A genetic toolkit for tagging intronic MiMIC containing genes

Sonal Nagarkar-Jaiswal[1†], Steven Z DeLuca[2†], Pei-Tseng Lee[3], Wen-Wen Lin[3], Hongling Pan[1], Zhongyuan Zuo[3], Jiangxing Lv[3], Allan C Spradling[2], Hugo J Bellen[1,3,4,5,6*]

[1]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, United States; [2]Department of Embryology, Howard Hughes Medical Institute, Carnegie Institution for Science, Baltimore, United States; [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, United States; [4]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, United States; [5]Department of Neuroscience, Baylor College of Medicine, Houston, United States; [6]Program in Developmental Biology, Baylor College of Medicine, Houston, United States

Abstract Previously, we described a large collection of Minos-Mediated Integration Cassettes (MiMICs) that contain two phiC31 recombinase target sites and allow the generation of a new exon that encodes a protein tag when the MiMIC is inserted in a codon intron (*Nagarkar-Jaiswal et al., 2015*). These modified genes permit numerous applications including assessment of protein expression pattern, identification of protein interaction partners by immunoprecipitation followed by mass spec, and reversible removal of the tagged protein in any tissue. At present, these conversions remain time and labor-intensive as they require embryos to be injected with plasmid DNA containing the exon tag. In this study, we describe a simple and reliable genetic strategy to tag genes/proteins that contain MiMIC insertions using an integrated exon encoding GFP flanked by FRT sequences. We document the efficiency and tag 60 mostly uncharacterized genes.
DOI: 10.7554/eLife.08469.001

*For correspondence: hbellen@bcm.edu

†These authors contributed equally to this work

Introduction

One of the most powerful techniques for characterizing gene function is to generate transgenic animals in which an epitope tag such as GFP has been fused to the gene at its normal genomic location (*Ross-Macdonald et al., 1999*; *Morin et al., 2001*; *Skarnes et al., 2004*). These tagged proteins are extremely useful as they permit determination of protein localization in vivo as well as conditional, tissue specific, temporal and reversible removal of the tagged proteins (*Nagarkar-Jaiswal et al., 2015*). However, previous methods for generating protein trap alleles in *Drosophila* have allowed only about 800 genes to be successfully tagged (*Kelso et al., 2004*; *Buszczak et al., 2007*; *Quinones-Coello et al., 2007*; *Aleksic et al., 2009*; *Lowe et al., 2014*).

We previously developed a flexible system for engineering the *Drosophila* genome using the Minos-Mediated Integration Cassette (MiMIC) transposable element. We generated 15,660 strains with a single MiMIC inserted at random within the fly genome and mapped their insertion site (*Bellen et al., 2011*; *Venken et al., 2011*; *Nagarkar-Jaiswal et al., 2015*). MiMIC carries sequences that function as a gene and protein trap when inserted in the proper orientation in a coding intron. Moreover, its content can be replaced by Recombination-Mediated Cassette Exchange (RMCE) leading to the introduction of any desired DNA, such as an artificial exon that encodes a protein tag. This approach can potentially be used to tag thousands of genes. Currently, 2854 existing insertions

are located within the coding introns of 1862 distinct genes (*Nagarkar-Jaiswal et al., 2015*), and MiMIC-like elements can now be placed in any gene of interest by CRISPR (*Zhang et al., 2014*). Unfortunately, the RMCE method needed to convert these insertions into functional protein traps requires embryonic injections of an appropriate donor DNA and screening of many offspring to identify the desired events, a labor and cost-intensive procedure that does not scale easily. We therefore developed a more efficient and economical in vivo genetic tagging methodology that can in principle be used to generate protein trap alleles of all *Drosophila* genes.

Results and discussion

We developed a genetic strategy that allows the desired RMCE event to take place efficiently without the need for microinjection. The method uses FLP recombinase to release a genomically integrated DNA flanked by FRT sites into the nucleoplasm where it can efficiently undergo phiC31 integrase-mediated cassette exchange, as shown by *Gohl et al. (2011)*. As shown in *Figure 1A*, we engineered three donor cassettes, one for each reading frame. The core, which contains a splice acceptor (SA) followed by a (GGS)$_4$ flexible linker, multiple tags (EGFP-FlAsH-StrepII-TEVcs-3xFlag {GFSTF}), another (GGS)$_4$ flexible linker, and a splice donor (SD), is flanked by two inverted *attB* sites for phiC31-mediated RMCE (*Venken et al., 2011*). We then cloned this cassette core between tandem FRT sites in a *P-element* transformation vector (*Gong and Golic, 2003*). FLP-mediated recombination between the tandem FRT sites excises a circular donor DNA molecule from its initial genetic locus, promoting its efficient recombination with a distal target site (*Golic et al., 1997*). A mini-white eye color marker gene between our donor cassette and one of the FRT sites allows us to monitor the presence or absence of the donor cassette in FLP recombinase-containing stocks.

We created 6 stocks (*Figure 1—source data 2*), each harboring one of the three donor transgenes located on the second or third chromosome, and a heat shock-inducible FLP recombinase and a germ line-expressed phiC31 integrase on the X-chromosome. Because the heat shock-inducible FLP recombinase is somewhat leaky at 18°C, the donor transgene is lost from these stocks at a low frequency, resulting in rare white-eyed flies, which we periodically discard.

To initiate RMCE, we crossed the appropriate donor flies to MiMIC-containing flies and heat shocked the resulting embryos and larvae (*Figure 2*). Within the primordial germ cells of some of the embryos and larvae, phiC31 integrase catalyzed recombination between *attB* sites in the donor and *attP* sites in the MiMIC transposon. The positive RMCE events were selected based on the loss of the *y* + marker present in the original MiMIC (*Figure 2*). We confirmed the integration and orientation of the donor cassette by PCR as described in *Venken et al. (2011)*. Typically, 50% of the integration events are in the proper orientation.

We observed one to ten RMCE events in 93 out of 113 attempts in our initial trial when we set up 3–7 crosses (Cross 2 in *Figure 2*). After PCR screening, 60/93 of the tested MiMICs allowed integration of at least one donor in the proper orientation to tag the endogenous gene (*Supplementary file 1*). In summary, we set up 3–7 vials for each starting cross and obtained 60/113 tagged genes. Since the efficiency of RMCE and the ease of detecting yellow⁻ progeny vary between different starting sites, we propose to set up 10–20 vials and to score more progeny to improve the success rate. The method has been found to work for a wide variety of genes including a gene located in a telomeric region (*lethal giant larvae (l(2)gl)*), suggesting that there may be few limitations in its applicability.

To ensure that the expression pattern and protein distribution correspond to the endogenous protein, we costained two tagged lines with GFP for which specific monoclonal antibodies are available: Eyes shut (Eys) (mAb 21A6,) and Delta (Dl) (mAb C594.9B) (*Figure 3A*). In both cases, the protein recognized by the mAb colocalizes with the GFP and match the described expression patterns (*Das et al., 2013*; *Haltom et al., 2014*). Note, however, that the GFP tagged Eys protein is present in the cytoplasm of the photoreceptors and the inter-rhabdomere spaces (IRS) of the photoreceptors, whereas the mAb against Eys mostly localizes to the IRS (*Figure 3A*). These data are in agreement with what we previously observed for numerous tagged proteins (*Venken et al., 2011*; *Nagarkar-Jaiswal et al., 2015*).

We stained third instar larval brains and discs for the 60 tagged gene/proteins. The examples, shown in *Figure 3B*, include *lethal (2) giant larvae (l(2)gl)* (a), *Delta* (*Dl*) (b), and *twins* (*tws*) (c) whose expression patterns are consistent with published data (*Kooh et al., 1993*; *Albertson and Doe, 2003*; *Chabu and Doe, 2009*). Similarly, *kayak/fos* (*kay*) is expressed in wing disc nuclei (d) as described earlier (*Zeitlinger and Bohmann, 1999*). The expression pattern of the remaining genes has not been previously described (*Figure 3B*): *Saposin-related* (*Sap-r*) is expressed in a subset of cells in larval
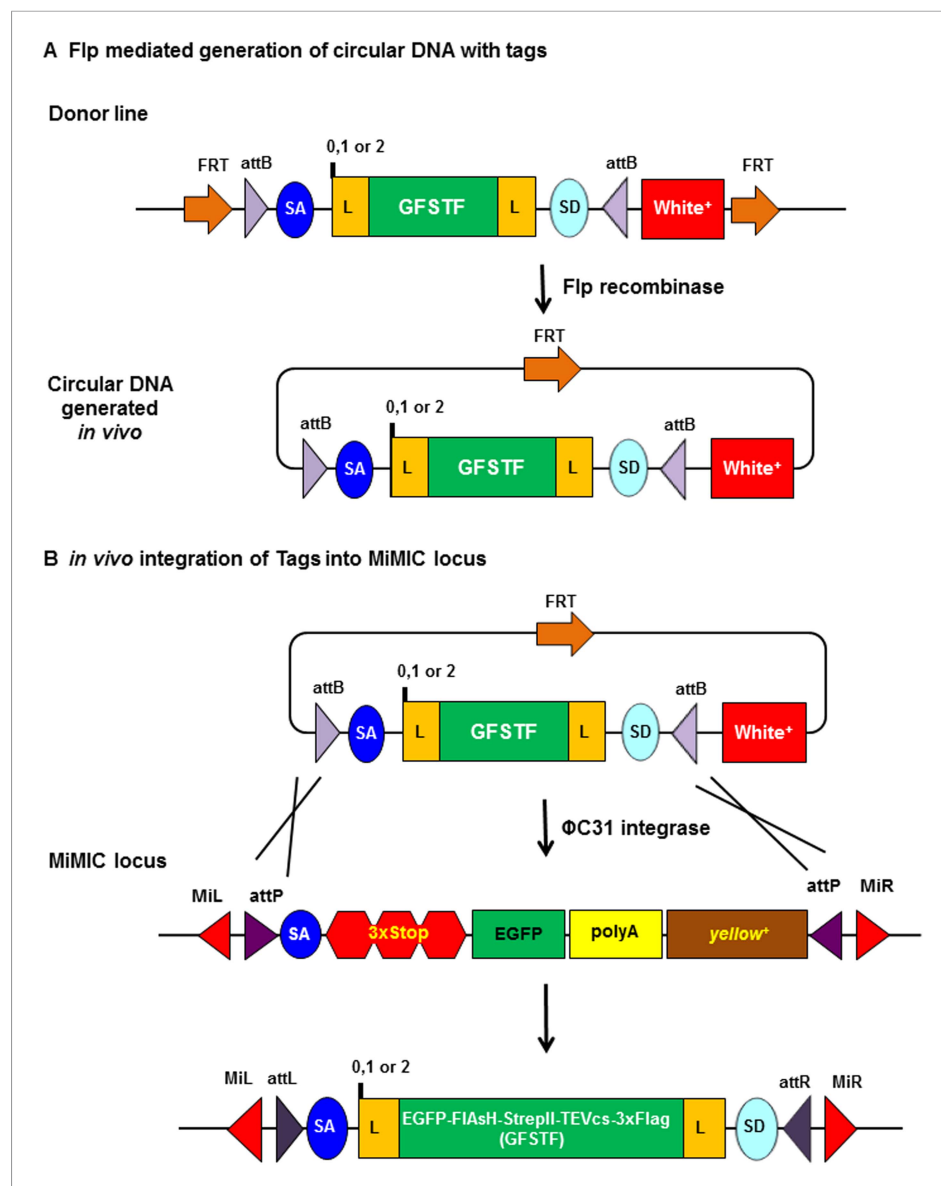
Figure 1. Schematic of flippase and phiC31 integrase mediated in vivo protein tagging. (**A**) Flp mediated generation of circular DNA with tag sequences. Donor line carries a source of Flp recombinase and phiC31 integrase, and the core cassette for a specific phase (0, 1, or 2) encoding a $(GGS)_4$ flexible linker (L), multiple tags EGFP-FlAsH-StrepII-TEVcs-3xFlag (GFSTF), another $(GGS)_4$ flexible linker, and a splice donor (SD), which is flanked by two inverted *attB* and two FRT sites oriented in that same direction. Upon Flp expression, the core is flipped out as a circular DNA containing one FRT site. (**B**) In vivo integration of tag sequence into Minos-Mediated Integration Cassette (MiMIC) locus. At a MiMIC locus in a coding intron, the MiMIC gene trap cassette is replaced by L-EGFP-FlAsH-StrepII-TEVcs-3xFlag-L sequence by phiC31 integrase resulting in loss of *yellow+* marker.

The following source data are available for figure 1:

Source data 1. List of constructs.
Source data 2. List of fly strains generated.

brain (e), *Rad, Gem/Kir family member 3* (*Rgk3*) is enriched in mushroom body in L3 larval brain (f), *Heterogeneous nuclear ribonucleoprotein at 98DE* (*Hrb98DE*) is expressed in L3 larval brain (g), *CG10086* is expressed in hindgut (h), and *CG5656* is expressed in the cells of the cuticle (i).
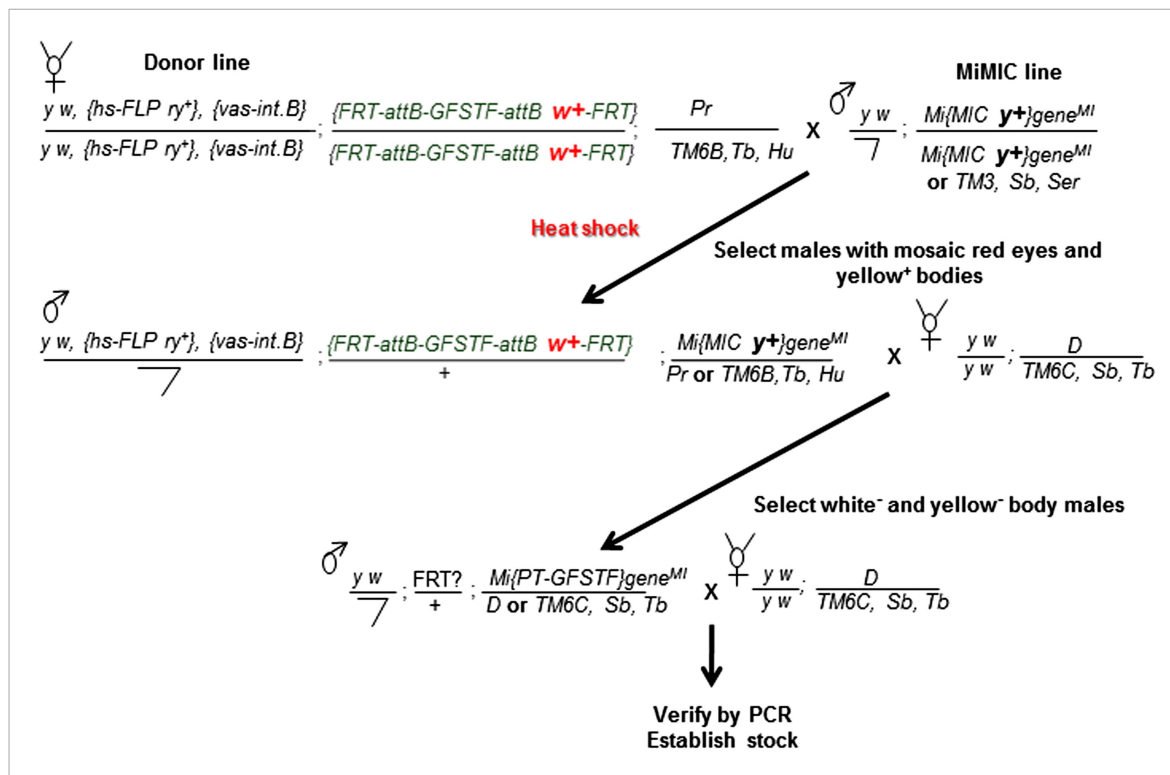
Figure 2. Crossing scheme for generating EGFP tagged MiMIC lines. Females carrying the hs-FLP and *vasa-phiC31* integrase on the X-chromosome, and the FRT flanked multiple tag (GFSTF) cassette on the II-chromosome are crossed to males carrying a MiMIC insertion in a coding intron of a gene on the III-chromosome. The resulting embryos are heat shocked, and adult progeny with the desired genotype are crossed with flies carrying appropriate balancers. The resulting progeny is screened for the loss of *white*+ and *yellow*+ and crossed to flies carrying appropriate balancers to establish stocks, which are then verified by PCR assay.
DOI: 10.7554/eLife.08469.005

The expression patterns of all these genes as well as all the genes listed in *Supplementary file 1* are documented in the MiMIC RMCE database at http://flypush.imgen.bcm.tmc.edu/pscreen/rmce/.

In summary, we developed a genetic tagging strategy that will greatly facilitate the EGFP tagging of nearly 2000 genes that already carry MiMIC insertions (*Nagarkar-Jaiswal et al., 2015*). The same strategies can also be used for tagging genes with other protein tags. In addition, a similar strategy based on *lox* sites instead of *FRT* cassettes has recently been developed to integrate an artificial exon carrying the *GAL4* gene in MiMICs inserted in coding introns (*Diao et al., 2015*). These insertions are mutagenic but permit the expression of the endogenous wild-type and mutant cDNAs of *Drosophila* as well as other species under the control of UAS. Moreover, these tagging methods can now be combined with CRISPR directed integration of *attP* carrying cassettes similar to MiMIC in coding introns to tag almost every gene in *Drosophila* (*Zhang et al., 2014*).

Material and methods

Cloning

The core cassettes (attB-SA- phase 0/1/2-[GGS]$_4$-EGFP-FlAsH-StrepII-TEVcs-3xFlag-[GGS]$_4$-SD-attB) for Phase 0,1, or 2 were excised from pBS-KS-attB1-2-PT-SA-SD-0-EGFP-FlAsH-StrepII-TEVcs-3xFlag, pBS-KS-attB1-2-PT-SA-SD-1-EGFP-FlAsH-StrepII-TEVcs-3xFlag, or pBS-KS-attB1-2-PT-SA-SD-2-EGFP-FlAsH-StrepII-TEVcs-3xFlag as NheI/NsiI fragments and subcloned into *P-element* vector pW35 (DGRC) between PstI/AvrII to create final donor vectors pW35-FRT-attB-SA-phase 0-(GGS)$_4$-EGFP-FlAsH-StrepII-TEVcs-3xFlag-(GGS)$_4$-SD-attB-white+-FRT, pW35-FRT-attB-SA-phase 1-(GGS)$_4$-EGFP-FlAsH-StrepII-TEVcs-3xFlag-(GGS)$_4$-SD-attB-white+-FRT and pW35-FRT-attB-SA-phase 2-(GGS)$_4$-EGFP-FlAsH-StrepII-TEVcs-3xFlag-(GGS)$_4$-SD-attB-white+-FRT.
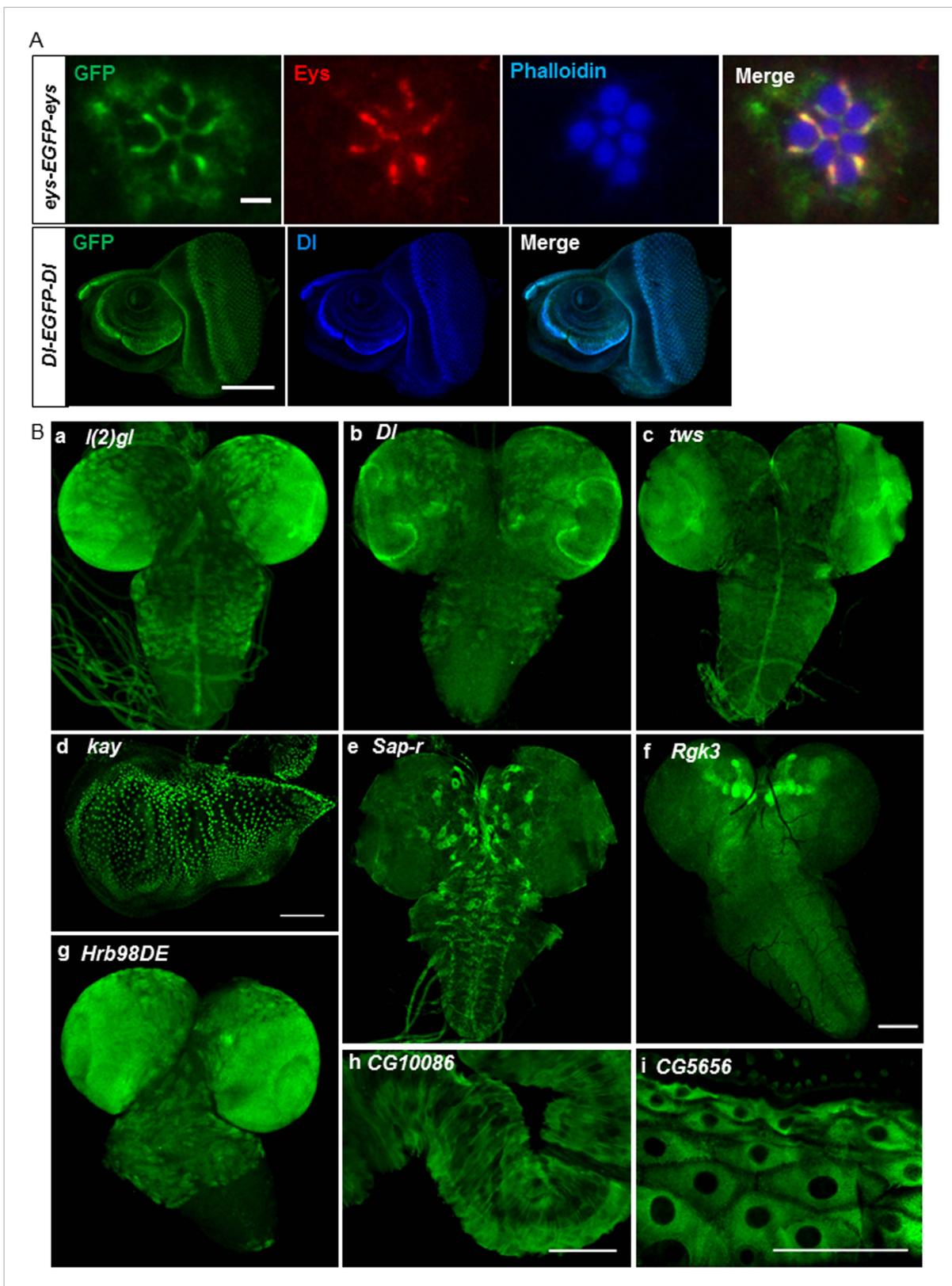
Figure 3. Expression of EGFP tagged protein in various tissues. (**A**) Colocalization of GFP tagged proteins with gene specific antibodies: anti-GFP signals (green) show colocalization with anti-Eys (red) in inter-rhabdomere space in adult eye (top panel) and anti-Delta (blue) in L3 eye imaginal disc in bottom panel. Scale bars, 2 µm and 5 µm. (**B**) Examples of EGFP expression pattern in different tissues from third instar larvae: brain; l(2)gl (**A**), Dl (**B**), twins (tws) (**C**), Sap-r (**E**), Rgk3 (**F**), and Hrb98DE (**G**), wing imaginal disc: kay (**D**), hindgut: CG10086 (**H**), and cells of the cuticle: CG5656 (**I**). Scale bars 100 µm. (**A–I**, except **G**), 25 µm (**G**). Eys, eyes shut.
DOI: 10.7554/eLife.08469.006

In vivo tagging

Around 15–20 phase-specific homozygous donor females (P{ry$^{+t7.2}$= hsFLP}12, y^{1}w*M{vas-int.B}ZH-2A; P{FRT-attB-{GFSTF}-attB(w$^+$)-FRT}; Pri^1/TM6B, Tb^1 or P{ry$^{+t7.2}$ = hsFLP}12, y^1 w*M{vas-int.B}ZH-2A; S^1/CyO; P{FRT-attB-{GFSTF}-attB (w$^+$)-FRT}) were crossed with 5–10 males carrying MiMIC insertion in coding intron (y^1 w*; Mi[MIC y$^+$] gene^{MI}/Mi[MIC y$^+$] gene^{MI} or balancer). Crosses were transferred to new vials every third day and constantly kept at 18°C. Vials with progeny embryos and larvae were heat shocked on day 3, 4, and 5 for 20 minutes at 37°C in a water bath and raised at 25°C. About 5–7 vials with a pool of 5 F1 males with mosaic red eyes and yellow body were crossed with 10–15 virgins of y^{1}w^{67c23}; In(2LR)Gla, wg^{Gla-1}/SM6a or y^{1}w*; D/TM6b, Hu, Tb. Transgenic F2 progeny were screened for loss of yellow$^+$ (yellow-phenotype) and subsequently crossed to virgins of y^{1}w^{67c23}; In(2LR)Gla, wg^{Gla-1}/ SM6a or y^{1}w*; D/TM3,Sb, Tb to establish stocks. Correct RMCE events were confirmed by PCR assay as described in *Nagarkar-Jaiswal et al., 2015*.

Immunostaining

Briefly, third instar larvae were dissected for larval brains, imaginal discs, and gut in 1xPBS and fixed in 3.7% formaldehyde for 30 min at room temperature and washed in 0.2% Triton X-100 (*Nagarkar-Jaiswal et al., 2015*). They were then incubated for 1 hr at RT in 10% NGS-PBS-0.2% Triton X-100 and stained with primary antibodies diluted in 10% NGS-PBS-0.2% Triton X-100 overnight at 4°C. The samples were washed and incubated with secondary antibodies for 2 hr at RT. The samples were then washed and mounted in Vectashield (Vector Labs, Burlingame, CA) and imaged with a Zeiss LSM710 confocal microscope and processed using Adobe Photoshop (Adobe Systems Inc., San Jose, CA, USA).

Antibodies used

Primary antibodies used: rabbit anti-GFP 1:1000 (Life Technologies, A11122), mouse anti-Delta 1:1000 (C594.9B, DSHB [*Qi et al., 1999*]), and mouse anti-Eys 1:250 (21A6, DSHB [*Fujita et al., 1982*]). Secondary antibodies used: Alexa 488 (Invitrogen, Life Technologies, Grand Island, NY), Cy5 and Cy3 conjugated antibodies (Jackson ImmunoResearch, West Grove, PA) were used at 1:500.

Additional information

Author contributions

SN-J, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; SZDL, ACS, HJB, Conception and design, Drafting or revising the article; P-TL, W-WL, HP, ZZ, JL, Acquisition of data, Drafting or revising the article

Additional files

Supplementary file

• Supplementary file 1. Total EGFP tagged genes created: MI: MiMIC insertion, GT: Gene Trap, PT: Protein Trap, Y/N: Yes/No, and L/V: Lethal/Viable
DOI: 10.7554/eLife.08469.007

References

Albertson R, Doe CQ. 2003. Dlg, Scrib and Lgl regulate neuroblast cell size and mitotic spindle asymmetry. *Nature Cell Biology* **5**:166–170. doi: 10.1038/ncb922.

Aleksic J, Lazic R, Muller I, Russell SR, Adryan B. 2009. Biases in *Drosophila melanogaster* protein trap screens. *BMC Genomics* **10**:249. doi: 10.1186/1471-2164-10-249.

Bellen HJ, Levis RW, He Y, Carlson JW, Evans-Holm M, Bae E, Kim J, Metaxakis A, Savakis C, Schulze KL, Hoskins RA, Spradling AC. 2011. The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics* **188**:731–743. doi: 10.1534/genetics.111.126995.

Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A, Wilhelm JE, Murphy TD, Levis RW, Matunis E, Srivali N, Hoskins RA, Spradling AC. 2007. The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* **175**:1505–1531. doi: 10.1534/genetics.106.065961.

Chabu C, Doe CQ. 2009. Twins/PP2A regulates aPKC to control neuroblast cell polarity and self-renewal. *Developmental Biology* **330**:399–405. doi: 10.1016/j.ydbio.2009.04.014.

Das S, Chen QB, Saucier JD, Drescher B, Zong Y, Morgan S, Forstall J, Meriwether A, Toranzo R, Leal SM. 2013. The *Drosophila* T-box transcription factor midline functions within the Notch-Delta signaling pathway to specify sensory organ precursor cell fates and regulates cell survival within the eye imaginal disc. *Mechanisms of Development* **130**:577–601. doi: 10.1016/j.mod.2013.08.001.

Diao F, Ironfield H, Luan H, Diao F, Shropshire WC, Ewer J, Marr E, Potter CJ, Landgraf M, White BH. 2015. Plug-and-play genetic access to *Drosophila* cell types using exchangeable exon cassettes. *Cell Reports* **10**:1410–1421. doi: 10.1016/j.celrep.2015.01.059.

Fujita SC, Zipursky SL, Benzer S, Ferrus A, Shotwell SL. 1982. Monoclonal antibodies against the *Drosophila* nervous system. *Proceedings of the National Academy of Sciences of USA* **79**:7929–7933. doi: 10.1073/pnas.79.24.7929.

Gohl DM, Silies MA, Gao XJ, Bhalerao S, Luongo FJ, Lin CC, Potter CJ, Clandinin TR. 2011. A versatile in vivo system for directed dissection of gene expression patterns. *Nature Methods* **8**:231–237. doi: 10.1038/nmeth.1561.

Golic MM, Rong YS, Petersen RB, Lindquist SL, Golic KG. 1997. FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucleic Acids Research* **25**:3665–3671. doi: 10.1093/nar/25.18.3665.

Gong WJ, Golic KG. 2003. Ends-out, or replacement, gene targeting in *Drosophila*. *Proceedings of the National Academy of Sciences of USA* **100**:2556–2561. doi: 10.1073/pnas.0535280100.

Haltom AR, Lee TV, Harvey BM, Leonardi J, Chen YJ, Hong Y, Haltiwanger RS, Jafar-Nejad H. 2014. The protein O-glucosyltransferase Rumi modifies eyes shut to promote rhabdomere separation in *Drosophila*. *PLOS Genetics* **10**:e1004795. doi: 10.1371/journal.pgen.1004795.

Kelso RJ, Buszczak M, Quinones AT, Castiblanco C, Mazzalupo S, Cooley L. 2004. Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster*. *Nucleic Acids Research* **32**:D418–D420. doi: 10.1093/nar/gkh014.

Kooh PJ, Fehon RG, Muskavitch MA. 1993. Implications of dynamic patterns of Delta and Notch expression for cellular interactions during *Drosophila* development. *Development* **117**:493–507.

Lowe N, Rees JS, Roote J, Ryder E, Armean IM, Johnson G, Drummond E, Spriggs H, Drummond J, Magbanua JP, Naylor H, Sanson B, Bastock R, Huelsmann S, Trovisco V, Landgraf M, Knowles-Barley S, Armstrong JD, White-Cooper H, Hansen C, Phillips RG, UK Drosophila Protein Trap Screening Consortium, Lilley KS, Russell S, St Johnston D. 2014. Analysis of the expression patterns, subcellular localisations and interaction partners of *Drosophila* proteins using a pigP protein trap library. *Development* **141**:3994–4005. doi: 10.1242/dev.111054.

Morin X, Daneman R, Zavortink M, Chia W. 2001. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proceedings of the National Academy of Sciences of USA* **98**:15050–15055. doi: 10.1073/pnas.261408198.

Nagarkar-Jaiswal S, Lee PT, Campbell ME, Chen K, Anguiano-Zarate S, Gutierrez MC, Busby T, Lin WW, He Y, Schulze KL, Booth BW, Evans-Holm M, Venken KJ, Levis RW, Spradling AC, Hoskins RA, Bellen HJ. 2015. A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in *Drosophila*. *eLife* **4**:e05338. doi: 10.7554/eLife.05338.

Qi H, Rand MD, Wu X, Sestan N, Wang W, Rakic P, Xu T, Artavanis-Tsakonas S. 1999. Processing of the notch ligand delta by the metalloprotease Kuzbanian. *Science* **283**:91–94. doi: 10.1126/science.283.5398.91.

Quinones-Coello AT, Petrella LN, Ayers K, Melillo A, Mazzalupo S, Hudson AM, Wang S, Castiblanco C, Buszczak M, Hoskins RA, Cooley L. 2007. Exploring strategies for protein trapping in *Drosophila*. *Genetics* **175**:1089–1104. doi: 10.1534/genetics.106.065995.

Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L, Heidtman M, Nelson FK, Iwasaki H, Hager K, Gerstein M, Miller P, Roeder GS, Snyder M. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**:413–418. doi: 10.1038/46558.

Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, Cox T, Young SG, Ruiz P, Soriano P, Tessier-Lavigne M, Conklin BR, Stanford WL, Rossant J, International Gene Trap Consortium. 2004. A public gene trap resource for mouse functional genomics. *Nature Genetics* **36**:543–544. doi: 10.1038/ng0604-543.

Venken KJ, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M, Carlson JW, Levis RW, Spradling AC, Hoskins RA, Bellen HJ. 2011. MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nature Methods* **8**:737–743. doi: 10.1038/nmeth.1662.

Zeitlinger J, Bohmann D. 1999. Thorax closure in *Drosophila*: involvement of Fos and the JNK pathway. *Development* **126**:3947–3956.

Zhang X, Koolhaas WH, Schnorrer F. 2014. A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in *Drosophila*. *G3* **4**:2409–2418. doi: 10.1534/g3.114.013979.