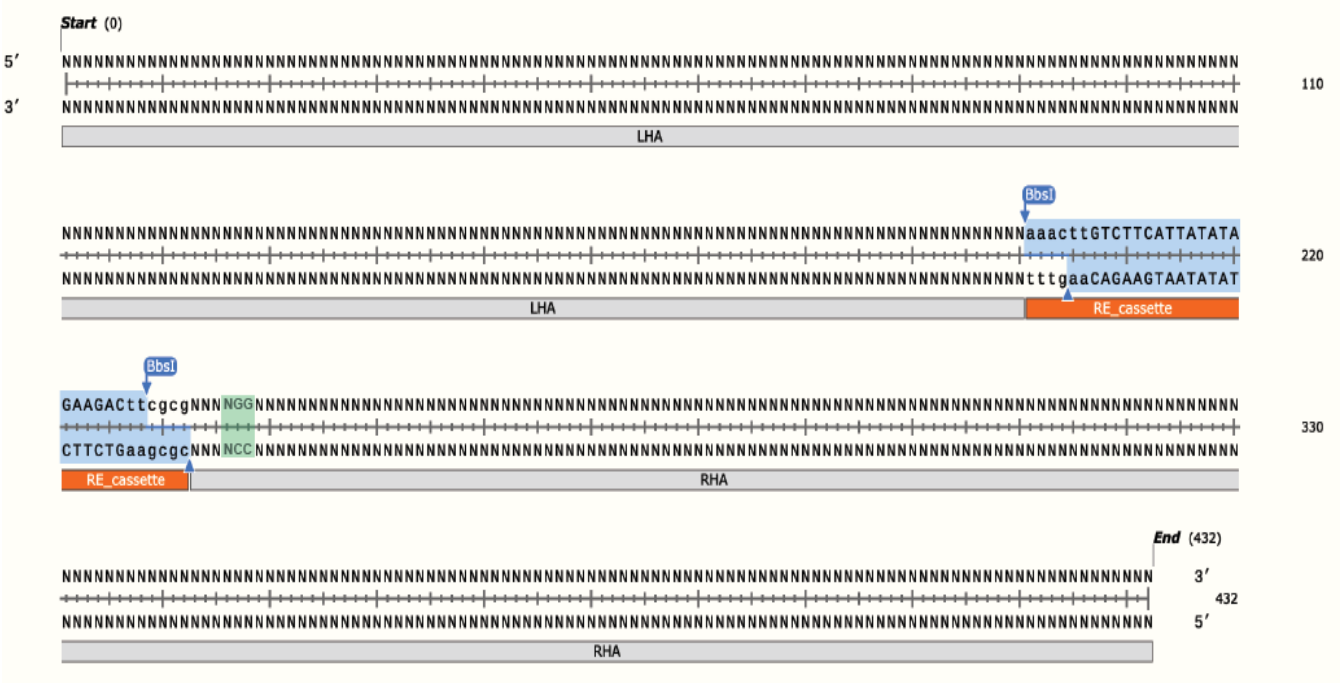
Design of new CRIMIC synthesis constructs:

Int200 strategy

Design region:



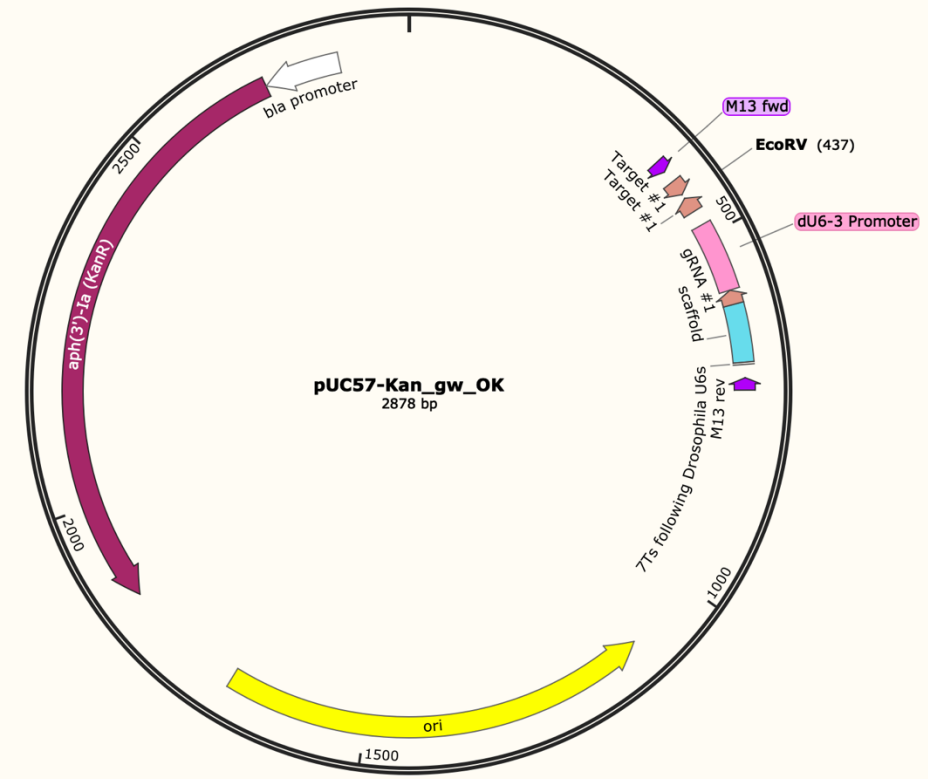
- 200 nucleotides on both sides of the cut site are used for synthesis using the template DNA file (int200_scaffold_BbsI file, in case the homology arms contain BbsI sites, BsaI-HF can be used for cloning. In that case int200_scaffold_BsaI scaffold file should be used for ordering the synthesis)



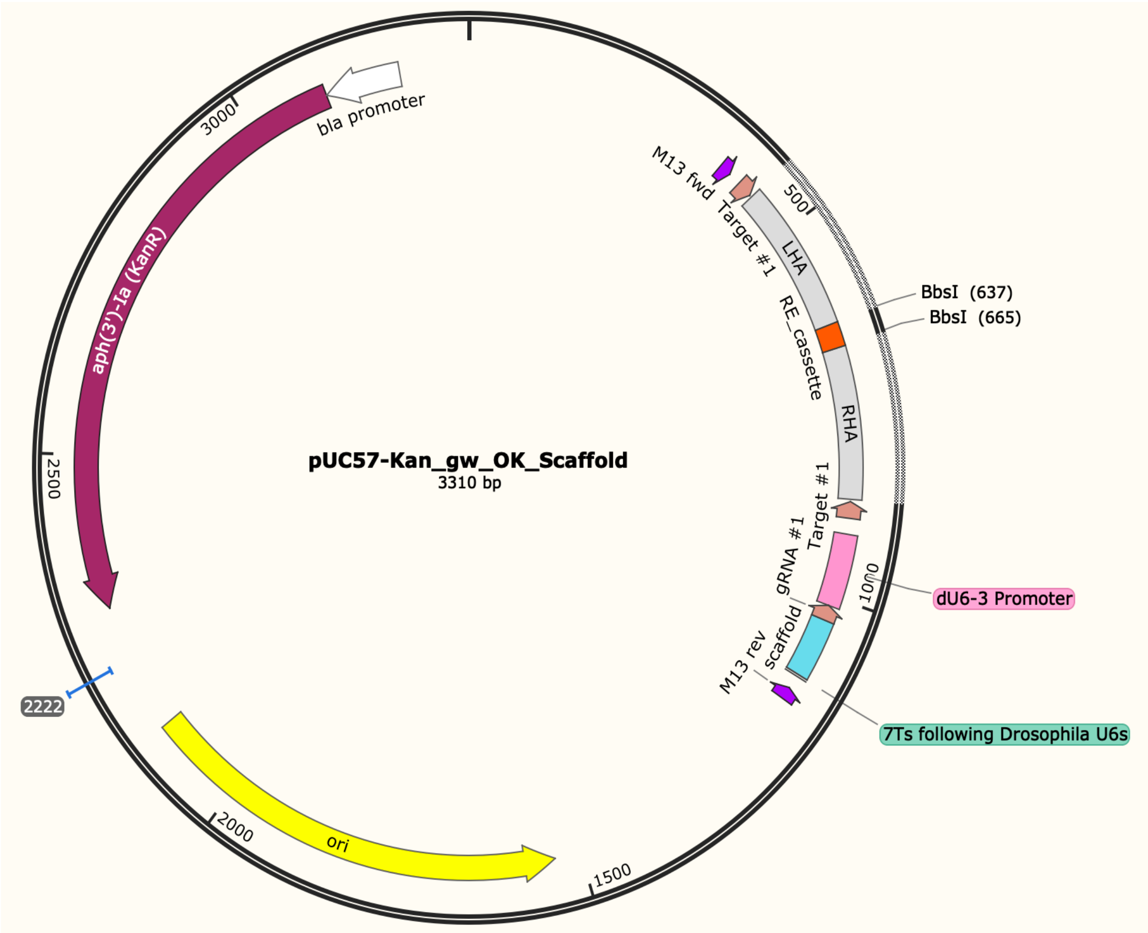
- Synthesis is ordered through Genewiz as Value gene synthesis (at 4nM scale). In order notes indicate:
    - The constructs should be cloned in pUC57_Kan_gw_OK vector in EcoRV site. This vector has the gRNA1 target sites on either side of EcoRV cut site and hence places the homology arms in between

gRNA1 cut sites. In the backbone there is a U6:gRNA1 that linearizes the construct *in vivo*.

- No need to order the glycerol stocks. Lyophilised constructs facilitates the reaction set up.



After synthesis the construct looks like below



gRNA_int200 strategy

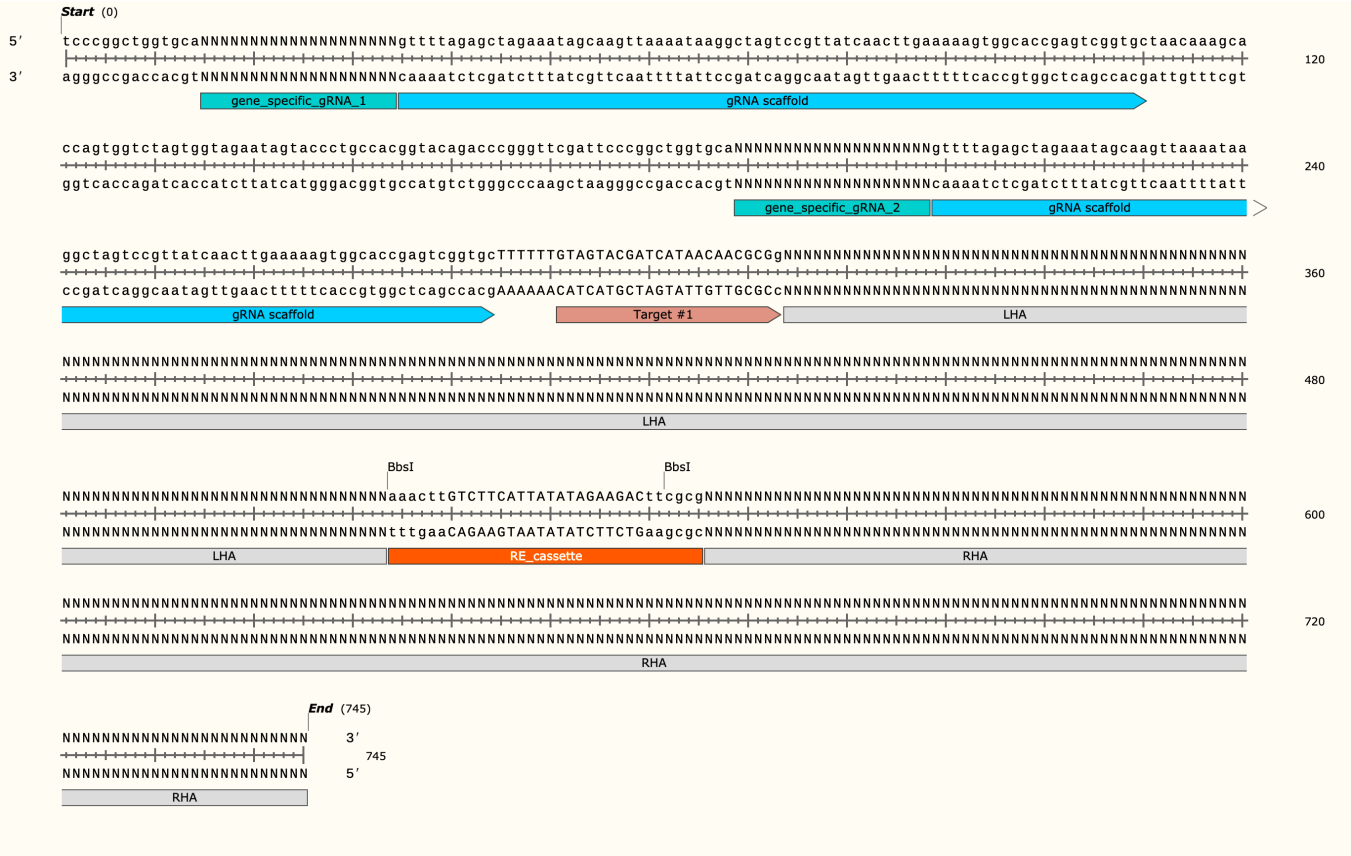
The ordering for gRNA_int200 strategy is similar to int200 strategy with the exception of including the target specific sgRNA sequences in the synthesis order and the synthesis should be conducted on pUC57_Kan_gw_OK2 backbone that contains the rest of the components of the homology donor construct.
For *KozakGAL4* constructs 2 sgRNAs are added in the synthesis reaction.

Cloning of constructs:

-When the constructs arrive, resuspend them in 53µl dH_2O (Do not resuspend in TE buffer. Genewiz lyophylises the constructs in TE buffer. Resuspending in TE buffer decreases efficacy of downstream cloning applications).

-Select the proper vector of pM37 (correct phase for T2AGAL4 or pM37_KozakGAL4 or pM37_SA_KozakGAL4) with BbsI-HF (or BsaI-HF if BsaI construct is being used). Set up the reaction (make a master mix for constructs of the same SIC if cloning multiple constructs):

- 1µl pM37-phase X* (290 ng/µl) or 1 µl pM37_KozakGAL4 (265 ng/µl)
- 2.5 µl 10X T4 DNA ligase buffer (NEB B0202S)
- 0.5 µl T4 DNA ligase (NEB M0202L)
- 1 µl Restriction enzyme (BbsI_HF (NEB R3733L) or BsaI_HFv2 (NEB R3559L)
- 19 µl of dH_2O



Distribute the master mix in PCR tubes and add 1 µl of reconstituted intermediate plasmid.

* Phase selection is done by following the reading frame of the gene at the end of the preceding exon. If the last codon in the preceeding exon is complete (NNN-intron-NNN), the phase is p0. If one of the nucleotide of the codon is in the preceeding exon and two are in the following exon (N-intron-NN) the phase is p1. If two nucleotides of the codon is in the preceeding exon and one in the following exon (NN-intron-N) the phase is p2. It is crucial to select the correct phase using incorrect phase cause a frameshift mutation.

-Incubate the reactions in a Thermocycler:

- 37°C 5 minutes
- 16°C 5 minutes
- Go to 1 30 times
- 65°C 20 minutes
- 8°C hold

The reactions can be left in the thermocycler overnight.

- An additional digestion step is done to remove self ligating plasmid backbones by adding:

- 19.5 µl dH_2O
- 5 µl 10X Cutsmart buffer
- 0.5 µl BbsI or BsaI_HFv2 (the enzyme used for the cloning reaction)

The restriction mix can also be prepared as a mastermix and distributed to each sample 25µl/sample. Incukbate the reactions for 30 minutes in 37°C incubator or thermocycler.

-Transform 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°C overnight.

-(Optional) Next day do colony PCR with primers M13F_Long_for CRIMIC_ch_rev

M13F_long_for gacgttgtaaaacgacggccag
CRIMIC_ch_rev gcggaagagagataaatcggttg

I use an 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_2O
12.5 ul OneTaq Quick-Load 2X Master Mix (M0486L)

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

-The positive colonies will show ~630 bps amplicon.

- The positive colonies are incubated in 5 ml of LB+Kanamycin and miiniprepped using Qiaprep Spin Miniprep kit (Qiagen 27106).

-Resulting DNA is sequenced using M13Reverse and intseq_forward primers (GTTCGATTCCCGGCCGATG)

Vector sequences:

pUC57_Kan_gw_OK

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCA
CAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGT
GTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
CACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
ACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTT
TTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGACGCGTATTGGGTAGTACGAT
CATAACAACGCGGATATCCGCGTTGTTATGATCGTACTACCCAATGGCGCGCCGAGCTTGG
AAGCAGAGAGGGCGCCAGTGCTCACTACTTTTTATAATTCTCAACTTCTTTTTCCAGACTCA
GTTCGTATATATAGACCTATTTTCAATTTAACGTCGTAGTACGATCATAACAACGGTTTTAGA
GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTTGCGGCCGCCTCGAGCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTAT
CCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGG
TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
CCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT
TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGAT
CTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTA
AAGTATATATGAGTAAACTTGGTCTGACAGTTAGAAAAACTCATCGAGCATCAAATGAAACTG
CAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGA
GAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGA
AATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAG
ACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTT
ATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTAC
AAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACC
TGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCAGGGATCGCAGTGGTGAGTA
ACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGT
CAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTT

CAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGC
CCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGC
GGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTAT
TATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

sgRNA1_target sgRNA1 M13Frw primer U6-3 promoter sgRNA scaffold

M13rev primer Ori KanR bla promoter underlined sequences are in -strand

pUC57_Kan_gw_OK2

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCA
CAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGT
GTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
CACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
ACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTT
TTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGACGCGTTTTTTTGCTCACCTGT
GATTGCTCCTACTCAAATACAAAAACATCAAATTTTCTGTCAATAAAGCATATTTATTTATATT
TATTTTACAGGAAAGAATTCCTTTTAAAGTGTATTTTAACCTATAATGAAAAACGATTAAAAAA
AATACATAAAATAATTCGAAAATTTTTGAATAGCCCAGGTTGATAAAAATTCATTTCATACGTT
TTATAACTTATGCCCCTAAGTATTTTTTGACCATAGTGTTTCAATTCTACATTAATTTTACAGA
GTAGAATGAAACGCCACCTACTCAGCCAAGAGGCGAAAAGGTTAGCTCGCCAAGCAGAGA
GGGCGCCAGTGCTCACTACTTTTTATAATTCTCAACTTCTTTTTCCAGACTCAGTTCGTATAT
ATAGACCTATTTTCAATTTAACGTCGGGGCTTTGAGTGTGTGTAGACATCAAGCATCGGTGG
TTCAGTGGTAGAATGCTCGCCTGCCACGCGGGCGGCCCGGGTTCGATTCCCGGCCGATGC
AGTAGTACGATCATAACAACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC
GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTAACAAAGCACCAGTGGTCTAGTGGT
AGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATATCCGCGTTGTTATGATCGTACT
ACGCGGCCGCCTCGAGCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAT
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATG
TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCC
ATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT
GAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAG
TTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAG
CAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTC
TGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
AACTTGGTCTGACAGTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC
AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG
GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCA
ATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTG
ACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGG
CCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATT

GCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGA
ATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATT
CTTCTAATACCTGGAATGCTGTTTTCCCAGGGATCGCAGTGGTGAGTAACCATGCATCATCA
GGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTC
TGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTG
GCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCG
AGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAG
ACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTA
ACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

Intseq_forward sgRNA1_target sgRNA1 part of tRNA sequence M13Frw primer
U6-3promoter sgRNA scaffold M13rev primer Ori KanR bla promoter
underlined sequences are in -strand

pM37_KozakGAL4

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGG
TTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAA
AGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGT
TTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGA
GCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGC
CGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTG
GGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGC
TGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACG
GCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGC
GTCTCTAAACGGTCTCTAAACGAAGACTTAAACCAAAATGAAGCTGCTGAGCAGCATCGAGC
AGGCCTGCGATATCTGCCGCCTGAAGAAGCTGAAGTGCAGCAAGGAGAAGCCCAAGTGCG
CCAAGTGCCTGAAGAACAACTGGGAGTGCCGCTACAGCCCCAAGACCAAGCGCAGCCCCC
TGACCCGCGCCCACCTGACCGAGGTGGAGAGCCGCCTGGAGCGCCTGGAGCAGCTGTTC
CTGCTGATCTTCCCCCGCGAGGATCTGGATATGATCCTGAAGATGGATAGCCTGCAGGATA
TCAAGGCCCTGCTGACCGGCCTGTTCGTGCAGGATAACGTGAACAAGGATGCCGTGACCG
ATCGCCTGGCCAGCGTGGAAACCGATATGCCCCTGACCCTGCGCCAGCACCGCATCAGCG
CCACCAGCAGCAGCGAGGAGAGCAGCAACAAGGGCCAGCGCCAGCTGACCGTGAGCATC
GATAGCGCCGCCCACCACGATAACAGCACCATCCCCCTGGATTTCATGCCCCGCGATGCC
CTGCACGGCTTCGATTGGAGCGAGGAGGATGATATGAGCGATGGCCTGCCCTTCCTGAAAA
CCGATCCCAACAACAACGGCTTCTTCGGCGATGGCAGCCTGCTGTGCATCCTGCGCAGCAT
CGGCTTCAAGCCCGAGAACTACACCAACAGCAACGTGAACCGCCTGCCCACCATGATCACC
GATCGCTACACCCTGGCCAGCCGCAGCACCACCAGCCGCCTGCTGCAGAGCTACCTGAAC
AACTTCCACCCCTACTGCCCCATCGTGCACAGCCCCACCCTGATGATGCTGTACAACAACC
AGATCGAGATCGCCAGCAAGGATCAGTGGCAGATCCTGTTCAACTGCATCCTGGCCATCGG
CGCCTGGTGCATCGAGGGCGAGAGCACCGATATCGATGTGTTCTACTACCAGAACGCCAA
GAGCCACCTGACCAGCAAGGTGTTCGAGAGCGGCAGCATCATCCTGGTGACCGCCCTGCA
CCTGCTGAGCCGCTACACCCAGTGGCGCCAAAAGACCAACACCAGCTACAACTTCCACAGC
TTCAGCATCCGCATGGCCATCAGCCTGGGCCTGAACCGCGATCTGCCCAGCAGCTTCAGC
GATAGCAGCATCCTGGAGCAGCGCCGCCGCATCTGGTGGAGCGTGTACAGCTGGGAGATC
CAGCTGAGCCTGCTGTACGGCCGCAGCATCCAGCTGAGCCAGAACACCATCAGCTTCCCC
AGCAGCGTGGATGATGTGCAGCGCACCACCACCGGCCCCACCATCTACCACGGCATCATC
GAGACTGCCCGCCTGCTGCAGGTGTTCACCAAGATCTACGAGCTGGATAAGACCGTGACC
GCCGAGAAGAGCCCCATCTGCGCCAAGAAGTGCCTGATGATCTGCAACGAGATCGAGGAG
GTGAGCCGCCAGGCCCCCAAGTTCCTGCAGATGGATATCAGCACCACCGCCCTGACCAAC
CTGCTGAAGGAGCACCCCTGGCTGAGCTTCACCCGCTTCGAGCTGAAGTGGAAGCAGCTG

AGCCTGATCATCTACGTGCTGCGCGATTTCTTCACCAACTTCACCCAGAAGAAGAGCCAGC
TGGAGCAGGATCAGAACGATCACCAGAGCTACGAGGTGAAGCGCTGCAGCATCATGCTGA
GCGATGCCGCCCAGCGCACCGTGATGAGCGTGAGCAGCTACATGGATAACCACAACGTGA
CCCCCTACTTCGCCTGGAACTGCAGCTACTACCTGTTCAACGCCGTGCTGGTGCCCATCAA
GACCCTGCTGAGCAACAGCAAGAGCAACGCCGAGAACAACGAAACGGCCCAGCTGCTGCA
GCAGATCAACACCGTGCTGATGCTGCTGAAGAAGCTGGCCACCTTCAAGATCCAGACCTGC
GAGAAGTACATCCAGGTGCTGGAGGAGGTGTGCGCCCCCTTCCTGCTGAGCCAGTGCGCC
ATCCCCCTGCCCCACATCAGCTACAACAACAGCAACGGCAGCGCCATCAAGAACATCGTGG
GCAGCGCCACCATCGCCCAGTACCCCACCCTGCCCGAGGAGAACGTGAACAACATCAGCG
TGAAGTACGTGAGCCCCGGCAGCGTGGGACCCAGCCCCGTGCCCCTGAAGAGCGGCGCC
AGCTTCAGCGATCTGGTGAAGCTGCTGAGCAACCGCCCCCCCAGCCGCAACAGCCCCGTG
ACCATCCCCCGCAGCACCCCCAGCCACCGCAGCGTGACCCCCTTCCTGGGCCAGCAGCAG
CAGCTGCAGAGCCTGGTGCCCCTGACCCCCAGCGCCCTGTTCGGCGGCGCCAACTTCAAC
CAGAGCGGCAACATCGCCGATAGCAGCCTGAGCTTCACCTTCACCAACAGCAGCAACGGC
CCCAACCTGATCACCACCCAGACCAACAGCCAGGCCCTGAGCCAGCCCATCGCCAGCAGC
AACGTGCACGATAACTTCATGAACAACGAGATCACCGCCAGCAAGATCGATGATGGCAACA
ACAGCAAGCCCCTGAGCCCCGGCTGGACCGATCAGACCGCCTACAACGCCTTCGGCATCA
CCACCGGCATGTTCAACACCACCACCATGGATGATGTGTACAACTACCTGTTCGATGATGA
GGATACCCCCCCCCAACCCCAAGAAGGAGTAATAAGGCGCGCCCAAAGATCCAGACATGATA
AGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACA
ACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAG
TAAAACCTCTACAAATGTGGTATGGCTGATTATGATCATAATTCGAATTCGAAGTTCCTATTC
TCTAGAAAGTATAGGAACTTCGCCCGGGGATCTAATTCAATTAGAGACTAATTCAATTAGAG
CTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGA
GGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGC
GAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGACTAGAGCCGGTC
GCCGGCCGGCCACCATGGTGTCCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCAATTC
TGGTGGAGCTGGATGGCGACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGACGCCACCTATGGAAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC
GTGCCATGGCCAACCCTCGTGACCACCCTGACCTATGGCGTGCAGTGCTTCAGCCGCTAC
CCCGATCACATGAAGCAGCACGATTTCTTCAAGAGCGCCATGCCCGAGGGCTACGTGCAG
GAGCGCACCATCTTTTTCAAGGATGACGGCAACTACAAGACCCGCGCCGAAGTGAAGTTCG
AGGGCGATACCCTCGTGAACCGCATCGAGCTGAAGGGCATCGATTTCAAGGAGGATGGAA
ACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCATGGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGATGGCGG
CGTGCAGCTGGCCGATCACTACCAGCAGAACACCCCAATCGGCGACGGCCCAGTGCTGCT
GCCCGATAACCATTACCTGAGCACCCAGAGCGCCCTGAGCAAGGATCCCAACGAGAAGCG
CGACCACATGGTGCTGCTGGAGTTTGTGACCGCCGCCGGCATTACCCTGGGCATGGATGA
GCTGTACAAGTAGGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTA
GAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCAT
TATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGG
GGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATG
ATCAGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTCGCGCTCGCGCGACTGACGGTC
GTAAGCACCCGCGTACGTGCGCGTTGTCTTCCGCGTGAGACCCGCGTGAGACGGATATCA
AGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGT
GAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTAT
CCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG

CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGG
TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
CCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT
TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGAT
CTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTA
AAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTC
AGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGA
TACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGTGACCCACGCTCAC
CGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT
TCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTC
GTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCC
CCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCG
CTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTAC
TTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA
AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGG
GTTCCGCGCACATTTCCCCGAAAAGTGCCAC

FRT SV40polyA GFP 3XP3-Hs70 promoter M13Frw primer KozakGAL4 AmpR

pM37_SA_KozakGAL4

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGG
TTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAA
AGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGT
TTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGA
GCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGC
CGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTG
GGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGC
TGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACG
GCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGC
GTCTCTAAACGGTCTCTAAACGAAGACTTAAACAGTCGATCCAACATGGCGACTTGTCCCAT
CCCCGGCATGTTTAAATATACTAATTATTCTTGAACTAATTTTAATCAACCGATTTATCTCTCT
TCCGCAGCCAAAATGAAGCTGCTGAGCAGCATCGAGCAGGCCTGCGATATCTGCCGCCTG
AAGAAGCTGAAGTGCAGCAAGGAGAAGCCCAAGTGCGCCAAGTGCCTGAAGAACAACTGG
GAGTGCCGCTACAGCCCCAAGACCAAGCGCAGCCCCCTGACCCGCGCCCACCTGACCGA
GGTGGAGAGCCGCCTGGAGCGCCTGGAGCAGCTGTTCCTGCTGATCTTCCCCCGCGAGGA
TCTGGATATGATCCTGAAGATGGATAGCCTGCAGGATATCAAGGCCCTGCTGACCGGCCTG
TTCGTGCAGGATAACGTGAACAAGGATGCCGTGACCGATCGCCTGGCCAGCGTGGAAACC
GATATGCCCCTGACCCTGCGCCAGCACCGCATCAGCGCCACCAGCAGCAGCGAGGAGAGC
AGCAACAAGGGCCAGCGCCAGCTGACCGTGAGCATCGATAGCGCCGCCCACCACGATAAC
AGCACCATCCCCCTGGATTTCATGCCCCGCGATGCCCTGCACGGCTTCGATTGGAGCGAG

GAGGATGATATGAGCGATGGCCTGCCCTTCCTGAAAACCGATCCCAACAACAACGGCTTCT
TCGGCGATGGCAGCCTGCTGTGCATCCTGCGCAGCATCGGCTTCAAGCCCGAGAACTACA
CCAACAGCAACGTGAACCGCCTGCCCACCATGATCACCGATCGCTACACCCTGGCCAGCC
GCAGCACCACCAGCCGCCTGCTGCAGAGCTACCTGAACAACTTCCACCCCTACTGCCCCAT
CGTGCACAGCCCCACCCTGATGATGCTGTACAACAACCAGATCGAGATCGCCAGCAAGGAT
CAGTGGCAGATCCTGTTCAACTGCATCCTGGCCATCGGCGCCTGGTGCATCGAGGGCGAG
AGCACCGATATCGATGTGTTCTACTACCAGAACGCCAAGAGCCACCTGACCAGCAAGGTGT
TCGAGAGCGGCAGCATCATCCTGGTGACCGCCCTGCACCTGCTGAGCCGCTACACCCAGT
GGCGCCAAAAGACCAACACCAGCTACAACTTCCACAGCTTCAGCATCCGCATGGCCATCAG
CCTGGGCCTGAACCGCGATCTGCCCAGCAGCTTCAGCGATAGCAGCATCCTGGAGCAGCG
CCGCCGCATCTGGTGGAGCGTGTACAGCTGGGAGATCCAGCTGAGCCTGCTGTACGGCCG
CAGCATCCAGCTGAGCCAGAACACCATCAGCTTCCCCAGCAGCGTGGATGATGTGCAGCG
CACCACCACCGGCCCCACCATCTACCACGGCATCATCGAGACTGCCCGCCTGCTGCAGGT
GTTCACCAAGATCTACGAGCTGGATAAGACCGTGACCGCCGAGAAGAGCCCCATCTGCGC
CAAGAAGTGCCTGATGATCTGCAACGAGATCGAGGAGGTGAGCCGCCAGGCCCCCAAGTT
CCTGCAGATGGATATCAGCACCACCGCCCTGACCAACCTGCTGAAGGAGCACCCCTGGCT
GAGCTTCACCCGCTTCGAGCTGAAGTGGAAGCAGCTGAGCCTGATCATCTACGTGCTGCGC
GATTTCTTCACCAACTTCACCCAGAAGAAGAGCCAGCTGGAGCAGGATCAGAACGATCACC
AGAGCTACGAGGTGAAGCGCTGCAGCATCATGCTGAGCGATGCCGCCCAGCGCACCGTGA
TGAGCGTGAGCAGCTACATGGATAACCACAACGTGACCCCCTACTTCGCCTGGAACTGCAG
CTACTACCTGTTCAACGCCGTGCTGGTGCCCATCAAGACCCTGCTGAGCAACAGCAAGAGC
AACGCCGAGAACAACGAAACGGCCCAGCTGCTGCAGCAGATCAACACCGTGCTGATGCTG
CTGAAGAAGCTGGCCACCTTCAAGATCCAGACCTGCGAGAAGTACATCCAGGTGCTGGAG
GAGGTGTGCGCCCCCTTCCTGCTGAGCCAGTGCGCCATCCCCCTGCCCCACATCAGCTAC
AACAACAGCAACGGCAGCGCCATCAAGAACATCGTGGGCAGCGCCACCATCGCCCAGTAC
CCCACCCTGCCCGAGGAGAACGTGAACAACATCAGCGTGAAGTACGTGAGCCCCGGCAGC
GTGGGACCCAGCCCCGTGCCCCTGAAGAGCGGCGCCAGCTTCAGCGATCTGGTGAAGCTG
CTGAGCAACCGCCCCCCCAGCCGCAACAGCCCCGTGACCATCCCCCGCAGCACCCCCAGC
CACCGCAGCGTGACCCCCTTCCTGGGCCAGCAGCAGCAGCTGCAGAGCCTGGTGCCCCTG
ACCCCCAGCGCCCTGTTCGGCGGCGCCAACTTCAACCAGAGCGGCAACATCGCCGATAGC
AGCCTGAGCTTCACCTTCACCAACAGCAGCAACGGCCCCAACCTGATCACCACCCAGACCA
ACAGCCAGGCCCTGAGCCAGCCCATCGCCAGCAGCAACGTGCACGATAACTTCATGAACAA
CGAGATCACCGCCAGCAAGATCGATGATGGCAACAACAGCAAGCCCCTGAGCCCCGGCTG
GACCGATCAGACCGCCTACAACGCCTTCGGCATCACCACCGGCATGTTCAACACCACCACC
ATGGATGATGTGTACAACTACCTGTTCGATGATGAGGATACCCCCCCCAACCCCAAGAAGG
AGTAATAAGGCGCGCCCAAAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACC
ACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTG
TAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGT
TCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCT
GATTATGATCATAATTCGAATTCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCG
ATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTC
AATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAG
CTGAACAAGCTAAACAATCGGACTAGAGCCGGTCGCCGGCCGGCCACCATGGTGTCCAAG
GGCGAGGAGCTGTTCACCGGCGTGGTGCCAATTCTGGTGGAGCTGGATGGCGACGTGAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGACGCCACCTATGGAAAGCTGAC
CCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCATGGCCAACCCTCGTGACCAC
CCTGACCTATGGCGTGCAGTGCTTCAGCCGCTACCCCGATCACATGAAGCAGCACGATTTC
TTCAAGAGCGCCATGCCCGAGGGCTACGTGCAGGAGCGCACCATCTTTTTCAAGGATGACG
GCAACTACAAGACCCGCGCCGAAGTGAAGTTCGAGGGCGATACCCTCGTGAACCGCATCG
AGCTGAAGGGCATCGATTTCAAGGAGGATGGAAACATCCTGGGCCACAAGCTGGAGTACAA
CTACAACAGCCACAACGTGTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAAC
TTCAAGATCCGCCACAACATCGAGGATGGCGGCGTGCAGCTGGCCGATCACTACCAGCAG
AACACCCCAATCGGCGACGGCCCAGTGCTGCTGCCCGATAACCATTACCTGAGCACCCAG
AGCGCCCTGAGCAAGGATCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTTGTG
ACCGCCGCCGGCATTACCCTGGGCATGGATGAGCTGTACAAGTAGGATCCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTG

AAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAA
CAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGT
AAAACCTCTACAAATGTGGTATGGCTGATTATGATCAGAAGTTCCTATTCTCTAGAAAGTATA
GGAACTTCTCGCGCTCGCGCGACTGACGGTCGTAAGCACCCGCGTACGTGCGCGTTGTCT
TCCGCGTGAGACCCGCGTGAGACGGATATCAAGCTTATCGATACCGTCGACCTCGAGGGG
GGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC
CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG
TTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTG
ACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCA
AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTAT
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCC
GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC
GAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT
TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTT
ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGTGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCC
ATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTC
GGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAG
CACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT
CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT
CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAG
GAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACT
CTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
GAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

FRT SV40polyA GFP 3XP3-Hs70 promoter M13Frw primer Splice Acceptor
KozakGAL4 AmpR