Novel Genome Engineering Approaches to Understanding Synaptic Function

— by —
Scott J. Gratz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

— in —
GENETICS

— at the —
UNIVERSITY OF WISCONSIN — MADISON

2015

Date of final oral examination: November 20th, 2015

This dissertation is approved by the following committee members:

Grace Boekhoff-Falk — Associate Professor — Cell & Regenerative Biology
Mary Halloran — Professor — Zoology
Allen Laughon — Professor — Genetics
Kate O'Connor-Giles — Assistant Professor — Genetics
Jerry Yin — Professor — Genetics

This work is dedicated to the love of my life, Amanda.

Without your unending love and support, none of this would have been possible.

Abstract

The nervous system is organized in neural circuits that regulate behaviors from simple reflexes to complex traits like learning and memory. These circuits are defined by their synaptic connections where information is transferred from one neuron to another. Each synapse — composed of a presynaptic active zone where neurotransmitter-filled synaptic vesicles are released in response to an action potential and a postsynaptic specialization where neurotransmitter receptors are clustered — is a finely-tuned machine that is independently regulated and modulated throughout the life of the neuron. I am interested in elucidating the molecular mechanisms that underlie the regulation of synapse formation, growth and plasticity. In pursuit of these goals, I have worked to characterize the function of *Fife*, a conserved active zone scaffolding protein, in neurotransmitter release. This work, described in chapters one and two, demonstrates a role in organizing the presynaptic active zone to tightly couple Ca^{2+} channels and synaptic vesicles, a critical parameter for modulating neurotransmitter release probability. I have also developed new genome engineering techniques to facilitate our studies of synaptic function. Chapters three and four describe my work developing the CRISPR-Cas9 system in *Drosophila* for generating targeted modifications, deletions and insertions in a robust and efficient way. Finally, chapter five describes my work applying the CRISPR-Cas9 system to endogenously tag presynaptic proteins for understanding Ca^{2+} channel localization and the establishment of distinct release parameters at individual active zones of single neurons.

Table of Contents

Section 1

Characterizing the role of *Fife* in regulating neurotransmitter release.

Our lab performed an RNAi screen of PDZ-domain containing genes to discover structural proteins important for the organization and function of synapses. This screen uncovered *Fife*, a conserved, but previously uncharacterized, active zone constituent and positive regulator of neurotransmitter release. Upon characterizing the full-length gene, we found that *Fife* bore significant sequence similarity to Piccolo, an active zone protein previously thought to be exclusive to mammalian synapses. This provided us a unique entry point into studies of the functional conservation of the molecular mechanisms that control synaptic function. In collaboration with Joseph Bruckner, a fellow graduate student in our lab, I have worked to characterize *Fife* and its role at synapses. For chapter 1, Joseph and I worked together to design and interpret the experiments. I helped create null alleles of *Fife*, generated transgenic flies for localization and rescue experiments, and performed the light-level phenotypic characterization *Fife* neuromuscular junctions and of Fife localization. For chapter two, I participated in experimental design and analysis. I performed superresolution imaging and light-level analysis of *Fife* mutant active zones and Ca^{2+} channel localization. In ongoing experiments, I am performing live functional imaging of Ca^{2+} influx in *Fife* mutants to extend my analysis of Ca^{2+} channel levels.

Chapter 1

Fife, a Drosophila Piccolo-RIM homolog, promotes active zone organization and neurotransmitter release

Joseph J. Bruckner[1,*], Scott J. Gratz[2,*], Jessica K. Slind[3], Richard R. Geske[3], Alexander M. Cummings[3], Samantha E. Galindo[3], Laura K. Donohue[3] and Kate M. O'Connor-Giles[1-4,**]

[1] Cell and Molecular Biology Training Program, University of Wisconsin-Madison, Madison, WI 53706

[2] Genetics Training Program, University of Wisconsin-Madison, Madison, WI 53706

[3] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

[4] Laboratory of Cell and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706

[*]These authors contributed equally to this work.

This chapter is adapted from the publication:

Bruckner, J J, S J Gratz, J K Slind, R R Geske, A M Cummings, S E Galindo, L K Donohue, and K M O'Connor-Giles. 2012. "Fife, a Drosophila Piccolo-RIM Homolog, Promotes Active Zone Organization and Neurotransmitter Release." *The Journal of Neuroscience* 32 (48): 17048–58. doi:10.1523/JNEUROSCI.3267-12.2012.

Abstract

Neuronal communication depends on the precisely orchestrated release of neurotransmitter at specialized sites called active zones (AZs). A small number of scaffolding and cytoskeletal proteins comprising the cytomatrix of the active zone (CAZ) are thought to organize the architecture and functional properties of AZs. The majority of CAZ proteins are evolutionarily conserved, underscoring the fundamental similarities in neurotransmission at all synapses. However, core CAZ proteins Piccolo and Bassoon have long been believed exclusive to vertebrates, raising intriguing questions about the conservation of the molecular mechanisms that regulate presynaptic properties. Here, we present the identification of a *piccolo-rim*-related gene in invertebrates, together with molecular phylogenetic analyses that indicate the encoded proteins may represent Piccolo orthologs. In accordance, we find that the Drosophila homolog, Fife, is neuronal and localizes to presynaptic AZs. To investigate the in vivo function of Fife, we generated a deletion of the *fife* locus. We find that evoked neurotransmitter release is substantially decreased in *fife* mutants and loss of *fife* results in motor deficits. Through morphological analysis of *fife* synapses, we identify underlying AZ abnormalities including pervasive presynaptic membrane detachments and reduced synaptic vesicle clustering. Our data demonstrate the conservation of a Piccolo-related protein in invertebrates and identify critical roles for Fife in regulating AZ structure and function. These findings suggest the CAZ is more conserved than previously thought, and open the door to a more complete understanding of how CAZ proteins regulate presynaptic structure and function through genetic studies in simpler model systems.

Introduction

Communication within neural circuits occurs primarily at chemical synapses comprising an active zone (AZ) in the signal-sending cell and postsynaptic density (PSD) in the signal-receiving cell. AZs are specialized sites for the regulated release of neurotransmitter via Ca^{2+}-dependent synaptic vesicle fusion. The cytomatrix of the active zone (CAZ) comprises cytoskeletal and scaffolding proteins. This highly interconnected molecular machine is thought to coordinate the organization of AZs and the docking and priming of fusion-competent synaptic vesicles to establish release dynamics. In vertebrates, core CAZ proteins include Piccolo, Bassoon, Rab3-interacting molecules RIM1 and RIM2, UNC-13 and CAST (Jin and Garner, 2008). The Drosophila genome encodes a CAST ortholog (Bruchpilot), UNC-13 (DUNC-13), and a single RIM (DRIM) (Aravamudan et al., 1999; Wang and Sudhof, 2003; Wagh et al., 2006). In contrast, homologs of the related Piccolo and Bassoon proteins have not previously been identified in invertebrates.

Piccolo is a large scaffolding protein containing two N-terminal zinc finger (ZF) domains, a large central region with three coiled-coil (CC) domains, and a C-terminal PDZ domain – all of which mediate protein-protein interactions, as well as two C-terminal C2 domains (C2A and C2B) that confer lipid binding and Ca^{2+} sensitivity to Piccolo function (See Fig. 2B) (Wang et al., 1999; Fenster et al., 2000). Piccolo is related to Bassoon, which shares its large size, ZF and CC domains, and to RIM-family proteins (Wang et al., 1997; tom Dieck et al., 1998). Vertebrate genomes encode two full-length RIM proteins containing ZF, PDZ, C2A and C2B domains with significant similarity to the corresponding Piccolo domains.

Studies of Piccolo function in mice have been inconclusive, with RNAi experiments suggesting a unique role in synaptic vesicle recruitment and knockout studies indicating a redundant role with Bassoon in organizing synaptic vesicle pools (Leal-Ortiz et al., 2008; Mukherjee et al., 2010; Waites et al., 2011). Bassoon regulates AZ structure, synaptic vesicle clustering/docking, and the localization of Ca^{2+} channels at ribbon synapses (Dick et al., 2003; Frank et al., 2010; Hallermann et al., 2010). Recent studies, including the double knockdown of Piccolo and Bassoon in cortical neurons, suggest a role in regulating vesicle dynamics at central synapses as well (Hallermann et al., 2010; Mukherjee et al., 2010). RIM proteins play key roles in regulating vesicle docking and priming and Ca^{2+} channel clustering at both invertebrate and vertebrate synapses (Koushika et al., 2001; Weimer et al., 2006; Gracheva et al., 2008; Deng et al., 2011; Han et al., 2011; Kaeser et al., 2011; Stigloher et al., 2011).

Here we describe the identification of a Piccolo-RIM-related protein in invertebrates. Through genetic studies in Drosophila, we demonstrate a critical role for the Drosophila homolog, Fife, in determining the architecture and function of presynaptic terminals. Expressed in neurons, Fife colocalizes with the CAZ protein Bruchpilot at AZs. Flies lacking *fife* exhibit ultrastructural abnormalities at AZs and deficits in synaptic vesicle clustering. Evoked neurotransmission is reduced to a third of wild-type levels at *fife* NMJs and motor function is significantly impaired in the absence of *fife*.

Materials and methods

Molecular characterization of *fife*

Genome annotations of the *fife* locus predicted three candidate genes: CG12187, CG16976, and CG14950 (Tweedie et al., 2009). RT-PCR of products spanning the junctions of the three candidate genes revealed that they are a single transcriptional unit. Specifically, we purified wild-type RNA using the RNeasy Mini Kit (Qiagen). Following, DNase digestion, isolated RNA was used to make cDNA with an Oligo$(dT)_{20}$ primer (Invitrogen) and SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. The resulting cDNA was used as a template for PCR with primer pairs spanning CG12187 to CG16976, CG12187 to CG14950, and CG16976 to CG14950. To determine the full-length sequence of *fife*, we employed 5' and 3' RACE using RNA purified from flies at multiple developmental stages. To generate 1st strand cDNA tagged with specific sequences at the 5' and 3' ends, we conducted RNA ligase-mediated rapid amplification of 5' and 3' cDNA ends (RLM-RACE) with the GeneRacer Kit (Invitrogen). Sequence analysis identified two previously unknown 5' UTR exons and confirmed that CG12187, CG16976 and CG14950 comprise a single gene. We also identified alternative isoforms of *fife* that encode proteins lacking the C2B domain (Isoform 2) and both C2 domains (Isoform 3). Note that since we conducted our analysis of the *fife* locus, the Drosophila genome annotation has been independently updated to reflect the link between the three candidate genes and the locus has been renamed CG43395 (FB2012_04, released July 6th, 2012) (McQuilton et al., 2012).

Molecular phylogeny

Invertebrate Piccolo/Fife-related proteins were identified through BLAST searches combined with genomic analysis to identify synteny in the arrangement of CG12187, CG16976 and CG14950-related genes. Functional domains were identified and

delineated using SMART and Pfam 25.0 (Schultz et al., 1998; Letunic et al., 2012; Punta et al., 2012). We have so far not identified sequences related to C2B-encoding CG14950 in *Anopheles gambiae*. The *Tribolium castaneum* CG16976-related sequence lacks a 23-amino acid conserved region of the C2A domain raising the possibility that an additional exon may await identification. We used T-Coffee to align sequences (Notredame, 2010). Phylogeny was determined from aligned sequences with gaps removed via maximum likelihood (PhyML), neighbor joining (BioNJ) and bayesian inference (MRBayes) (Ronquist and Huelsenbeck, 2003; Zelwer and Daubin, 2004; Guindon et al., 2005; Dereeper et al., 2008). Confidence levels are based on 500 bootstrap replicates.

Fly stocks

The following lines were used: w^{1118} (wild-type), *C155-Gal4* (Lin and Goodman, 1994) and Df(3L)BSC412 (Bloomington Drosophila stock center). All crosses were conducted at 25°C with 10 males and 10 females.

Generation of *fife* alleles

A line carrying a Minos element inserted 3' of *fife* isoform 3 in a w^{1118} background, Mi{ET1}CG43375^{MB10889}, was generated by the Drosophila Gene Disruption Project and obtained from the Bloomington Drosophila Stock Center (Bellen et al., 2011). We excised the element as described except that the excision was performed with the Minos element in trans to a *fife* deficiency chromosome (Metaxakis et al., 2005). Imprecise excisions were identified and mapped by PCR and subsequent sequence analysis. To assess transcripts in $fife^{ex1027}$, we purified RNA from adult heads of w^{1118} and $fife^{ex1027}$ and generated cDNA as described above. The resulting cDNA was used as a template for PCR with primers corresponding to *fife* exons 3 and 14, which flank the genomic excision

event, and products sequenced. A primer pair for *β-tubulin* was included in these reactions to serve as a loading control.

Transgenic flies

UAS-venus-fife flies were generated through Gateway cloning (Invitrogen) of the *fife* isoform 1 coding sequence into the pBI-UASC-VG vector (Wang et al., 2012) and PhiC31 integration into the attP2 site.

In situ hybridization and immunohistochemistry

RNA in situ hybridization was performed as described in Lehmann and Tautz (1994). NMJ dissections and antibody stains were performed as previously described (O'Connor-Giles et al., 2008). We used the following antibodies: chicken anti-GFP at 1:500 (Abcam #13790), mouse anti-Bruchpilot at 1:100 (NC82; generated by Erich Buchner and available through the Developmental Hybridoma Studies Bank), and rabbit anti-GluRIIC at 1:2500 (Marrus et al., 2004). Species-specific Alexa-488 and Alexa-568 secondary antibodies (Invitrogen) were used at 1:500. Anti-HRP conjugated to Cy3 or Alexa Fluor 647 was used at 1:500 and 1:100, respectively (Jackson ImmunoResearch).

Confocal imaging and analysis

Confocal images were obtained on a Zeiss LSM510 and processed in the Fiji distribution of ImageJ (Schindelin et al., 2012). Overall brightness and/or contrast were adjusted in Photoshop. For morphological analysis, dissections were triple labeled for Brp, GluRIIC, and HRP and all genotypes were stained in the same dish. All genotypes for an individual experiment were imaged with identical settings. A total of 8-9 NMJs from at least 6 animals were compared for each genotype. Data are presented as the mean ± SEM of three independent experiments. For consistency, analysis was limited to NMJ 4 in

segment A3. To determine AZ/PSD area and intensity, non-synaptic structures including axons were removed from the images using freehand selection and fill. Channels were separated and a threshold was applied to remove irrelevant lower intensity pixels. Separation of individual AZs and PSDs was facilitated by the Find Maxima tool in Fiji. Intensity and area data were collected using the Analyze Particles tool. Results were exported to Excel for analysis. HRP intensity and area were analyzed similarly except that no segmentation was performed.

Electron microscopy and analysis

Third instar larvae were dissected and fixed overnight in 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1M sodium cacodylate buffer at 4°C. Samples were then washed in sodium cacodylate and postfixed in 2% osmium tetroxide for 1 hour at room temperature. Following ethanol dehydration, samples were infiltrated and flat embedded in Epon resin. Trimmed blocks were sectioned on a Leica EM UC6 ultramicrotome. Ultrathin (grey-silver) sections were collected on pioloform-coated copper slot grids and stained with uranyl acetate and Reynold's lead citrate. Images were collected on either a Phillips CM120 or JEOL 100CX transmission electron microscope. To quantify mean vesicle density, total synaptic vesicles were counted in a central cross-section and bouton area was measured in Fiji. The AZ was defined as the length of presynaptic membrane in close opposition to a postsynaptic density. Total AZ length was measured summing the lengths of five line segments tracing the curved presynaptic membrane. The distance of each synaptic vesicle to the closest point on any of the five line segments was measured, and vesicles within 200 nm were counted. Floating t bars and ruffling of the presynaptic membrane were scored only when observable in at least two consecutive sections.

Electrophysiology

Third-instar larvae were dissected in Ca^{2+}-free HL-3 saline (70 mM NaCl, 5 mM KCl, 20 mM $MgCl_2$, 10 mM $NaHCO_3$, 115 mM Sucrose, 5 mM Trehalose, 5 mM HEPES, pH 7.2). For recordings, dissections were bathed in HL-3 containing 0.5 mM Ca^{2+}. Using glass microelectrodes filled with 3 M KCl, recordings of spontaneous miniature excitatory junction potentials (mEJPs) and evoked potentials (EJPs) were obtained from muscle six of the third abdominal segment. For each cell, 60 consecutive mEJPs were collected with pClamp (Molecular Devices) and analyzed using MiniAnalysis (Synaptosoft). mEJPs with slow rise times were excluded from the analysis as background from neighboring muscles. mEJPs were then averaged to obtain mean amplitude and frequency. EJPs were evoked by suctioning the cut end of the appropriate segmental nerve into a stimulating electrode filled with bath solution and applying stimulus at half-second intervals. For each cell, stimulus amplitude was adjusted so that both the 1s and 1b nerve inputs were recorded, and EJPs were analyzed in pClamp to determine mean EJP amplitude. All cells measured had a mean resting potential between -60 and -80 mV and input resistance greater than 5 MΩ. At least 6 animals were analyzed for each genotype.

Behavioral assays

Locomotion was assessed 24-48 hours after eclosion. Single flies with clipped wings were placed in a petri dish marked with a 2 cm^2 grid and allowed to acclimate for 30 seconds. The fly was then tapped to the center of the dish and allowed to walk freely for 30 s while the number of line crosses was recorded. This analysis was repeated three times for each fly and averaged (Wagh et al., 2006). Ten flies per genotype were assessed.

Statistical analyses

Single comparisons were conducted by Student's t test with Welch's correction where appropriate. For multiple comparisons, we performed analysis of variance (ANOVA)

followed by post hoc tests with Bonferroni correction to maintain an experimentwise significance level of 0.05. We report the significance levels at the corresponding experimentwise levels of less than 0.05, less than 0.01, less than 0.001, or less than 0.0001 by one, two, three and four stars, respectively. Significance levels refer to the comparison of the indicated genotype to wild type, except in the case of rescue, which reflects the comparison to *fife*^{ex1027}*/Df*. Error bars denote SEM.

Results

Identification of invertebrate Piccolo-related proteins

As part of an effort to analyze the role of Drosophila PDZ domain proteins at synapses, we noted a previously uncharacterized Drosophila candidate gene that encodes a protein with sequence similarity to the ZF and PDZ domains of Piccolo, previously thought to be present only in vertebrates (CG12187; Fig. 1A,B). The Drosophila CG12187 PDZ domain shares 43% identity and 64% similarity with the human Piccolo PDZ domain (Table 1). Given the significant similarity, we explored the possibility that *piccolo* was not absent from the Drosophila genome, but rather overlooked due to misannotation. Consistent with this possibility, we identified two downstream candidate genes predicted to encode proteins with sequence similarity to the C-terminal C2A and C2B domains of Piccolo-RIM proteins (CG16976 and CG14950; Fig. 1B). Through RT-PCR and 5'- and 3'-RACE, we determined that the three annotated genes are in fact a single transcript that encodes a full-length protein with significant similarity to both Piccolo and RIM proteins (Fig. 1C,D and Table 1). We identified two additional isoforms including isoform 2, which lacks only the C2B domain and corresponds to the most prevalent isoform of vertebrate Piccolo (Wang et al., 1999) (Fig. 1 C,D). However, our comprehensive computational and molecular investigations of the locus have not identified any coding sequence corresponding to the large CC-containing region of Piccolo. These findings are supported by genome-wide RNA-SEQ data, so we believe we have identified the full coding sequence and conclude that the Drosophila genome encodes a Piccolo-RIM-related protein lacking Piccolo's CC domains, but sharing the ZF, PDZ, C2A and C2B functional domains that uniquely define Piccolo-RIM proteins (Drysdale, 2008; Celniker et al., 2009).

We next sought to identify the novel *piccolo-RIM*-related gene in other invertebrates. In addition to a single previously identified RIM ortholog, UNC-10, the *C. elegans* and *C. briggsae* genomes each encode uncharacterized Piccolo-related proteins that contain a PDZ domain and a single C2 domain (F45E4.3 and CBG05923, respectively). Homologs with a similar size and genomic organization are also found in members of other nematode genera, including *Pristionchus pacificus* and *Haemonchus contortus* (James Rand, personal communication). Turning to insects, we looked for related sequences in the mosquito *Anopheles gambiae*, which diverged from Drosophila 250 million years ago, the honeybee *Apis mellifera* and the beetle *Tribolium castaneum*, both of which are approximately 300 million years diverged from Drosophila. With the exception of *Anopheles*, in which we did not identify the C2B-encoding CG14950-related gene, the full-length CG12187/CG16976/CG14950-related gene is conserved (Fig. 2A,B). Thus *piccolo-RIM*-related genes are broadly conserved in invertebrate genomes.

Piccolo, RIM1 and RIM2 exhibit significant sequence similarity in their common functional domains, but their overall structure differs due to the presence of a second ZF and large CC-containing region in Piccolo (Fig. 2B). This results in a substantial size difference between Piccolo and RIM proteins: the largest human Piccolo isoform contains more than 5,000 amino acids, while full-length human RIM proteins comprise 1350-1700 amino acids. Lacking the second ZF and CC regions of vertebrate Piccolo, CG12187/CG16976/CG14950 encodes a 1300-amino acid protein more similar in size to vertebrate RIM proteins (Fig. 2B). However, outside of the conserved domains RIM proteins themselves vary significantly in size. For example, at nearly 3,000 amino acids, Drosophila RIM is roughly twice the size of human RIM proteins. Thus, interdomain sequences have diverged significantly. Nonetheless, the structural considerations together with the significant homology CG12187/CG16976/CG14950 exhibits with both

Piccolo and RIM proteins raise the question of whether the newly identified protein more likely represents a second invertebrate RIM or a Piccolo ortholog.

To investigate these evolutionary alternatives, we conducted a molecular phylogenetic analysis of conserved Piccolo-RIM-Fife domains. Using sequences from representative vertebrates and invertebrates, we aligned synthetic proteins constructed via concatemerization of the conserved ZF/PDZ/C2A/C2B domains (T-Coffee) (Notredame, 2010). We then employed maximum likelihood (PhyML), neighbor joining (BioNJ) and bayesian inference (MRBayes) approaches to construct phylogenetic trees using the PDZ-domain-containing Dishevelled family of proteins as the outgroup, with the three approaches yielding the same results (Ronquist and Huelsenbeck, 2003; Zelwer and Daubin, 2004; Guindon et al., 2005; Dereeper et al., 2008). This analysis places Fife proteins in the same clade as Piccolo with bootstrap support of 78%. Bootstrap values are a conservative estimate of accuracy, and empirical tests indicate that bootstrap values above 70% are usually accurate with >95% confidence (Hillis et al., 1993). For estimating phylogeny, the most conserved sequences yield the best alignments and most accurate trees. Because the individual conserved domains have different levels of similarity to vertebrate Piccolo-RIM proteins, we also constructed phylogenetic trees from separate alignments of the PDZ, C2A and C2B domains with the Dishevelled PDZ domain, Synaptotagmin I C2A or Synaptotagmin I C2B domains, respectively, as outgroups. Trees generated from alignment of the well-conserved PDZ domains place Fife proteins in the same clade as Piccolo with robust bootstrap support of 89% (Fig. 2C). Trees based on alignments of the less conserved C2 domains place Fife closer to RIM proteins (C2A), or in the Piccolo clade with lower bootstrap support of 59% (C2B). Taken together, these results are most consistent with the model that the divergence of the Piccolo and RIM gene families is ancestral to the divergence of vertebrates and invertebrates, and that

CG12187/CG16976/CG14950 is the Drosophila ortholog of vertebrate Piccolo. While these analyses leave open the possibility of more complex models, this model is consistent with the previous identification of DRIM as the sole Drosophila RIM ortholog and the conclusion that RIM1 and RIM2 arose from a later duplication in vertebrates (Wang and Sudhof, 2003). Finally, an investigation of conserved sequences outside of the ZF/PDZ/C2A/C2B domains leads to a similar conclusion: (i) Piccolo and CG12187/CG16976/CG14950 contain glutamine-rich N-terminal sequences not found in RIM proteins, (ii) CG12187/CG16976/CG14950 does not contain the N-terminal α-helix, that mediates Rab3-binding and is conserved in invertebrate and vertebrate RIM proteins, and (iii) CG12187/CG16976/CG14950 lacks the conserved SH3-binding domain that mediates binding between RIMs and RIM-Binding Proteins (Fig. 2B). While a final determination of evolutionary and functional orthology awaits further study, we have named CG12187/CG16976/CG14950 *fife* to reflect its homology to Piccolo-RIM proteins and its smaller size relative to Piccolo.

Fife is an active-zone protein

Only a small number of proteins localize to the CAZ. To determine if Fife is among these, we monitored cellular expression and subcellular localization during development. Through in situ hybridization we determined that *fife* is broadly expressed in the CNS, with expression initiating in embryonic post-mitotic neurons at approximately stage 15 and continuing through embryogenesis (Fig. 3A). Attempts to raise antibodies against Fife for assessing protein localization have been unsuccessful, so to determine the subcellular localization of Fife, we generated *UAS-venus-fife* transgenic flies. Neuronal expression of Venus-Fife rescues behavioral deficits in *fife* mutants indicating that the tagged protein localizes to the site of endogenous Fife function in neurons (Fig. 8F, $p < 0.001$). At the larval neuromuscular junction (NMJ), Venus-Fife overlaps significantly

with Bruchpilot, an integral component of electron-dense cytoplasmic projections, called T bars, that cluster synaptic vesicles at NMJ AZs (Fig. 3B). We observe similar results using the smaller Flag tag (data not shown). Thus, Fife, like its vertebrate counterparts, is a CAZ protein. This places Fife in the company of a small number of AZ-localized scaffolding proteins that mediate the organization and function of synapses and suggests Fife may play an important role in these processes.

Synapse assembly is not significantly impaired in *fife* mutants

To study the in vivo function of Fife, we generated loss-of-function alleles by screening for imprecise excisions of a Minos element inserted in the *fife* locus (Fig. 1B). We identified a 16.5-kb deletion, $fife^{ex1027}$, that removes all but the first 11 amino acids of Fife isoform 3 and the splice donor site for isoforms 1 and 2 (Fig. 1B (black bar),E). Based on the extent of the deletion, $fife^{ex1027}$ is likely a strong hypomorph or null allele. To test this, we compared the $fife^{ex1027}$ homozygous phenotype to its phenotype over a chromosomal deficiency (Df(3L)BSC412, which removes *fife* and four other genes) in morphological, functional and behavioral assays. These phenotypes are not enhanced over the deficiency (Figs. 4, 5 and 8D-F; $p > 0.05$ in all cases); thus, $fife^{ex1027}$ behaves as a genetic null.

While our genetic analysis demonstrates that Fife function is lost in $fife^{ex1027}$, the nature of the deletion leaves open the possibility that the C2A and C2B-encoding exons may be expressed. We performed RT-PCR to assess this possibility and found that exons encoding both C2 domains are in fact transcribed. Two models could explain this result: (1) our analysis missed a C2A/C2B-only isoform of *fife* with a start site outside of the deleted region or (2) by bringing exons 3 and 13 in close proximity, the genomic deletion results in the generation a ectopic isoform. While we did not find evidence to support the first model – either in our experimental analysis of the locus or in RNA-SEQ data

(Drysdale, 2008; Celniker et al., 2009), our RT-PCR results confirm the second model (Fig. 1E). In wild type, RT-PCR with primers designed to amplify exons 3 through 14, which flank the genomic deletion in *fife*^{ex1027}, yields a 2130 base-pair product corresponding to full-length *fife* isoforms. The same reaction in *fife*^{ex1027} fails to amplify the wild-type product, confirming the absence of endogenous *fife* isoforms, but instead yields a 240 base-pair product. Sequence analysis demonstrates that this product is the result of an ectopic splice event between exons 3 and 13. This ectopic splice event is frame conserving and appears to generate stable RNA. Because the 5' UTR and start site of endogenous *fife* are preserved, the ectopic splice isoform could be translated into a truncated protein containing only the C2A and C2B domains. Interestingly, the fact that *fife*^{ex1027} is a genetic null indicates that any truncated protein produced is non-functional, possibly for failing to localize or fold properly.

As a first step towards understanding the in vivo role of Fife, we assessed synaptic growth in *fife* mutants. NMJs of third instar larvae were labeled with HRP, which stains all neuronal membranes, and total synaptic area determined (Fig. 4). We observed no difference in synaptic area between wild-type and *fife* mutant larvae (wild type: 243.8 μm^2 vs. *fife*^{ex1027}*/Df*: 232.3 μm^2; $p > 0.05$). Thus, overall synaptic growth proceeds normally in the absence of Fife, indicating that Fife either acts redundantly with other CAZ proteins to coordinate synaptic growth or has distinct functions downstream of the initial formation of synaptic connections.

As a scaffolding protein present at AZs, Fife is well situated to participate in the assembly of presynaptic terminals during synapse development. To assess the assembly of AZs in the absence of Fife, we labeled wild-type and *fife* mutant NMJs with antibodies to the CAZ protein Bruchpilot. Bruchpilot is a late-arriving molecular component of AZs whose

localization depends on other CAZ proteins and thus a good readout of AZ assembly (Fouquet et al., 2009; Graf et al., 2009; Owald et al., 2010). We measured AZ number by counting Bruchpilot puncta and found no significant difference between wild-type and *fife* mutants (Fig. 5A-C; wild type: 278 ± 8 AZs/NMJ vs. $fife^{ex1027}/Df$: 338 ± 16; $p > 0.05$). To assess the molecular composition of *fife* AZs, we measured Bruchpilot intensity relative to HRP levels at wild-type and *fife* AZs. While we noted a tendency towards lower Bruchpilot levels, no significant difference was observed (Fig. 5A-B,D; wild type: .62 ± 0.05 a.u. vs. $fife^{ex1027}/Df$: .46 ± 0.09 a. u.; $p > 0.05$). Thus, as assessed by the accumulation of Bruchpilot at the level of light microscopy, we find that the assembly of presynaptic terminals is not significantly impaired in *fife* mutants.

A critical aspect of synapse formation is the accumulation of PSD proteins in direct apposition to AZs to facilitate the rapid transfer of information. At the Drosophila NMJ, glutamate receptors (GluRs) comprising four heteromeric subunits cluster at PSDs. By co-labeling NMJs for Bruchpilot and GluRIIC, an obligate GluR subunit also known as GluRIII, we found that this process is grossly normal in *fife* mutants (Fig. 5A, B). However, we do observe some disorganization, including unapposed PSDs (Fig. 5B, arrow), unapposed AZs (filled arrowhead), and a failure of 1:1 apposition between AZs and PSDs (open arrowhead). It is notable that these mild abnormalities are diverse, which may explain the failure to identify significant differences by any single measure.

To assess PSD development, we examined the number and relative intensity of GluRIIC puncta and, as with AZ number, observed no significant difference between wild-type and *fife* mutant synapses. Wild-type NMJs contain 267 ± 9 PSDs, while $fife^{ex1027}/Df$ NMJs have 293 ± 10 (Figure 5E; $p > 0.05$). We again observe an insignificant tendency towards reduced receptor levels; however, the intensity of GluRIIC at *fife* PSDs is not

significantly different than at wild-type PSDs (wild type: .62 ± 0.06 a.u. vs. *fife*^{ex1027}*/Df*: .43 ± 0.08 a.u.; $p > 0.05$), indicating that receptor clusters are able to accumulate in the absence of Fife (Fig. 5A-B,F) Together our morphological data indicate that, at the level of resolution afforded by light microscopy, synapse assembly is not significantly impaired in the absence of Fife and hint at more subtle deficits below the resolution of these assays.

Fife regulates active zone architecture

Because the size of critical AZ components falls well below the 200-nm resolution limit of light microscopy, we turned to electron microscopy to investigate synaptic architecture in greater detail. Consistent with our light-level analysis, average AZ length is normal in *fife* mutants (483 nm ± 29.1 nm in wild type vs. 490 nm ± 35.1 nm in *fife* mutants; $p > 0.05$). However, we immediately noticed multiple abnormalities at *fife* AZs. First, we occasionally observe non-membrane-associated electron-dense structures of variable morphology that are capable of clustering synaptic vesicles away from AZs (Fig. 6A-C and E-G; wild type: 0%, n=20 vs. *fife*^{ex1027}*/Df* : 12.1%, n = 33). Second, at nearly all *fife* AZs we observe detachments of the presynaptic membrane, or membrane ruffling (Fig. 6 D,H; wild type: 0%, n=13 vs. *fife*^{ex1027}*/Df* : 91.7%, n = 12). On average *fife* AZs with membrane defects are ruffled along 44% of their length (44.1 ± 5.3%, n=11), indicating severely compromised AZs in which large portions of the synaptic membrane are disrupted. Together, these findings reveal a critical role for Fife in determining the precise architecture of AZs.

Fife promotes synaptic vesicle clustering

A fundamental property of presynaptic organization and synaptic function is the clustering of synaptic vesicles in close proximity to the presynaptic membrane and, in

many cases, at electron densities that extend from the membrane into the cytoplasm. CAZ proteins are critical regulators of this process, so we examined a possible role for Fife in the organization of synaptic vesicle pools by quantifying vesicles within 200 nm of the plasma membrane at wild-type and *fife* mutant AZs. We found a decrease of approximately 20% in vesicle number at *fife* AZs (Fig. 7; wild type: 40.4 ± 2.9 synaptic vesicles vs. $fife^{ex1027}/Df$: 31.8 ± 1.9; $p < 0.02$). In contrast, total vesicle density is not significantly altered in *fife* boutons (wild type: 152.9 ± 10.34 vesicles/μm^2 vs. $fife^{ex1027}/Df$: 187.7 ± 15.92; $p > 0.05$). Thus, in addition to its role in regulating synapse architecture, Fife promotes synaptic vesicle clustering at AZs.

Neurotransmitter release is severely reduced in *fife* mutants

The abnormalities in synaptic architecture and vesicle clustering observed at *fife* NMJs suggest neurotransmission may be impaired in the absence of *fife*. To assess the role of Fife in synaptic transmission, we measured spontaneous and evoked synaptic potentials through intracellular recordings at *fife* NMJs, and observed significant deficits in exocytosis (Fig. 8A-E). Miniature excitatory junction potential (mEJP) frequency and amplitude are not significantly affected by loss of *fife* (wild type: 3.7 ± 0.4 Hz vs. $fife^{ex1027}/Df$: 3.4 ± 0.5 Hz; $p > 0.05$ and wild type: 1.07 ± 0.06 mV vs. $fife^{ex1027}/Df$: 0.89 ± 0.09 mV; $p > 0.05$, respectively). However, evoked excitatory junction potentials (EJPs) are decreased by approximately 65% (Fig. 8A,B,D; wild type: 27.4 ± 1.6 mV vs. $fife^{ex1027}/Df$: 9.2 ± 1.5 mV; $p < 0.0001$). Decreased evoked potentials were rescued by neuronal expression of full-length *fife* (Fig. 8C-E) indicating a presynaptic role for Fife consistent with its cellular and subcellular expression (Rescue: 20.6 ± 1.7 mV vs. $fife^{ex1027}/Df$: 9.2 ± 1.5 mV; $p < 0.001$). We calculated quantal content (EJP amplitude divided by mEJP amplitude) to estimate the number of vesicles released per action potential and found a 60% decrease at *fife* NMJs (Fig. 8E; wild type: 27.0 ± 2.0 vesicles

vs. *fife*^{ex1027}*/Df*: 10.6 ± 1.7 vesicles; $p < 0.0001$). Together, these results reveal a critical requirement for Fife in evoked neurotransmitter release consistent with the defects we observe in the organization of presynaptic terminals of *fife* mutants.

Loss of Fife disrupts motor behavior

A 60% decrease in neurotransmitter release and severe ultrastructural abnormalities would be expected to have behavioral consequences for the fly. Indeed, *fife* mutant flies have multiple behavioral deficits. Homozygous *fife* ^{ex1027} and *fife*^{ex1027}*/Df* females produce, but fail to lay, eggs. This may reflect deficits in neurotransmission at motorneurons or a disruption of the complex circuitry that links mating to egg laying upstream of motor function. *fife* mutants also exhibit locomotor defects, moving only half as much as wild-type flies (Fig. 8F). In line-crossing assays, we find that wild-type flies cross an average of 17.6 ± 2.5 lines per 30-second trial, while *fife*^{ex1027}*/Df* flies only cross 9.1 ± 0.6 lines ($p < 0.05$). Neuronal expression of *fife* fully rescues the locomotion deficit (*C155-Gal4>UAS-fife*: 19.5 ± 2.3 line crosses/trial; $p < 0.01$). Thus, Fife functions at presynaptic terminals to organize AZs, support evoked neurotransmitter release, and generate normal motor behavior.

Discussion

The CAZ is a conserved molecular machine responsible for coordinating the structure and function of presynaptic terminals. With the notable exceptions of Piccolo and Bassoon, representatives of all major vertebrate CAZ protein families have been found in invertebrates. Here we present the identification of a previously overlooked invertebrate Piccolo-RIM homolog. We uncover a requirement for the Drosophila homolog, Fife, in locomotor behavior and neurotransmitter release, and define a role for Fife in AZ organization and synaptic vesicle clustering at the NMJ.

Conservation of the CAZ

Is Fife a Piccolo ortholog or a second RIM ortholog? Our phylogenetic analyses support the former model. However, Fife, which exhibits significant sequence homology to both Piccolo and RIM proteins, is structurally more similar to RIMs. Because molecular phylogenetic approaches rely on conserved sequences, the non-conserved CC-regions are not considered in these analyses and, thus, this finding is not conclusive. Determining orthology based on function is likely to prove similarly complicated in this case because CAZ proteins share many physical interactions and often function redundantly in carrying out core CAZ functions. Future studies of protein-protein interactions may prove informative. Previous work in yeast has demonstrated that combining sequence-level analysis with protein-protein interaction data improves the ability to identify functional orthologs (Bandyopadhyay et al., 2006). While RIMs and Piccolo have many binding partners in common, they also have unique associations (Schoch and Gundelfinger, 2006). For example, RIM proteins bind synaptic vesicle protein Rab3, while Piccolo associates with the Rab3-interacting protein PRA1 (Wang et al., 1997; Fenster et al., 2000). Determining which Piccolo/RIM associations are conserved in Fife

will be an important step towards understanding the conservation of CAZ function in general and Fife function specifically.

As noted, our molecular and computational searches in Drosophila have not identified sequences that encode the large CC-domain-containing region of Piccolo – either at the *fife* locus or elsewhere in the genome – suggesting that this sequence is specific to vertebrates or was lost in the insect lineage. Interestingly, while insect Fife proteins may have lost their CC domains, the opposite is true of the Drosophila CAST homolog Bruchpilot, which colocalizes with Fife at AZs and has a much larger CC-containing region than its vertebrate counterparts (Wagh et al., 2006). These observations raise the intriguing possibility that the expansion of Bruchpilot's CC-regions may have reduced selective pressure to maintain these regions in Fife, and that evolutionary conservation may be found at the level of the intertwined complex of CAZ proteins as well as at the level of individual proteins.

The CC domains of Piccolo mediate physical interactions with CAZ proteins CAST and UNC-13. However, it is interesting to note that both of these physical interactions are maintained in RIM proteins despite the lack of CC domains as RIM binds CAST through its PDZ domain and UNC-13 through its ZF domain. Future studies will be important for determining if physical interactions mediated by the CC domains of vertebrate Piccolo are carried out by other domains in Fife or by other CAZ proteins such as Bruchpilot.

The CC-domain containing region of Piccolo gives the protein its large size. If Piccolo's 5100 amino acids were arranged in an extended α–helical conformation it could in principle span 750 nm (Limbach et al., 2011). It has long been speculated that this potential, also present in Bassoon, has important implications for Piccolo-Bassoon

function. However, in a recent immuno-electron microscopy study, Limbach and colleagues determined the localization of 11 Piccolo epitopes spanning the length of the protein to understand its spatial organization at AZs. They determined that Piccolo adopts a compact conformation with its N-terminus localizing near dense projections approximately 75 nm from the plasma membrane and its C-terminus localizing within 30-40 nm of the plasma membrane in close proximity to binding partners RIM and UNC-13. This finding is supported by a study employing stochastic optical reconstruction microscopy (STORM), and argues that the functions of Piccolo are not dependent upon its large size but rather, like RIMs, on its many interactions with other AZ proteins – a role that is likely maintained in Fife (Dani et al., 2010).

CAZ regulation of synaptic architecture and function

To investigate Fife function at Drosophila synapses, we generated deletion alleles. Here we have focused on *fife*^{ex1027}, a deletion mutant that behaves as a genetic null in that its phenotype over a chromosomal deficiency mirrors its homozygous phenotype in morphological, functional and behavioral assays. Through phenotypic analysis of *fife*, we found loss of Fife results in a 50% reduction in locomotion and 60% decrease in evoked neurotransmitter release. To uncover the underlying cause of these deficits, we conducted a morphological analysis of *fife* synapses that revealed crucial roles for Fife in the regulation of AZ structure. Absolute number and gross morphology of synapses is largely normal in *fife* mutants. This is consistent with functional studies of Piccolo and RIM proteins demonstrating normal formation of synapses and the accumulation of wild-type levels of synaptic proteins in their absence (Koushika et al., 2001; Schoch et al., 2006; Leal-Ortiz et al., 2008; Mukherjee et al., 2010). However, *fife* AZs often exhibit separations of pre- and post-synaptic membranes as revealed by ruffling of the presynaptic membrane in electron micrographs. As we observe this defect along much of

the length of affected AZs, it likely disrupts presynaptic release sites and thus contributes to the deficit in exocytosis observed at *fife* NMJs. This phenotype has previously been linked to decreased cell adhesion, raising the possibility that Fife might regulate the levels or localization of transsynaptic cell adhesion molecules at synapses (Banovic et al., 2010; Mosca et al., 2012). Less frequently we observed floating electron densities, a phenotype highly reminiscent of the floating ribbons observed at *bassoon* mutant ribbon synapses in the vertebrate retina (Dick et al., 2003). At ribbon synapses, electron dense ribbon-shaped structures extend from the AZ membrane into the presynaptic cytoplasm and, like T bars, cluster synaptic vesicles (Zhai and Bellen, 2004). In Drosophila, floating electron densities have also been observed at low frequency in the absence of Rim-Binding Protein, a CAZ-associated protein with an essential role in neurotransmitter release (Liu et al., 2011). It will be important to determine whether Fife functions with Rim-Binding Protein and/or other CAZ proteins to control the structural integrity of AZs.

We also identified a role for Fife in clustering synaptic vesicles. At *fife* synapses, 20% fewer synaptic vesicles reside within 200 nm of the AZ membrane than at control synapses. In their study of *piccolo* knockout mice, Mukherjee et al. (2010) identified a similar deficit in synaptic vesicle clustering at cortical synapses lacking both Piccolo and Bassoon, as did a recent study of Bassoon function at mouse hair cell ribbon synapses (Frank et al., 2010). It will be of interest to determine if this deficit at *fife* AZs includes a decrease in the number of docked (membrane-bound) vesicles as has been observed in *rim* and *bassoon* mutants and in *piccolo, bassoon* double knockdown neurons (Weimer et al., 2006; Frank et al., 2010; Mukherjee et al., 2010). We quantified docked vesicles in our EM samples, which were prepared by conventional aldehyde fixation methods, and observed no significant difference between wild type and *fife* mutants (JB, RG and KOCG, unpublished data). However, recent studies in mice and *C. elegans* have made it clear

that aldehyde fixation masks vesicle-docking defects (Siksou et al., 2007; Gundelfinger and Fejtova, 2009; Gracheva et al., 2010). Thus, future studies employing high pressure freeze fixation methods will be required to determine whether Fife functions in the docking of vesicles for rapid release.

The number and nature of vertebrate Piccolo/RIM-binding partners hint at additional roles for Fife at AZs. However, it is currently unknown if many of these vertebrate interactions are biologically relevant. With the identification of a Piccolo-related protein in invertebrates and the generation of *fife* alleles in Drosophila, this question can now be readily addressed through double mutant analysis. Studies of CAZ components in Drosophila and *C. elegans* have been critical to our understanding of AZ organization and function (Owald and Sigrist, 2009). This study reveals greater conservation of CAZ proteins than previously believed and opens the door to a more complete understanding of CAZ function through in vivo studies of this fundamental molecular machine and its core components in an amenable genetic model system.

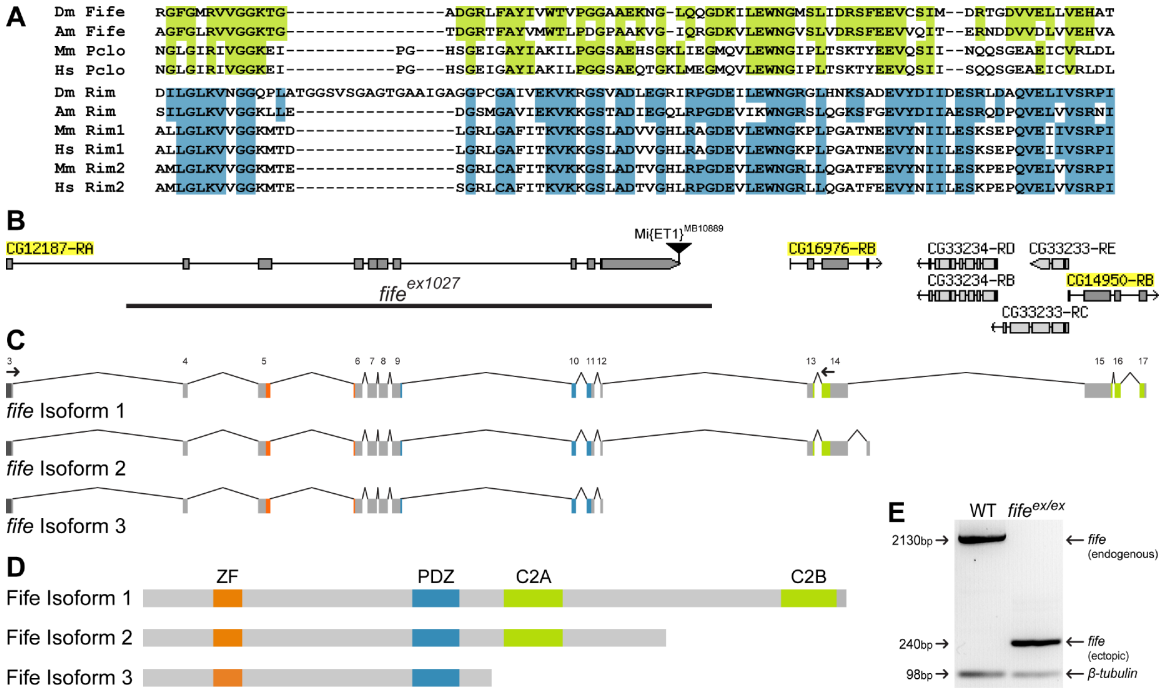

Figure 1. Molecular identification of a *piccolo-rim*-related gene in Drosophila. ***A,*** PDZ domain of Fife/Piccolo (top) and RIM (bottom) family proteins are aligned to illustrate conservation among the families. Highlighted regions indicate identity between the Drosophila protein and vertebrate Piccolo (green) or RIM (blue). (Am) *Apis mellifera*, (Dm) *Drosophila melanogaster*, (Hs) *Homo sapiens*, (Mm) *Mus musculus*. ***B,*** Full-length *fife* comprises 17 exons corresponding to a predicted open reading frame previously annotated as 3 separate candidate genes, CG12187, CG16976, and CG14950. Two 5' UTR exons, located 10 and 20 kb upstream, are not depicted. $fife^{ex1027}$ was generated by excision of the Minos element $Mi\{ET1\}CG43375^{MB10889}$. The location of the 16.5 deletion is indicated (black bar). ***C,*** *fife* encodes 3 transcripts. Exons are numbered above isoform 1. ***D,*** Full-length Fife (Isoform 1) has four conserved functional domains: ZF (orange), PDZ (blue), C2A (green), and C2B (green). Alternative isoforms lack either the C2B domain (Isoform 2) or both C2 domains (Isoform 3). ***E,*** Real-time PCR products from amplification of exons 3-14 in wild type and $fife^{ex1027}$. Primer locations are indicated by small arrows in panel C. A full-length *fife* product of 2130 base pairs is amplified from wild-type, but not $fife^{ex1027}$, cDNA. Conversely, a 240 base-pair band representing an aberrant splicing event between exons 3 and 13 is present in $fife^{ex1027}$ but not in wild type. *β-tubulin* was also amplified in these reactions as a loading control. Image ***B*** is modified from FlyBase (McQuilton et al., 2012).

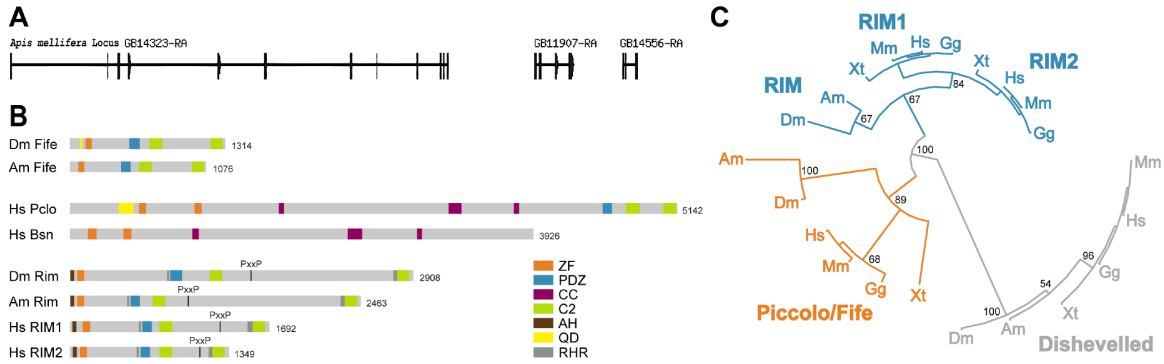

Figure 2. Fife may encode an invertebrate Piccolo ortholog. *A,* Gene order of the three annotated genes that comprise *fife* is conserved in *Apis mellifera* and conceptual translation of the three genes produces a protein comprising ZF, PDZ, C2A, and C2B domains. ***B,*** Domain structures of related Fife, Piccolo, Bassoon (Bsn) and RIM proteins in insects and humans drawn to scale. ZF (orange), PDZ (blue), CC (purple), C2 (green), α-helix (AH, brown), and glutamine-rich (QD, yellow) domains are represented. Inter-domain homology regions, which are conserved in RIM family proteins but not found in Fife or Piccolo, are indicated in dark grey. ***C,*** Circle tree indicating the most likely evolutionary relationships of insect and vertebrate Piccolo, RIM and Fife proteins. Phylogeny was determined by maximum likelihood from alignments of the well-conserved PDZ domains of each protein. The PDZ-domains of the Dishevelled protein family were used as an outgroup. Confidence was assessed by the bootstrapping method with 500 replications. Branches with > 50% support are shown. (Am) *Apis mellifera*, (Dm) *Drosophila melanogaster*, (Hs) *Homo sapiens*, (Gg) *Gallus gallus*, (Mm) *Mus musculus*, and (Xt) *Xenopus tropicalis*.

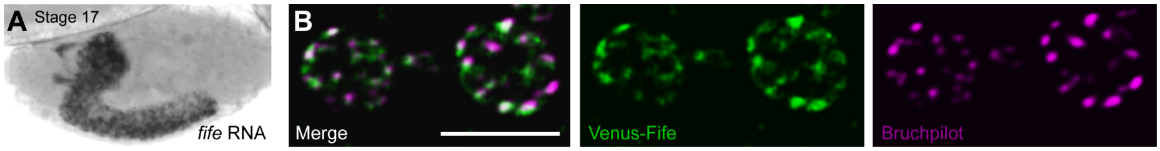

Figure 3. Fife is an active zone protein. ***A,*** In situ hybridization detects *fife* RNA in post-mitotic neurons of stage 17 embryos. ***B,*** Single optical section (0.1 μm) of *C155GAL4/Y; UAS-Venus-fife/+* NMJ boutons co-labeled with antibodies to GFP and Bruchpilot. Scale bar = 5 μm.

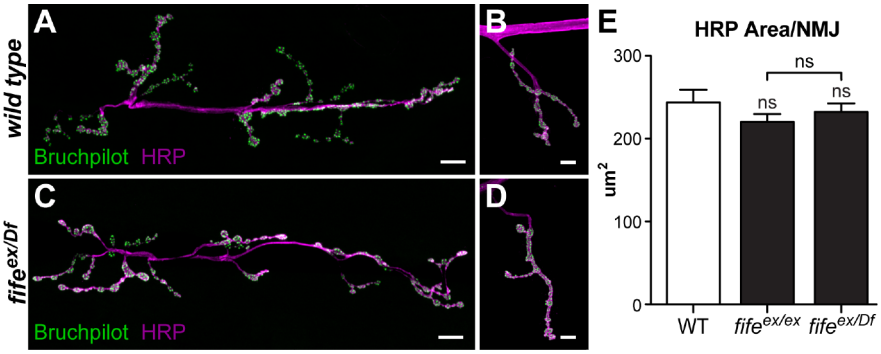

Figure 4. Synaptic growth is not altered in the absence of Fife. ***A-D,*** Confocal z-projections of NMJ 6/7 (***A, C***) and NMJ 4 (***B, D***) co-labeled with antibodies to Bruchpilot and HRP in wild-type (***A, B)*** and *fife*$^{ex/Df}$ *(fife*^{ex1027}*/Df (3L) BSC412)* (***C, D)*** larvae. Quantification of the area of HRP signal at NMJ 4 of segment A3 shows no significant difference between wild type (n = 14 NMJs) and *fife*$^{ex/ex}$ (n = 19 NMJs) or *fife*$^{ex/Df}$ (n = 17 NMJs). Scale bar = 10 μm.

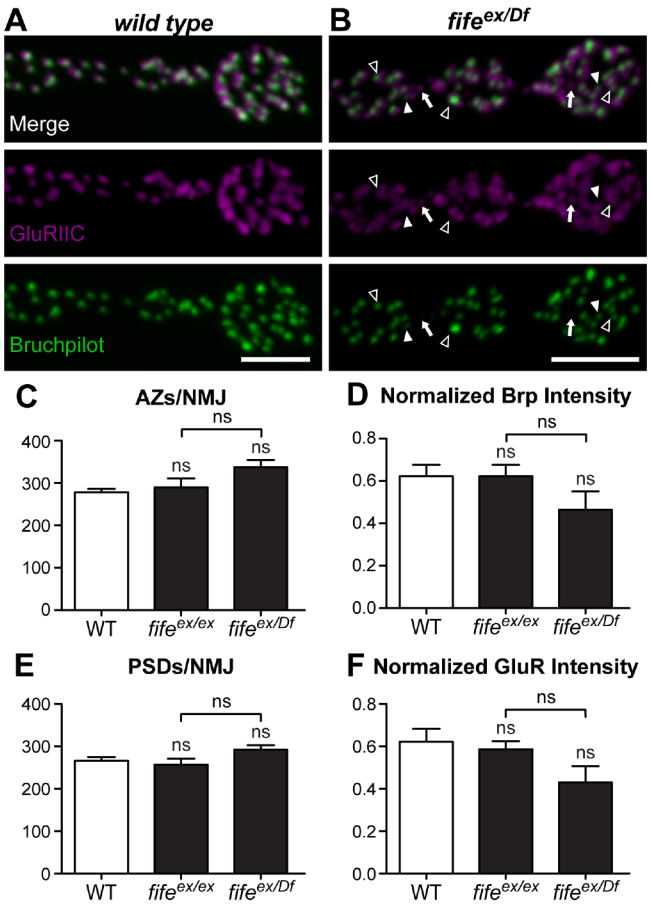

Figure 5. Synapse assembly is grossly normal in *fife* mutants. ***A, B,*** Confocal z-projections of boutons co-labeled with antibodies to GluRIIC and Bruchpilot at wild-type (***A***) and *fife*$^{ex/Df}$ *(fife*^{ex1027}*/Df (3L) BSC412)* (***B***) NMJs. Mild disruptions of AZ architecture observed at *fife*$^{ex/Df}$ NMJs include unopposed GluRIIC fields (arrows), unopposed Bruchpilot puncta (filled arrowhead), and failure of 1:1 apposition of GluRIIC and Bruchpilot clusters (open arrowhead). ***C,*** The number of Bruchpilot puncta does not differ between wild-type (n = 8) and *fife*$^{ex/ex}$ (n = 9) or *fife*$^{ex/Df}$ (n = 9) NMJs. ***D,*** The intensity of Bruchpilot puncta normalized to HRP is not significantly different at wild-type (n = 2224) and *fife*$^{ex/ex}$ (n = 2609) or *fife*$^{ex/Df}$ (n = 3041) AZs. ***E,*** The number of GluRIIC puncta is not altered between wild-type (n = 8) and *fife*$^{ex/ex}$ (n = 9) or *fife*$^{ex/Df}$ (n = 9) NMJs. ***F,*** The intensity of GluRIIC puncta normalized to HRP is not significantly different at wild-type (n = 2132) and *fife*$^{ex/ex}$ (n = 2315) or *fife*$^{ex/Df}$ (n = 2634) AZs. Scale bar = 5 μm.

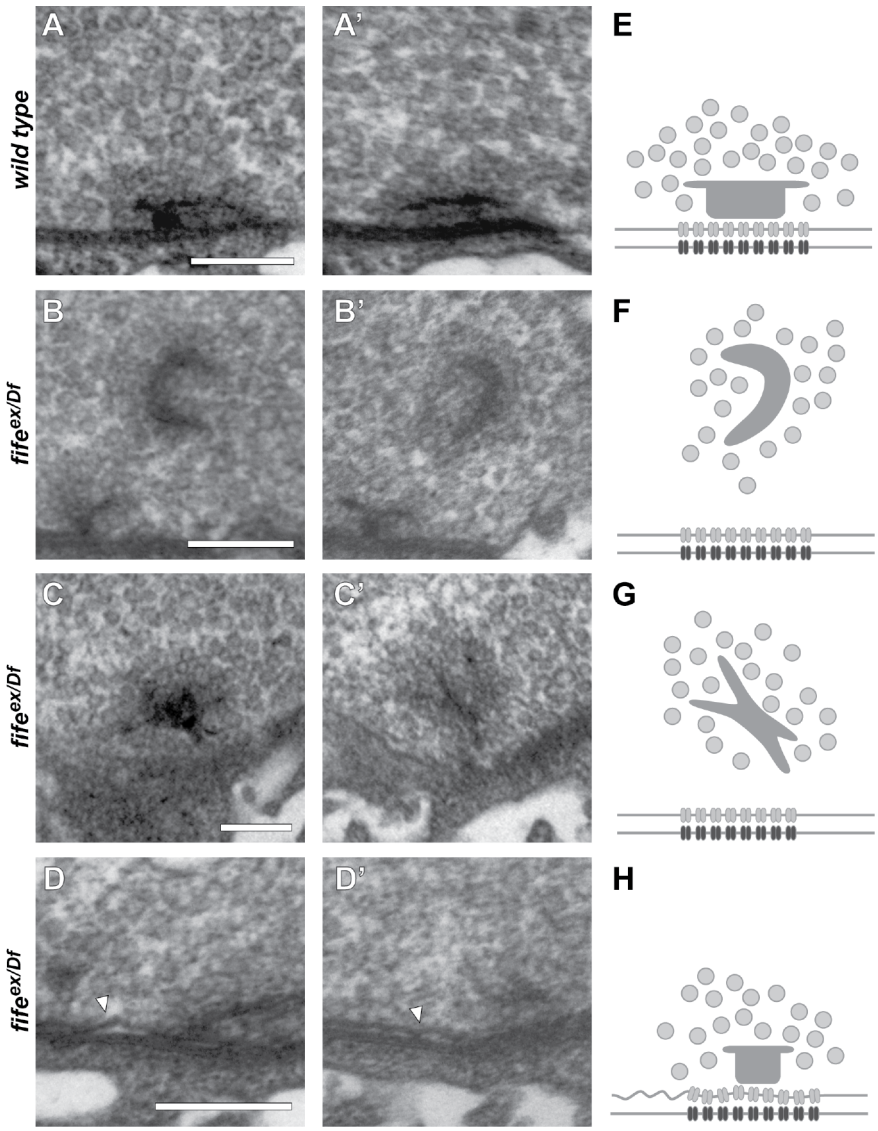

Figure 6. Fife is required for normal active zone architecture. *A-D',* Representative transmission electron micrographs of serial sections of wild-type (***A, A'***) and *fife*$^{ex/Df}$ *(fife*^{ex1027}*/Df (3L) BSC412)* (***B-D'***) synapses. ***A, A',*** Two serial sections of a wild type synapse reveal a platform-shaped electron dense projection (T-bar) that clusters synaptic vesicles at the presynaptic membrane. Tight apposition of the pre- and postsynaptic is maintained across the length of synapse. ***B, B',*** Two serial sections of a *fife*$^{ex/Df}$ synapse reveal an ectopic, electron dense structure distant from the presynaptic membrane. This T-bar-like malformation still maintains local clustering of synaptic vesicles. ***C, C',*** A second example of an ectopic electron density in two serial sections of a *fife*$^{ex/Df}$ synapse. ***D, D',*** Two serial sections of a typical *fife*$^{ex/Df}$ synapse demonstrate loss of the close apposition of pre- and postsynaptic membranes in *fife* mutants (arrowhead). ***E-H,*** Schematic representations of the AZs and ectopic densities depicted in A-D' including synaptic vesicles (light grey circles), T-bars (dark grey structures), calcium channels (grey) and postsynaptic receptors (black). Scale bar = 200 nm.

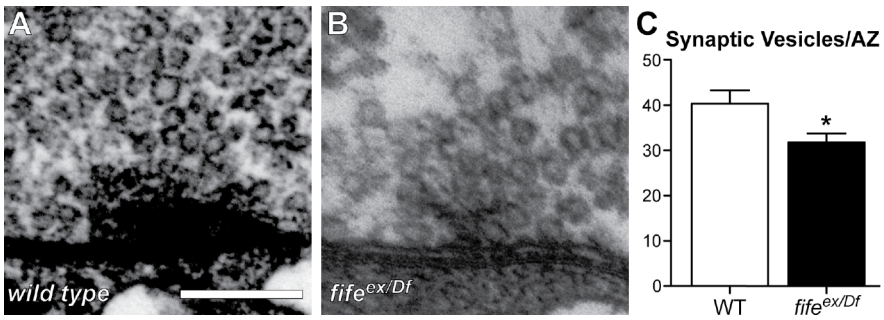

Figure 7. Fife promotes synaptic vesicle clustering. ***A, B,*** Representative transmission electron micrographs of wild-type (***A***) and *fife*$^{ex/Df}$ *(fife*^{ex1027}*/Df (3L) BSC412)* (***B***) synapses ***C,*** Quantification of the number of synaptic vesicles clustered at AZs reveals a significant reduction in the number of vesicles within 200 nm of the AZ in *fife*$^{ex/Df}$ (n = 25 AZs) compared to wild type (n = 23 AZs). Scale bar = 200 nm.

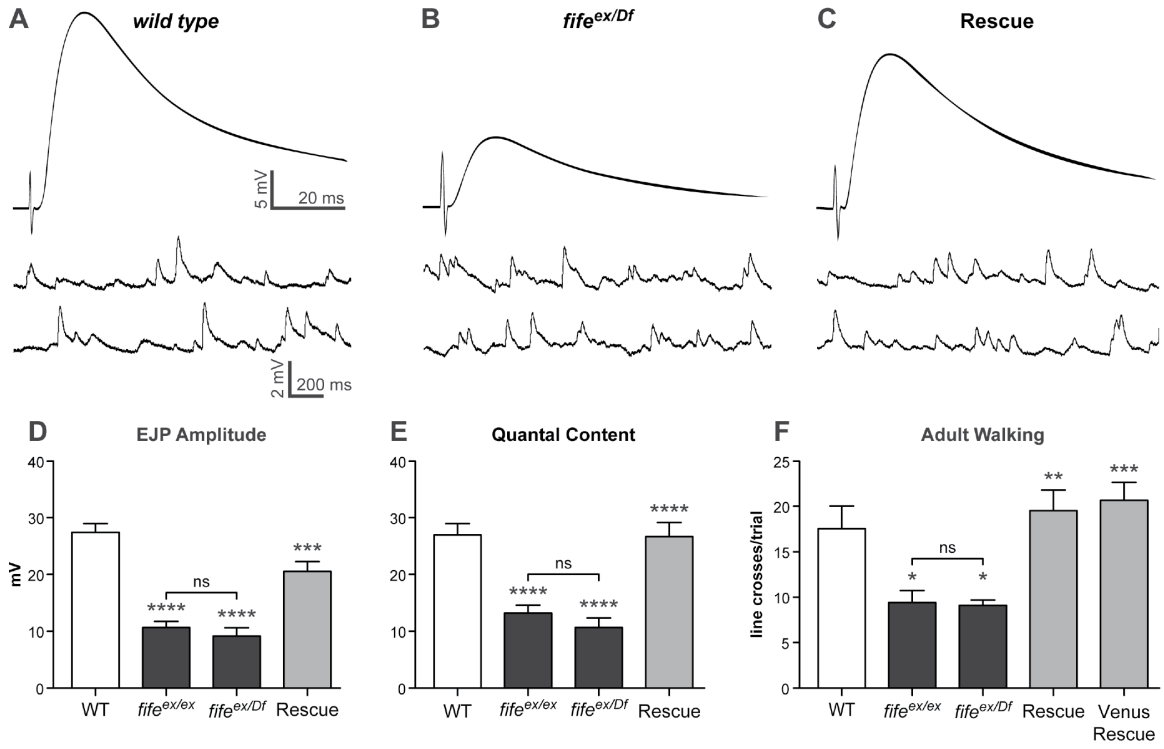

Figure 8. Neurotransmitter release is severely reduced in *fife* mutants. *A-C,* Representative traces of EJPs and mEJPs from wild-type (***A***) *fife*$^{ex/Df}$ (*fife*^{ex1027}*/Df (3L) BSC412)* (***B***), and Rescue (*C155-Gal4/Y; UAS-fife, Df (3L) BSC412/fife*^{ex1027}) (***C***) NMJs. Recordings were obtained from muscle 6 of abdominal segment 3 at 0.5mM extracellular Ca^{2+}. ***D,*** Mean EJP amplitude is decreased significantly in *fife*$^{ex/ex}$, n = 11 muscles and *fife*$^{ex/Df}$ (n = 9 muscles) compared to wild type (n = 19 muscles). EJP amplitude is rescued by neuronal *fife* expression (Rescue, n = 9 muscles). ***E,*** Mean quantal content is significantly reduced in *fife*$^{ex/ex}$ and *fife*$^{ex/Df}$ compared to wild type and rescued by neuronal *fife* expression (Rescue). ***F,*** Locomotion is severely impaired in both *fife*$^{ex/ex}$ and *fife*$^{ex/Df}$ flies compared to wild-type flies. This motor deficit is rescued in both Rescue and Venus Rescue (*C155-Gal4/Y; UAS-venus-fife, Df (3L) BSC412/fife*^{ex1027}) flies.

Table 1. Identity and similarity of Fife to human Piccolo and RIM proteins.

	Hs Piccolo			Hs RIM1			Hs RIM2			Dm Fife		
	PDZ	C2A	C2B	PDZ	C2A	C2B	PDZ	C2A	C2B	PDZ	C2A	C2B
Dm Fife												
Identity	43	25	31	39	34	30	41	34	35	100	100	100
Similarity	64	51	52	57	50	55	55	50	54	100	100	100
Dm RIM												
Identity	35	39	33	57	53	65	53	53	66	40	33	32
Similarity	58	56	54	77	69	79	70	70	79	54	51	49

Numbers indicate the percentage of identical (top) and similar (bottom) amino acid residues from alignments of the PDZ, C2A and C2B domains of Fife, Piccolo and RIM proteins. (Dm) *Drosophila melanogaster*, (Hs) *Homo sapiens*.

References

Aravamudan B, Fergestad T, Davis WS, Rodesch CK, Broadie K (1999) Drosophila UNC-13 is essential for synaptic transmission. Nat Neurosci 2:965-971.

Bandyopadhyay S, Sharan R, Ideker T (2006) Systematic identification of functional orthologs based on protein network comparison. Genome Res 16:428-435.

Banovic D, Khorramshahi O, Owald D, Wichmann C, Riedt T, Fouquet W, Tian R, Sigrist SJ, Aberle H (2010) Drosophila neuroligin 1 promotes growth and postsynaptic differentiation at glutamatergic neuromuscular junctions. Neuron 66:724-738.

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.

Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM, Micklem G, Piano F, Snyder M, Stein L, White KP, Waterston RH (2009) Unlocking the secrets of the genome. Nature 459:927-930.

Dani A, Huang B, Bergan J, Dulac C, Zhuang X (2010) Superresolution imaging of chemical synapses in the brain. Neuron 68:843-856.

Deng L, Kaeser PS, Xu W, Sudhof TC (2011) RIM proteins activate vesicle priming by reversing autoinhibitory homodimerization of Munc13. Neuron 69:317-331.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465-469.

Dick O, tom Dieck S, Altrock WD, Ammermuller J, Weiler R, Garner CC, Gundelfinger ED, Brandstatter JH (2003) The presynaptic active zone protein bassoon is essential for photoreceptor ribbon synapse formation in the retina. Neuron 37:775-786.

Drysdale R (2008) FlyBase : a database for the Drosophila research community. Methods Mol Biol 420:45-59.

Fenster SD, Chung WJ, Zhai R, Cases-Langhoff C, Voss B, Garner AM, Kaempf U, Kindler S, Gundelfinger ED, Garner CC (2000) Piccolo, a presynaptic zinc finger protein structurally related to bassoon. Neuron 25:203-214.

Fouquet W, Owald D, Wichmann C, Mertel S, Depner H, Dyba M, Hallermann S, Kittel RJ, Eimer S, Sigrist SJ (2009) Maturation of active zone assembly by Drosophila Bruchpilot. J Cell Biol 186:129-145.

Frank T, Rutherford MA, Strenzke N, Neef A, Pangrsic T, Khimich D, Fejtova A, Gundelfinger ED, Liberman MC, Harke B, Bryan KE, Lee A, Egner A, Riedel D, Moser T (2010) Bassoon and the synaptic ribbon organize Ca(2)+ channels and vesicles to add release sites and promote refilling. Neuron 68:724-738.

Gracheva EO, Hadwiger G, Nonet ML, Richmond JE (2008) Direct interactions between C. elegans RAB-3 and Rim provide a mechanism to target vesicles to the presynaptic density. Neurosci Lett 444:137-142.

Gracheva EO, Maryon EB, Berthelot-Grosjean M, Richmond JE (2010) Differential Regulation of Synaptic Vesicle Tethering and Docking by UNC-18 and TOM-1. Front Synaptic Neurosci 2:141.

Graf ER, Daniels RW, Burgess RW, Schwarz TL, DiAntonio A (2009) Rab3 dynamically controls protein composition at active zones. Neuron 64:663-677.

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557-559.

Gundelfinger ED, Fejtova A (2009) Neurotransmitter release: a docking role for UNC-13 proteins (Commentary on Siksou et al.). Eur J Neurosci 30:47-48.

Hallermann S, Fejtova A, Schmidt H, Weyhersmuller A, Silver RA, Gundelfinger ED, Eilers J (2010) Bassoon speeds vesicle reloading at a central excitatory synapse. Neuron 68:710-723.

Han Y, Kaeser PS, Sudhof TC, Schneggenburger R (2011) RIM determines Ca(2)+ channel density and vesicle docking at the presynaptic active zone. Neuron 69:304-316.

Hillis DM, Allard MW, Miyamoto MM (1993) Analysis of DNA sequence data: phylogenetic inference. Methods Enzymol 224:456-487.

Jin Y, Garner CC (2008) Molecular mechanisms of presynaptic differentiation. Annu Rev Cell Dev Biol 24:237-262.

Kaeser PS, Deng L, Wang Y, Dulubova I, Liu X, Rizo J, Sudhof TC (2011) RIM proteins tether Ca2+ channels to presynaptic active zones via a direct PDZ-domain interaction. Cell 144:282-295.

Koushika SP, Richmond JE, Hadwiger G, Weimer RM, Jorgensen EM, Nonet ML (2001) A post-docking role for active zone protein Rim. Nat Neurosci 4:997-1005.

Leal-Ortiz S, Waites CL, Terry-Lorenzo R, Zamorano P, Gundelfinger ED, Garner CC (2008) Piccolo modulation of Synapsin1a dynamics regulates synaptic vesicle exocytosis. J Cell Biol 181:831-846.

Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40:D302-305.

Limbach C, Laue MM, Wang X, Hu B, Thiede N, Hultqvist G, Kilimann MW (2011) Molecular in situ topology of Aczonin/Piccolo and associated proteins at the mammalian neurotransmitter release site. P Natl Acad Sci USA 108:E392-401.

Liu KS, Siebert M, Mertel S, Knoche E, Wegener S, Wichmann C, Matkovic T, Muhammad K, Depner H, Mettke C, Buckers J, Hell SW, Muller M, Davis GW,

Schmitz D, Sigrist SJ (2011) RIM-binding protein, a central part of the active zone, is essential for neurotransmitter release. Science 334:1565-1569.

Marrus SB, Portman SL, Allen MJ, Moffat KG, DiAntonio A (2004) Differential localization of glutamate receptor subunits at the Drosophila neuromuscular junction. J Neurosci 24:1406-1415.

McQuilton P, St Pierre SE, Thurmond J (2012) FlyBase 101--the basics of navigating FlyBase. Nucleic acids research 40:D706-714.

Metaxakis A, Oehler S, Klinakis A, Savakis C (2005) Minos as a genetic and genomic tool in Drosophila melanogaster. Genetics 171:571-581.

Mosca TJ, Hong W, Dani VS, Favaloro V, Luo L (2012) Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. Nature 484:237-241.

Mukherjee K, Yang X, Gerber SH, Kwon HB, Ho A, Castillo PE, Liu X, Sudhof TC (2010) Piccolo and bassoon maintain synaptic vesicle clustering without directly participating in vesicle exocytosis. Proc Natl Acad Sci U S A 107:6504-6509.

Notredame C (2010) Computing multiple sequence/structure alignments with the T-coffee package. Curr Protoc Bioinformatics Chapter 3:Unit 3 8 1-25.

O'Connor-Giles KM, Ho LL, Ganetzky B (2008) Nervous wreck interacts with thickveins and the endocytic machinery to attenuate retrograde BMP signaling during synaptic growth. Neuron 58:507-518.

Owald D, Sigrist SJ (2009) Assembling the presynaptic active zone. Current opinion in neurobiology 19:311-318.

Owald D, Fouquet W, Schmidt M, Wichmann C, Mertel S, Depner H, Christiansen F, Zube C, Quentin C, Korner J, Urlaub H, Mechtler K, Sigrist SJ (2010) A Syd-1 homologue regulates pre- and postsynaptic maturation in Drosophila. J Cell Biol 188:565-579.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res 40:D290-301.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676-682.

Schoch S, Gundelfinger ED (2006) Molecular organization of the presynaptic active zone. Cell Tissue Res 326:379-391.

Schoch S, Mittelstaedt T, Kaeser PS, Padgett D, Feldmann N, Chevaleyre V, Castillo PE, Hammer RE, Han W, Schmitz F, Lin W, Sudhof TC (2006) Redundant functions of RIM1alpha and RIM2alpha in Ca(2+)-triggered neurotransmitter release. EMBO J 25:5852-5863.

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95:5857-5864.

Siksou L, Rostaing P, Lechaire JP, Boudier T, Ohtsuka T, Fejtova A, Kao HT, Greengard P, Gundelfinger ED, Triller A, Marty S (2007) Three-dimensional architecture of presynaptic terminal cytomatrix. J Neurosci 27:6868-6877.

Stigloher C, Zhan H, Zhen M, Richmond J, Bessereau JL (2011) The presynaptic dense projection of the Caenorhabditis elegans cholinergic neuromuscular junction localizes synaptic vesicles at the active zone through SYD-2/liprin and UNC-10/RIM-dependent interactions. J Neurosci 31:4388-4396.

tom Dieck S, Sanmarti-Vila L, Langnaese K, Richter K, Kindler S, Soyke A, Wex H, Smalla KH, Kampf U, Franzer JT, Stumm M, Garner CC, Gundelfinger ED (1998) Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively localized at the active zone of presynaptic nerve terminals. J Cell Biol 142:499-509.

Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R, Zhang H (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res 37:D555-559.

Wagh DA, Rasse TM, Asan E, Hofbauer A, Schwenkert I, Durrbeck H, Buchner S, Dabauvalle MC, Schmidt M, Qin G, Wichmann C, Kittel R, Sigrist SJ, Buchner E (2006) Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in Drosophila. Neuron 49:833-844.

Waites CL, Leal-Ortiz SA, Andlauer TF, Sigrist SJ, Garner CC (2011) Piccolo regulates the dynamic assembly of presynaptic F-actin. J Neurosci 31:14250-14263.

Wang JW, Beck ES, McCabe BD (2012) A modular toolset for recombination transgenesis and neurogenetic analysis of Drosophila. PLoS One 7:e42102.

Wang X, Kibschull M, Laue MM, Lichte B, Petrasch-Parwez E, Kilimann MW (1999) Aczonin, a 550-kD putative scaffolding protein of presynaptic active zones, shares homology regions with Rim and Bassoon and binds profilin. J Cell Biol 147:151-162.

Wang Y, Sudhof TC (2003) Genomic definition of RIM proteins: evolutionary amplification of a family of synaptic regulatory proteins(small star, filled). Genomics 81:126-137.

Wang Y, Okamoto M, Schmitz F, Hofmann K, Sudhof TC (1997) Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion. Nature 388:593-598.

Weimer RM, Gracheva EO, Meyrignac O, Miller KG, Richmond JE, Bessereau JL (2006) UNC-13 and UNC-10/rim localize synaptic vesicles to specific membrane domains. J Neurosci 26:8040-8047.

Zelwer M, Daubin V (2004) Detecting phylogenetic incongruence using BIONJ: an improvement of the ILD test. Mol Phylogenet Evol 33:687-693.

Zhai RG, Bellen HJ (2004) The architecture of the active zone in the presynaptic nerve terminal. Physiology (Bethesda) 19:262-270.

Chapter 2

Fife regulates the readily releasable vesicle pool size and vesicle-calcium channel coupling to promote neurotransmission

Joseph J. Bruckner, Scott J. Gratz, Hong Zhan, Xiao Rao, Gregory Zilberg, Zhongmin Lu, Gregory T. Macleod and Kate M. O'Connor-Giles

This work is in preparation to be submitted for publication.

Abstract

A conserved active zone cytomatrix is thought to organize presynaptic terminals for efficient synaptic transmission. We previously identified *Drosophila* Piccolo-related Fife as an integral component of the active zone cytomatrix that regulates neurotransmission and motor behavior. Here, we present a comprehensive mechanistic study of Fife's role in promoting neurotransmission. Paired-pulse recordings and functional studies in the presence of the slow Ca^{2+} buffer EGTA suggest Fife promotes high-probability release by positionally coupling synaptic vesicles and voltage-dependent calcium channels. Cumulative quantal analysis and high-pressure freeze/freeze substitution electron microscopy demonstrate that Fife determines the number of release-ready synaptic vesicles. Through tomographic reconstruction of active zones and superresolution imaging, we find that Fife determines the ultrastructural and molecular organization of the active zone cytomatrix. Based on these findings, we propose that Fife promotes synaptic transmission by organizing the active zone cytomatrix to create synaptic vesicle release sites within nanometer distance of calcium channel clusters.

Introduction

Neurotransmitter release occurs at specialized domains of presynaptic terminals called active zones. At active zones, synaptic vesicles dock with the presynaptic membrane where they respond to Ca^{2+} influx by rapidly fusing with the membrane and releasing their contents into the synaptic cleft. A highly conserved complex of active zone-associated proteins makes up the active zone cytomatrix, and is thought to spatially organize the millisecond coupling of action potentials to neurotransmitter release. These proteins include ELKS-family proteins, Rab3-interacting molecules (RIMs), RIM-binding proteins, UNC-13 proteins, Bassoon, and Piccolo (Michel et al., 2015). The *Drosophila* active zone cytomatrix, which lacks much of the genetic redundancy of mammalian active zones, comprises a single representative of each family: ELKS homolog Bruchpilot, RIM, RBP, DUNC-13, and Piccolo homolog Fife (Aravamudan et al., 1999; Wang and Sudhof, 2003; Wagh et al., 2006; Mittelstaedt and Schoch, 2007; Bruckner, Gratz et al. 2012).

Synchronized neurotransmitter release depends on the detection of localized domains of increased intracellular Ca^{2+} by release-ready synaptic vesicles. This requires the regulation of Ca^{2+} channel levels, localization and function at active zones; the establishment and maintenance of synaptic vesicles docked at the membrane and molecularly primed for Ca^{2+}-triggered release, termed the readily releasable pool; and the micro- or nanodomain coupling of Ca^{2+} channels and release-ready vesicles to establish active zone release properties. Studies in *Drosophila*, *C. elegans* and mice are demonstrating the key roles played by active zone cytomatrix proteins in support of the functional requirements of regulated neurotransmission (Michel et al., 2015, Ackermann et al., 2015). Recent studies combining high-resolution imaging and functional analysis have begun to elucidate the links between active zone structure and release properties in

both baseline neurotransmission and synaptic plasticity (Bruckner et al., 2015; Gundelfinger, Fejtová, 2012; Chua et al., 2014).

We previously identified the long-overlooked invertebrate Piccolo-RIM related protein Fife (Bruckner, Gratz et al., 2012). Fife localizes to active zones where it promotes the close association of the active zone cytomatrix with the membrane, synaptic vesicle exocytosis and motor behavior. Here, we undertake a detailed functional and morphological analysis of the mechanism of Fife function in regulating neurotransmitter release. We demonstrate that Fife promotes release non-redundantly and in close association with RIM. Paired-pulse recordings and functional studies in the presence of the slow Ca^{2+} buffer EGTA suggest that Fife regulates the probability of neurotransmitter release through a role in coupling Ca^{2+} channels to synaptic vesicle release sites. Cumulative quantal analysis and high-pressure freeze/freeze substitution (HPF/FS) electron microscopy demonstrate that Fife determines the size of the readily releasable vesicle pool. Through tomographic reconstruction of active zone ultrastructure and super-resolution microscopy, we find that the active zone cytomatrix is reduced in size and adopts an abnormal structure the absence of Fife. Together, our findings reveal a central role for Fife in the regulation of two key parameters of synaptic efficacy, the size of the readily releasable pool of synaptic vesicles (n) and the probability of release (p_r). We propose that Fife organizes active zone nanodomains to position release-ready vesicles in close proximity to Ca^{2+} channels in support of high-probability neurotransmitter release.

Materials and methods

Genetics and genome engineering

The following fly lines are available at the Bloomington Drosophila Stock Center: w^{1118}, Df(3L)Exe16091 (referred to here as *Fife Df*), Df(3R)ED5785 (referred to here as *RIM Df*), $elav^{c155}$*-Gal4* (Lin and Goodman, 1994) and *vasa-Cas9* (Gratz, Ukken et al., 2014). $Fife^{ex1027}$ and UAS-Fife are described in Bruckner, Gratz et al., 2012. rim^{ex73} and rim^{103} were obtained from Aaron DiAntonio and Graeme Davis, respectively (Graf et al., 2012; Muller et al., 2012). $Fife^{AC}$ was generated via CRISPR-mediated HDR (Gratz, Ukken et al., 2014). Briefly, two targeting gRNAs and a donor plasmid containing 1-kb homology arms flanking an attP, 3xP3-DsRed cassette were injected into *vasa-Cas9* embryos. Engineered lines were identified by DsRed expression in the eye and confirmed molecularly. This resulted in a 25.6-kb deletion from 2,664,461 to 2,690,083 on chromosome 3L encompassing the first 12 coding exons of *Fife* while avoiding two nested genes on the opposite strand.

Electrophysiology

Current clamp recordings were performed as previously described (Bruckner, Gratz et al., 2012). Male third instar larvae were dissected in Ca^{2+}-free HL3 (70mM NaCl, 5mM KCl, 20mM MgCl2, 10mMNaHCO3, 115 mM sucrose, 5 mM trehalose, 5 mM HEPES, pH 7.2) (Stewart et al. 1994). Recordings were performed in HL3 saline containing 0.6 mM Ca^{2+} unless indicated. A sharp borosilicate electrode (resistance of 15-25 MΩ) filled with 3M KCl was used to impale muscle six of segments A3 and A4. All dissections and recordings were performed at 25°C, and all cells analyzed had an initial resting potential between -60 and -80 mV and input resistance ≥ 5 MΩ. For each cell, 60 consecutive miniature excitatory junction potentials (mEJPs) were collected with pClamp (Molecular Devices) and analyzed using MiniAnalysis (Synaptosoft) to obtain mean amplitude and frequency.

Excitatory junction potentials (EJPs) were stimulated by applying a 1-ms pulse to the cut end of the segmental nerve innervating the impaled muscle cell. Stimulus amplitude was adjusted to reliably recruit both the 1s and 1b inputs onto muscle 6. 100 consecutive EJPs were recorded for each cell and analyzed in pClamp to obtain mean EJP amplitude. Quantal content was calculated for each NMJ as mean EJP amplitude divided by mean mEJP amplitude.

For paired-pulse recordings, stimulus amplitude was adjusted to recruit 1s and 1b nerve inputs and five stimuli were applied at 20 Hz. Each NMJ was stimulated with ten 20-Hz trains, with five seconds of rest between each train. Trains were averaged for each cell and the mean amplitude of each EJP was measured.

The effect of the slow Ca^{2+}-chelation on neurotransmitter release was assessed by pre-treating dissected preparations with the cell permeable EGTA-AM (ThermoFisher). EGTA-AM was diluted to 25mM in DMSO, aliquoted, and stored at -20°C. Stock solutions were thawed immediately before use, diluted to 25 μM in Ca^{2+}-free HL3, and applied to the filleted larval body wall for 10 minutes. Following treatment, preparations were washed for 5 minutes with HL3 with Ca^{2+} before recording EJP and mEJP amplitudes as described above.

The size of the readily releasable vesicle pool was estimated using linear back extrapolation of cumulative quantal content during high-frequency stimulus trains in two-electrode voltage clamp recordings. Recordings were conducted in HL3 containing 1 mM Ca^{2+}, a concentration at which most *Fife* cells showed clear depression; those that did not were excluded from the analysis to prevent underestimation of pool size. Briefly, muscle 6 of abdominal segments A3 and A4 was impaled with voltage monitoring

(resistance 15-25 MΩ) and current passing (resistance 7-12 MΩ) sharp electrodes filled with 3M KCl. Measurements of initial resting potential did not differ by more than 5 mV between the two electrodes, and cells with an initial resting potential > -50 mV or input resistance ≤ 5 MΩ were rejected. Cells were clamped at -70 mV for both EJC and mEJC recordings and clamp gain was adjusted to achieve voltage error of 1-5%. Holding current did not exceed -5 nA. EJCs were evoked by applying 1-ms pulses to the cut end of the appropriate segmental nerve and stimulus amplitude was adjusted to reliably recruit the combined 1s and 1b inputs. 1-ms pulses were applied at 60 Hz, and the amplitude of each EJC from peak to baseline immediately before each stimulus was measured in Clampfit. Quantal content during the high-frequency train was calculated by dividing the amplitude of each EJC in the train by the average mEJC amplitude from that cell. The apparent size of the readily releasable pool was estimated for each NMJ by a linear fit to the last 10 of 30 quantal content measurements in each 60-Hz train, when a steady state of neurotransmitter release from the recycling pool had been reached, and back extrapolating to time zero.

Co-immunoprecipitation

Fife and *Rim* cDNAs were obtained by reverse transcription from wild-type flies using SuperScript III Reverse Transcriptase (Invitrogen), as described previously (Bruckner, Gratz et al., 2012). C2-domain fragments were subcloned into pAFW or pAHW expression vectors for expression in S2 insect cells (*Drosophila* Gateway™ Vector Collection, generated by Terence Murphy and available at the Drosophila Genomics Resource Center).

S2 cells were transfected with Fife and/or RIM expression vectors using Effectene per manufacturer's instructions (Qiagen). At 24 hours, S2 cells were lysed in IP buffer

(0.05% Triton X-100, 100 mM KCl, 20 mM HEPES [pH 7.5], 5% glycerol, and a protease inhibitor cocktail (Roche)) and centrifuged at 18,000g at 4°C for 10 min. Soluble fractions were incubated overnight with anti-FLAG (Sigma) or IgG, followed by incubation with prewashed protein A/G agarose beads (50% slurry, Thermo Scientific) for 3 hours at 4°C. Beads were washed extensively with lysis buffer and bound proteins eluted in Laemmli buffer for analysis by Western blotting with mouse monoclonal anti-FLAG (1:4000; Sigma #F1804) and rabbit monoclonal anti-HA (1:3000; Cell Signaling #3724). Interactions were confirmed in at least three independent experiments.

Immunohistochemistry

NMJ dissections and antibody stains were carried out as previously described (O'Connor-Giles et. al., 2008). The following antibodies were used in this study: mouse anti-Bruchpilot at 1:500 (NC82; developed by Erich Buchner and obtained from the Developmental Studies Hybridoma Bank), chicken anti-GFP at 1:500 (Abcam #13790), and rabbit anti-GluRIIC at 1:2500 (Marrus et al., 2004). Species-specific Alexa Fluor 488 and 568 secondary antibodies (Invitrogen) were used at 1:500.

Confocal imaging and analysis

Confocal images were acquired on a Zeiss LSM 510 and processed using the Fiji distribution of ImageJ (Schindelin et al., 2012). Overall brightness and/or contrast were adjusted in Photoshop. For analysis of Cacophony puncta intensity, all genotypes were stained together and imaged with identical settings. For consistency, analysis was limited to NMJ 4 in segments A2 and A3. To measure Cac-GFP and GluRIIC area and intensity, nonsynaptic structures including axons were removed from the images using freehand selection and fill. Z-stacks were flattened using the Maximum Intensity Z-projection function. Channels were separated and a threshold was applied to remove irrelevant

lower intensity pixels. Separation of individual puncta was facilitated by the Find Maxima tool in Fiji. Intensity and area data were collected using the Analyze Particles tool.

SIM imaging and analysis

SIM images were acquired on a Nikon N-SIM microscope with 100x (1.49 NA) Apo TIRF oil immersion objective (Nikon) and Andor iXon 897 EMCCD camera, and reconstructed using NIS-Elements Ar (Nikon) software. Following reconstruction, images were processed and analyzed in Fiji (Schindelin et al., 2012). Z-stacks were flattened using the Maximum Intensity Z-projection function. For each NMJ image, approximately 20 central Bruchpilot punctae in planar orientation were selected, blind to genotype, for analysis. For each active zone, a fluorescence intensity profile plot was generated along a one micron line drawn along the longest axis through the center of the Bruchpilot spot.

High-pressure freeze and freeze substitution

High-pressure freeze (HPF) and freeze substitution (FS) was adapted from previous work in both *C. elegans* and *Drosophila* (Weimer 2006, Stigloher et al., 2011, Matkovic et al., 2013). Late 2nd instar larvae were placed in the 200-µm cavity of aluminum specimen carriers (Type A, with the flat side of Type B carriers used as a lid, Technotrade International), filled with a mixture of 10% BSA (Sigma) and OP50 *E. coli* in PBS. Specimen carriers were pretreated with 1-Hexadecene (Sigma) to adequately seal the freezing chamber. HPF was performed with a HPM-010 (Bal-Tec) apparatus at a freezing speed >20, 000K/s and pressure >2000 bar. FS was performed in an EM AFS1 (Leica). Samples were incubated in 0.5% glutaraldehyde, 0.1% tannic acid, and 1% H_2O in anhydrous acetone for 98 hours at -90°C, rinsed with 1% H_2O in acetone for 2 h at -90°C, then incubated in 1% OsO4 and 1% H_2O in acetone for 7 h at -90°C. Samples were

warmed to -20°C over 14 hours (5°C/hour), then incubated at -20°C for 16 hours. Temperature was increased to 4°C over 2.4 hours (10°/h). Samples were then rinsed with 1% H_2O in acetone for 2 hours, transferred to room temperature, and embedded in Epon using standard procedures.

Electron microscopy and electron tomography

Ultrathin sections (gray-silver) were collected on pioloform-coated copper slot grids and stained with 8% uranyl acetate in 50% ethanol and Reynold's lead citrate. Images were collected on a Phillips CM120 transmission electron microscope at 80 KeV. For electron tomography, 250-nm sections were cut, collected onto pioloform-coated copper slot grids, and stained with 2% aqueous uranyl acetate and Reynold's lead. For fiducial-aided image alignment, grids were treated with 10-nm Aurion Gold solution (Electron Microscopy Sciences). Boutons were preselected using a Phillips CM120 transmission electron microscope at 80KeV. Freezing quality was evaluated by surveying the preservation of muscle cell nuclei that are sensitive to ice crystal formation, and damaged specimens were excluded. Serial tilt acquisition was automatically conducted from -60° to 60° in 1° increments at 300KeV using SerialEM software (Mastronarde et al., 2005) on an FEI TF-30 transmission electron microscope equipped with a GATAN 2k x 2k ultrascan camera. Tilt-image series were aligned and tomograms reconstructed using fiducial-guided alignment in ETomo/IMOD, and virtual slices of tomogram modes were generated (Kremer et al., 1996).

Image segmentation and quantification

Segmentation and quantification of membrane-proximal synaptic vesicles, dense projections, and cytosolic filaments was conducted using Amira 6.0.1 (FEI). Membrane-proximal synaptic vesicles were defined as the population of synaptic vesicles within 5

nm of the presynaptic membrane, including synaptic vesicles making direct morphological contact and linked by tethers <5 nm in length. Active zone length is unaltered at *Fife* synapses (Bruckner, Gratz et al., 2012;), so to account for variations in the extent of active zone captured in random serial sections, quantifications of RRP size from thin sections were normalized to active zone length in the quantified image.

Statistical analyses

Single comparisons were conducted by Student's *t* test. Welch's correction was used in cases of unequal variance. Multiple comparisons were performed by ANOVA followed by *post hoc* tests with Šidák correction to maintain an experimentwise significance level of 0.05. Chi-square tests were used to compare proportions. Excel, Prism and SigmaPlot wer used for data analysis, statsitical testing and curve fitting.

Results

Fife and RIM promote neurotransmitter release non-redundantly

Fife and RIM are closely related proteins that share a common domain structure, colocalize with Bruchpilot at active zones where they both promote neurotransmitter release, and, in mammals, have multiple binding partners in common (Bruckner, Gratz et al., 2012; Graf et al., 2012; Muller et al., 2012; Wang et al., 2012). The similarity between the two proteins raises the possibility that they may regulate neurotransmitter release redundantly. To test this possibility, we generated double mutants of a null *RIM* allele, RIM^{ex73}, and a precise deletion of *Fife* generated through CRISPR-mediated homology-directed repair, $Fife^{AC}$ (Materials and Methods; Graf et al., 2012). Similar to a previously characterized *Fife* excision allele, $Fife^{AC}$ displays a reduction of ~60% in neurotransmitter release and is rescued by neuronal expression of a full-length *Fife* cDNA (Figure 1A,B,F-H; control: 32.6 ± 1.7, n = 7 NMJs vs. $Fife^{AC}/Df$: 13.2 ± 1.7, n = 15 vs. *C155-Gal4/*Y; *UAS-Fife*, $Fife^{AC}/Df$ (Rescue): 42.5 ± 4.1, n = 9; Bruckner, Gratz et al., 2012). As previously observed, we find that neurotransmitter release at RIM^{ex73}/Df NMJs is decreased by approximately 65% (Figure 1A,C,F-H; 11.3 ± 2.1, n = 13 NMJs; Graf et al., 2012). Loss of both *Fife* and *RIM* yields viable flies and results in a neurotransmitter release deficit similar to loss of either gene alone, indicating that Fife and RIM function non-redundantly to promote neurotransmitter release (Figure 1D,F-H; $Fife^{AC}$, RIM^{ex73}*/Fife Df, RIM Df*: 12.9 ± 1.4, n = 14 NMJs). Further analysis confirms the prediction that the two active zone proteins function closely together. While larvae lacking one copy of either *RIM* or *Fife* exhibit normal neurotransmitter release, quantal content is decreased by nearly 65% in double heterozygous larvae (Figure 1E,F-H; $Fife^{AC}$/+: 32.2 ± 1.8, n = 9 NMJs vs. RIM^{103}/+: 24.0 ± 2.4, n = 12 vs. $Fife^{AC}/RIM^{103}$: 11.6 ± 1.3, n = 12). Non-allelic non-complementation is often indicative of proteins that function together within a complex. This is consistent with the colocalization of Fife and

RIM to the active zone cytomatrix at the *Drosophila* NMJ and in the mouse CNS, where Piccolo and RIM are both found within 20-50 nm of the presynaptic membrane (Dani et al., 2010; Bruckner, Gratz et al., 2012, Graf et al., 2012).

Piccolo and RIM proteins both contain C-terminal C2 domains through which they interact in rodents (Fujimoto et al., 2002; Shibasaki et al., 2004). Canonical C2 domains mediate Ca^{2+}-dependent lipid binding through five conserved aspartate residues as well as protein-protein interactions, whereas non-canonical C2 domains may confer Ca^{2+}-independent lipid binding (Cho and Stahelin, 2006). To determine if Fife and RIM interact through their C2 domains, we conducted coimmunoprecipitation experiments in *Drosophila* S2 cells. When we express FLAG-tagged versions of either the Fife C2A or C2B domain, we observe interactions with the RIM C2B domain (Figure 1I, lanes 1, 3, 5 and 7). The interaction is specific as no RIM is detected in the absence of FLAG antibodies or FLAG-Fife (Figure 1I lanes 2, 4, 6 and 8). While we cannot rule out an interaction, we do not detect interactions between Fife and RIM's C2A domain under these conditions (data not shown). Together, these data indicate that Fife and RIM carry out independent functions in close association at the active zone cytomatrix.

Fife regulates probability of release

To determine the mechanism of Fife-dependent regulation of neurotransmitter release, we conducted a systematic analysis of the three quantal parameters that determine synaptic efficacy: the size of the post-synaptic response to the release of an individual synaptic vesicle, or quantal size (q), the average probability of synaptic vesicle release in response to Ca^{2+} influx (p_r), and the total number of release-ready vesicles (n). Electrophysiological measurements of mEJP amplitude indicate that Fife does not

regulate quantal size, suggesting that Fife regulates neurotransmitter release through modulation of p_r, n, or both (Figure 1B,F; Bruckner, Gratz et al. 2012).

In low extracellular Ca^{2+} concentrations, *Fife* NMJs fail to respond to a stimulus significantly more frequently than control, suggesting that p_r may be impaired in the absence of Fife (wild type: 1.7 ± 1.1%, n = 6 NMJs vs. $Fife^{ex/ex}$: 24.1 ± 6.6%, n = 11 at 0.2 mM external Ca^{2+}; p = 0.007). To quantify the effect of loss of Fife on p_r, we applied closely spaced sequential presynaptic stimuli and measured the degree of facilitation. At active zones with low release probability, the accumulation of residual Ca^{2+} following a second, closely spaced stimulus enhances neurotransmitter release. Thus, paired-pulse facilitation is inversely correlated with p_r. In response to paired stimuli at 20 Hz in 0.6 mM external Ca^{2+}, wild-type NMJs exhibit almost no facilitation as expected, whereas *Fife* NMJs exhibit an approximately 60% increase in EJP amplitude (Figure 2A-C; paired-pulse ratio in wild type: 1.01 ± 0.04, n = 10 NMJs vs. $Fife^{AC}/Df$: 1.58 ± 0.16, n = 14; p = 0.003). Together, the increased failure rate and paired-pulse facilitation indicate significant impairment of p_r in the absence of Fife.

Ca^{2+} channels are reduced at *Fife* active zones

The levels and localization of Ca^{2+} channels at active zones are key determinants of p_r , and have been shown to be regulated by active zone cytomatrix proteins, including RIM (Graf et al., 2012). To investigate the possibility that altered Ca^{2+} channel levels underlie reduced p_r in the absence of Fife, we expressed GFP-tagged Cacophony, the pore-forming subunit of the Drosophila CaV2-type Ca^{2+} channel, in *Fife* and wild-type neurons (Kawasaki et al., 2002). We observe a 16% decrease in Cacophony levels, indicating a mild impairment (Figure 3A,B; wild type: 109.6 ± 6.0 a.u., n = 1998 active zones at 6 NMJs vs. $Fife^{ex}/Df$: 92.5 a.u. ± 3.2, n = 2093 active zones at 7 NMJs; p =

0.037). We previously observed a similar trend in Bruchpilot and GluR levels (Bruckner, Gratz et al., 2012), suggesting Fife may regulate Ca^{2+} channel levels in the context of a more general role establishing and/or maintaining the molecular organization of active zones. Consistent with this conclusion, when we normalize Cacophony levels to GluR levels measured in the same experiment, we no longer observe a difference between wild type and *Fife* (Figure 3A,C,D).

Fife promotes the coupling of synaptic vesicles and calcium channels

The slightly lower Ca^{2+} channel levels at *Fife* active zones combined with normal bouton-level Ca^{2+} influx points to a role for Fife in the local regulation of Ca^{2+} channel localization at active zones. Release-ready synaptic vesicles are poised to fuse with the membrane in response to transiently elevated Ca^{2+} concentrations generated near voltage-dependent Ca^{2+} channels that open in response to an action potential. Thus, p_r depends critically on the positional coupling of Ca^{2+} influx and synaptic vesicle release sites. High-probability synchronous release at many synapses, including those of the *Drosophila* NMJ, is achieved through the nanodomain (defined as < 100 nm) coupling of molecularly primed synaptic vesicles and Ca^{2+} channels (Kittel et al., 2006; Eggerman et al., 2012). To determine if Fife regulates the probability of neurotransmitter release by promoting nanodomain coupling, we investigated the effect of EGTA on mutant and wild-type synapses. Because EGTA is a slow Ca^{2+} chelator, at low concentrations its effects are limited to release-ready synaptic vesicles positioned at distances greater than 100 nm from Ca^{2+} channel clusters (Smith et al. 1984, Eggerman et al. 2012). Therefore, wild-type synapses exhibiting nanodomain coupling should be insensitive to EGTA, whereas synapses with impaired positional priming will exhibit sensitivity. We measured EJP and mEJP amplitude at *Fife* and wild-type NMJs with and without a 10 minute pre-treatment with 25 μM membrane permeable EGTA-AM. Consistent with nanodomain

coupling at the *Drosophila* NMJ, average EJP amplitude was unaffected at wild-type NMJs (Figure 4A,B,E; without EGTA: 30.2 ± 1.0 mV, n = 16 NMJs vs. with EGTA: 29.1 ± 2.0 mV, n = 12 NMJs, p = 0.60). Quantal content was similarly unaffected by exposure to EGTA in wild type (Figure 4F,G; without EGTA: 31.6 ± 1.7, n = 16 NMJs vs. with EGTA: 33.5 ± 2.1, n = 12 NMJs, p = 0.47). In contrast, at *Fife* NMJs EJP amplitude is reduced 44% following EGTA exposure (Figure 4C-E; without EGTA: 13.8 ± 1.4 mV, n = 16 NMJs vs. with EGTA: 7.7 ± 1.7 mV, n = 11 NMJs, $p < 0.05$). Quantal content is reduced 42% at *Fife* NMJs following EGTA exposure, consistent with a severe impairment in the positional priming of synaptic vesicles (Figure 4F,G; without EGTA: 15.0 ± 1.7, n = 16 NMJs vs. with EGTA: 8.7 ± 1.5, n = 11 NMJs, $p < 0.05$). Therefore, we conclude that Fife regulates p_r by promoting the nanodomain coupling of Ca^{2+} channels and readily releasable synaptic vesicles.

Fife regulates the size of the readily releasable pool of synaptic vesicles

We next sought to determine if Fife plays a role in establishing and/or maintaining the number of release-ready synaptic vesicles (n). Generally considered to be those vesicles molecularly primed to fuse in response to action potential-triggered Ca^{2+} influx, the readily releasable pool can be measured both physically and functionally (Rosenmund and Stevens 1996; Neher, 2015). Recent studies have employed high-frequency stimulus trains that result in short-term synaptic depression caused by depletion of the readily releasable pool to obtain physiological estimates and identify the molecular determinants of pool size (Schneggenburger et al. 1999, Hallermann et al., 2010; Miskiewicz et al., 2011; Weyhersmuller et al., 2011;). Following depletion, a steady-state of release representing the immediate release of vesicles trafficked from the recycling pool is observed. Fitting a line to the steady-state phase of cumulative quantal content

measurements during the high-frequency train enables back extrapolation of the readily releasable pool size.

To assess whether Fife regulates the number of readily releasable vesicles, we measured the amplitude of postsynaptic responses in two-electrode voltage clamp mode during 60-Hz stimulus trains at 1 mM $[Ca^{2+}]_e$. We then calculated quantal content and plotted cumulative quantal content. By fitting a line to the last 10 of 30 cumulative quantal content measurements for each NMJ and back extrapolating to time zero, we determined the average readily releasable vesicle pool size at wild-type and *Fife* active zones for direct comparison. The number of release-ready vesicles calculated at Fife NMJs is 38% lower than wild type (Figure 5A,B; wild type: 647 ± 67 vesicles, n = 9 NMJs vs. $Fife^{AC}/Df$: 403 ± 88 vesicles, n = 6 NMJs; p = 0.043), consistent with a requirement for Fife in determining the size of the readily releasable pool. In contrast, the average slope of the steady state period of cumulative quantal release is not significantly different between *Fife* and wild-type NMJs, indicating that replenishment of released synaptic vesicles occurs independently of Fife (Figure 5A,C; wild type: 85.9 ± 11.0, n = 9 NMJs vs. $Fife^{AC}/Df$: 89.9 ± 8.6, n = 6 NMJs; p = 0.80).

Fife regulates the number of morphologically docked synaptic vesicles

To further investigate Fife's role in establishing the readily releasable vesicle pool, we performed complementary imaging studies. Work in diverse organisms and cell types suggests that when visualized by transmission electron microscopy, the readily releasable pool comprises synaptic vesicles in direct contact with or tethered to the active zone membrane by short filaments (Fernandez-Busnadiego et al. 2010, Lanzavecchia et al. 2005, Zampighi et al. 2005, Chen et al. 2015). Our previous ultrastructural studies of aldehyde-fixed samples identified a 20% decrease in total synaptic vesicles clustered at

active zones, but could not address the size of the readily releasable pool as aldehyde fixation has been shown to alter membrane docking of synaptic vesicles (Siksou et al. 2009; Bruckner, Gratz et al., 2012;). To overcome this caveat, we quantified synaptic vesicles within 5 nm of the active zone membrane at wild-type and *Fife* synapses from second-instar larvae prepared for electron microscopy via HPF/FS, which enables instant immobilization of intact larvae at a near-to-native state without crystal ice formation. In agreement with our electrophysiological estimates, we found approximately 30% fewer docked vesicles at *Fife* active zones (Figure 5D-F; wild type: 3.9 ± 0.46 vesicles/600 nm active zone length, n = 30 vs. $Fife^{AC}/Df$: 2.7 ± 0.30 vesicles/100 nm active zone length, n = 25; p = 0.03).

For 3-D resolution, we turned to electron tomography of HPF/FS-prepared samples. The tomographic reconstruction of each active zone in 250-nm sections enables ultrahigh-resolution 3D reconstruction of cytoplasmic structures through semiautomatical segmentation (see Material and Methods). As has been observed in previous studies, the collapsed T-bar morphology of the electron-dense cytomatrix routinely observed in traditional preparations is resolved as a complex network of proteinaceous filaments that clusters synaptic vesicles at both wild-type and *Fife* active zones (Figure 6A-?; Matkovic et al., 2013; Fouquet et al., 2009; Kittelmann et al., 2013; Siksou et al., 2007; Zhao et al., 2012b; Helmprobst et al., 2015). Figure 7x provides a representative 3D model of a presynaptic active zone with synaptic vesicles surrounding the cytomatrix, or T-bar.

As previously observed at the frog and fly NMJ, the dense projection is linked to the presynaptic membrane by short regularly spaced “pegs” hypothesized to represent cytoplasmic complexes including Ca^{2+} channels (Figure 6A-?, Jiao et al., 2010; Harlow et al., 2001). Further, our tomographic reconstruction of HPF/FS-prepared samples

enables the visualization of a complex network of filaments linking synaptic vesicles clustered at the presynaptic terminal. The active zone-associated pool includes synaptic vesicles in direct morphological contact with the presynaptic membrane, linked to the membrane by tethers ranging from 3-55 nm in length, and clustered within the terminal either tethered to the dense projection or linked only to neighboring vesicles (Figure 6A-B). Similar to observations in other organisms, the majority of synaptic vesicles at the *Drosophila* NMJ are linked to their neighbors by one or two connecting filaments (Figure 6F-J, Siksou et al., 2007).

We quantified the number of membrane-associated synaptic vesicles in our tomograms. In agreement with our electron microscopic analysis of ultrathin sections, we note consistent trends toward fewer membrane-associated vesicles in our tomographic reconstructions of *Fife* active zones (Figure 6?; wild type: 3.6 ± 0.51 vs. $Fife^{AC}/Df$: 2.0 ± 0.55, $p = 0.065$, $n = 5$ tomograms per genotype). Thus, both electrophysiological and morphological measurements indicate that Fife is a critical determinant of the size of the readily releasable vesicle pool.

Fife determines active zone ultrastructure

Our functional and morphological analyses demonstrate that Fife promotes neurotransmission by regulating the size of the readily releasable vesicle pool and coupling release-ready vesicles to Ca^{2+} channels. While the active zone cytomatrix has long been hypothesized to play a central role in organizing presynaptic terminals for neurotransmission, genetic studies are now beginning to provide precise links between the structure and function of active zones (Chua et al., 2014). These studies suggest that a key role of the active zone cytomatrix may be to establish release slots where synaptic vesicles can dock in close proximity to voltage-dependent Ca^{2+} channels (Matkovic et al.,

2013; Catterall et al., 2008; Cao et al., 2004). Consistent with a role in organizing active zone ultrastructure, we observe a clear decrease in the complexity and size of the electron-dense cytomatrix in our tomographic reconstructions of *Fife* active zones (Figure 7A-F). Our previous electron microscopic analysis revealed the presence of occasional membrane-detached, or floating, electron densities that maintained the ability to cluster synaptic vesicles in *Fife* mutants (Bruckner, Gratz et al., 2012). At *Fife* active zones that contain T-bars, they are significantly reduced in size (Figure 7G,H; maximum T-bar width in wild-type: 121.0 ± 17.1, n = 13 vs. $Fife^{ex}/Df$: : 80.4 ± 7.8, n = 17; p= 0.04). Although it is difficult to quantify due to its filamentous nature, we see a qualitatively similar decrease in our HPF thin sections as well (Figure 7I-L). These data suggest Fife organizes active zone ultrastructure in support of its functional roles in establishing the readily releasable vesicle pool and Ca^{2+} channel-vesicle nanocomplexes to promote high probability neurotransmitter release.

Fife regulates the molecular organization of active zones

The CAST/ELKS homolog Bruchpilot is a major molecular component of T-bars at the Drosophila NMJ (Kittel et al., 2006). To determine if the decreased cytomatrix complexity we observe in our ultrastructural analyses is reflected in alterations to the molecular composition of the cytomatrix, we employed structured-illumination microscopy to assess Bruchpilot organization. Superresolution imaging studies employing stimulated emission depletion (STED) microscopy have shown that the C-terminus of Bruchpilot is arranged in a donut-shaped pattern at active zones (Kittel et al., 2006). Structured-illumination microscopy (SIM) reveals a similar organization at wild-type synapses (Figure 8E). In contrast, Bruchpilot is disordered in *Fife* mutants and rarely displays a clear donut shape (Figure 8F). To quantify this difference, we selected planar-oriented active zones in wild-type and *Fife* mutant NMJs, blind to genotype, and

analyzed active zone morphology using a fluorescence intensity profile along a line drawn through the center of each Bruchpilot spot. The average plot profile of *Fife* mutant Bruchpilot spots indicated the loss of the two peak profile seen in wild-type donut-shaped Bruchpilot spots (Figure 8G,H). In *Fife* mutants, we measured 70% fewer active zones with two peaks than in wild type (Figure 8I; wild type: 75.7 ± 5.4%, n= 198 active zones at ten NMJs vs. $Fife^{AC}/Df$: 23.4 ± 3.5%, n=200 active zones at ten NMJs; $p < 0.0001$). The precise organization of Bruchpilot was recently linked to the creation of synaptic vesicle release slots (Matkovic et al., 2013). Together, our findings indicate that Fife plays a key role in creating release sites where molecularly primed synaptic vesicles are positionally coupled to voltage-dependent Ca^{2+} channels, and may do so in part by organizing the molecular composition and ultrastructural organization of active zones.

Discussion

Precise regulation of synchronized neurotransmitter release underlies the patterned neural activity that controls thought and behavior. The active zone cytomatrix has been proposed to play a key role in organizing presynaptic terminals to establish and modulate synapse-specific neurotransmitter release properties. Here, we show that Fife determines the structure and molecular composition of the active zone cytomatrix, and regulates the number of readily releasable synaptic vesicles and their positional coupling to voltage-dependent Ca^{2+} channels to promote high-probability neurotransmitter release.

Ultrastructure of the active zone

Recent high-resolution reconstructions of presynaptic ultrastructure at a number of model synapses have begun to clarify the structural foundations of regulated neurotransmitter release (Stigloher et al., 2011; Leitinger et al., 2012; Helmprobst et al., 2015, Bruckner et al., 2015; Harlow et al., 2001). While the spatial details vary between synapses with diverse functional requirements, the active zone cytomatrix appears to play a general role in clustering release-ready synaptic vesicles and Ca^{2+} channels at the presynaptic membrane (Zhai and Bellen, 2004). Consistent with this emerging picture, our 3D tomographic reconstructions and serial section electron microscopy of samples preserved in a near native state with HPF/FS vividly reveal a proteinaceous active zone cytomatrix comprising a complex network of interconnected filaments radiating up to 200 nm from the membrane (Figures 6 and 7; Figure 7—figure supplements 1-3). In our tomograms, cytomatrix filaments often contact synaptic vesicles in direct morphological contact with the presynaptic membrane or clustered around the active zone, where they may act to facilitate trafficking of this vesicle population to available release slots. More than 60% of vesicles clustered at the active zone are linked by thin filaments to one or

more other vesicles (Figure 6). A subpopulation of synaptic vesicles is linked not only to neighboring vesicles, but also tethered to the presynaptic membrane by filaments similar in appearance. Whether the formation of tethers between synaptic vesicles is required upstream of docking and fusion remains unclear. It has been proposed that tethers may facilitate the loading and reloading of vesicles tightly coupled to Ca^{2+} channels and/or distinguish evoked vs. spontaneously released vesicles (Hallermann and Silver, 2013). Finally, we observe peg-like structures, thought to be Ca^{2+} channels, clustered between T-bars and the active zone membrane in thin sections of HPF/FS-preserved larvae (Figure 6; Fouquet et al., 2009; Jiao et al., 2010). Our tomography illustrates the highly structured nature of the active zone. Future studies coupling EM tomography and immuno-EM with mutant analyses will undoubtedly be key to elucidating the molecular nature of these AZ structures and to understanding the structural determinants of neurotransmitter release parameters.

Active zone organization

The active zone cytomatrix is smaller, less filamentous and clusters fewer synaptic vesicles at *Fife* mutant synapses (Figure 7). These changes are accompanied by disorganized localization of Bruchpilot, a critical molecular component of the *Drosophila* active zone cytomatrix (Figure 8). The ring of Bruchpilot apparent through superresolution imaging of wild-type active zones in planar orientation is collapsed at the majority of *Fife* active zones. Similar disorganization of Bruchpilot is observed in *RBP* null mutants and isoform-specific Bruchpilot mutants, $brp^{\Delta 170}$ and $brp^{\Delta 190}$ (Liu et al., 2011; Matkovic et al., 2013). Multiple studies have linked release dynamics to the overall size of the active zone cytomatrix across phyla (Holderith et al., 2012; Matkovic et al., 2013; Matz et al., 2010). Our results suggest that Fife may regulate neurotransmitter release by modifying both the size and the molecular organization of the cytomatrix.

The readily releasable vesicle pool

Through physiological and morphological measurements, we also demonstrate a role for Fife in determining the size of the readily releasable vesicle pool (Figures 5 and 7). Although the exact nature of the pool has been difficult to define, the conceptual framework of a small pool of rapidly depleted vesicles primed and available for immediate release upon Ca^{2+} influx holds up across many measures (Pan and Zucker, 2009; Neher, 2015). Electrophysiologically, this pool size can be determined by calculating the number of vesicles released during a high-frequency stimulus prior to steady state release from the recycling pool. The morphological correlate is thought to include the population of vesicles docked at the active zone membrane or linked by filaments of less than 5 nm (Siksou et al., 2009; Fernández-Busnadiego et al., 2013; Gracheva et al., 2010; Imig et al., 2014). In addition to their physical proximity to the active zone membrane, these vesicles are presumed to be molecularly primed, meaning that the release apparatus is assembled such that only Ca^{2+} binding is required to trigger fusion. The number of readily releasable vesicles at an active zone may be a fixed function of cytomatrix structure (Matkovic et al., 2013). If so, by regulating the size and structural organization of the active zone cytomatrix, *Fife* may participate in regulating the number of release sites. The size of the readily releasable pool is also decreased in *RBP*, $brp^{\Delta 170}$ and $brp^{\Delta 190}$ mutants. These mutants also share structural abnormalities with *Fife*, supporting the possibility that active zone cytomatrix proteins may regulate the size of the readily releasable pool of vesicles by creating physical slots for the membrane docking of vesicles.

Synaptic vesicle-Ca^{2+} channel nanocomplexes

In addition to regulating the number of release-ready vesicles, Fife regulates probability of release by coupling release-ready synaptic vesicles and voltage-dependent Ca^{2+} channels (Figures 2-4). Synaptic vesicle release is likely mediated by a very small number of Ca^{2+} channels clustered in close proximity to release-ready synaptic vesicles (Chen et al., 2015; Heidelberger et al., 1994; Cao et al., 2004; Sheng et al., 2012). Thus a key parameter in determining the release probability at active zones is the physical distance between voltage-dependent Ca^{2+} channels and synaptic vesicles primed for Ca^{2+}-dependent release, a parameter referred to as positional priming (Wadel et al., 2007; Neher et al., 2008). At high-probability release synapses, including the mammalian mature Calyx of held and the *Drosophila* NMJ, molecularly primed synaptic vesicles are believed to be positioned within 30-100 nms of voltage-dependent Ca^{2+} channels clusters in nanocomplexes that ensure their tight coupling and explain their observed release characteristics (Eggermann et al., 2012, Keller et al., 2015). It is possible that Fife and other cytomatrix proteins ensure such an arrangement by scaffolding the protein-protein interaction network that links Ca^{2+} channels and release-ready vesicles in a tightly coordinated manner.

As might be expected, the same active zone proteins that regulate *n*, including Bruchpilot, RBP and RIM, also play key roles in determining p_r. High-release probability active zones at the Drosophila NMJ contain higher levels of Bruchpilot, which in turn correlates with higher Ca^{2+} channel levels (Peled and Isacoff, 2011; Graf et al., 2009; Fouquet et al., 2009). At mouse hippocampal synapses, Bassoon levels scale with neurotransmitter release property, and active zone size correlates with both the number of docked vesicles and the size of Ca^{2+} channel clusters (Matz et al., 2010; Holderith et al., 2012). This correlation extends beyond development to synaptic plasticity. In response to decreases in glutamate receptor levels or function, *Drosophila* motorneurons rapidly

increase synaptic vesicle release to maintain postsynaptic excitation (Petersen et al., 1997; Davis, 2006). This homeostatic increase in presynaptic neurotransmission is accompanied by an increase in T-bars and Bruchpilot levels (Reiff et al., 2002; Weyhersmuller et al., 2011). Together, these studies all point to the conclusion that synapses with high release probability are associated with high levels of active zone cytomatrix proteins. An attractive model is that the structure of the active zone cytomatrix, as established by its component molecules and their protein-protein interactions, directly regulates both n and p_r by creating slots for synaptic vesicles to dock within 100 nm of Ca^{2+} channel clusters.

Overall, the emerging picture of the active zone cytomatrix points to a critical role in determining baseline presynaptic function and plasticity. The size, organization and composition of active zone cytomatrix appears tightly correlated with the number of release sites as well as the probability of release suggesting that these molecules determine active zone release parameters. Sorting out the common and unique contributions of each member of the active zone cytomatrix will be critical to understanding how release dynamics are established and modulated.

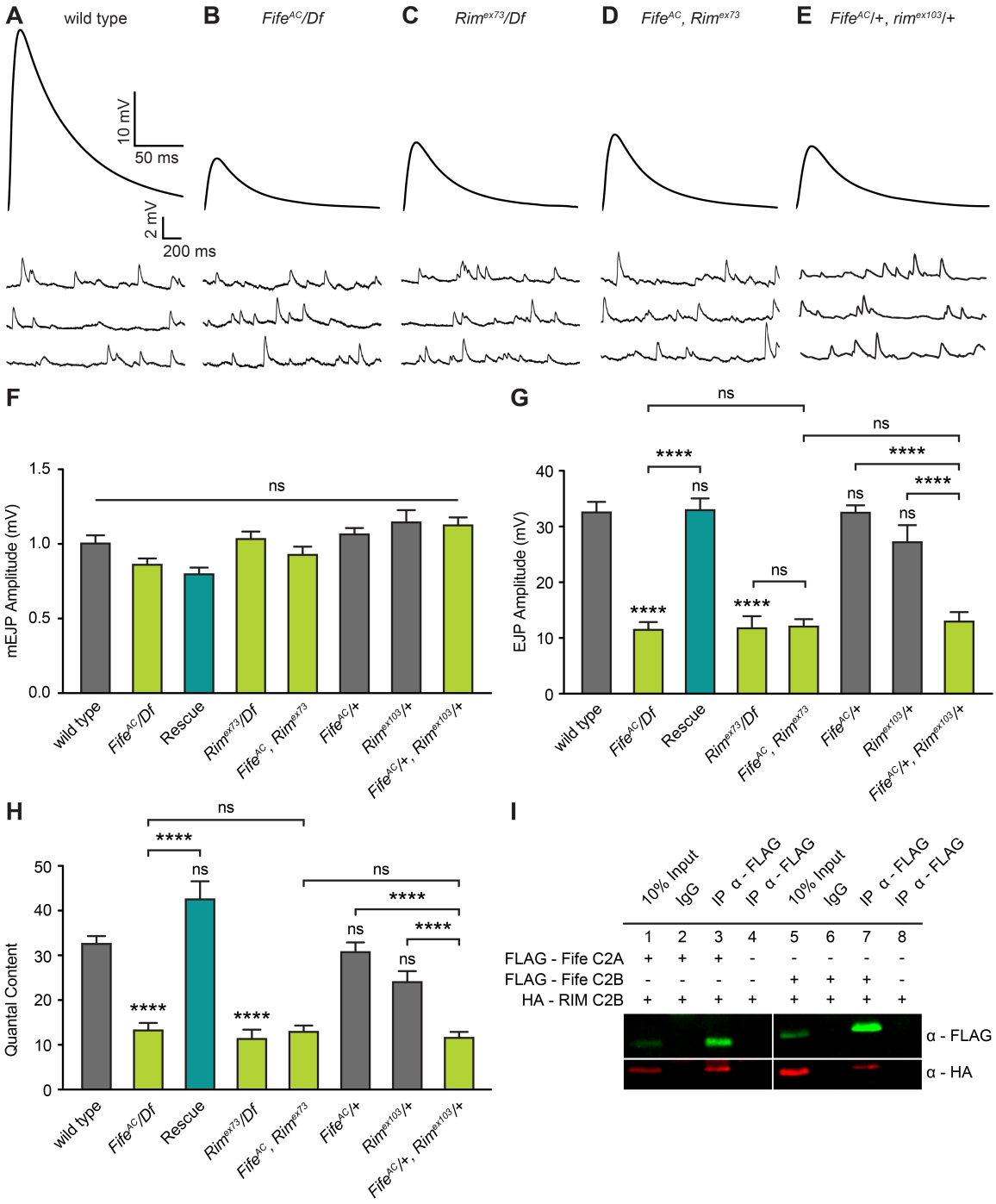

Figure 1. Fife and RIM promote neurotransmitter release non-redundantly. **(A-E)** Representative traces of EJPs and mEJPs recorded in 0.6 mM Ca^{2+} in wild type, n = 7 (A), *Fife^{AC}/Fife Df*, n = 15 (B), *Rim^{ex73}/Rim Df* n = 13 (C), *Fife^{AC}, Rim^{ex73}/Fife Df, Rim Df*, n = 14 (D), and *Fife^{AC}, Rim^{ex103}/+*, n = 12 (E). Stimulus artifacts have been removed for clarity. **(F-H)** Average mEJP amplitude (F), EJP amplitude (G), and quantal content (H) in the genotypes depicted in A-E and *C155-Gal4/*Y; *UAS-Fife*, *Fife^{AC}*/Fife *Df* (Rescue), n = 9; *Fife^{AC}*/+, n = 9; and *Rim^{ex103}*/+, n = 12. Statistical significance represents comparison to control unless otherwise indicated. ns, not significant; ***p<0.001, ****p<0.0001. Error bars represent s.e.m. mEJP amplitude is not significantly altered in

any genotype. EJP amplitude and quantal content are significantly decreased in *Fife* and *rim* single and double mutants, and the reduction in *Fife* is rescued by expression of *Fife* in neurons. EJP amplitude and quantal content are normal in *Fife* and *rim* heterozygotes, but significantly reduced in *Fife, Rim* double heterozygotes. **(I)** Coimmunoprecipitation of HA-tagged Rim C2B with FLAG-tagged Fife C2A or Fife C2B. HA-Rim C2B is coimmunoprecipitated by FLAG-Fife C2A (lane 3) or FLAG-Fife C2B (lane 7), but is not bound by IgG (lanes 2 and 6) or beads alone (lanes 4 and 8).). Input lanes (1, 5) contain lysate equal to 10% of the amount used for the coimmunoprecipitation assays.

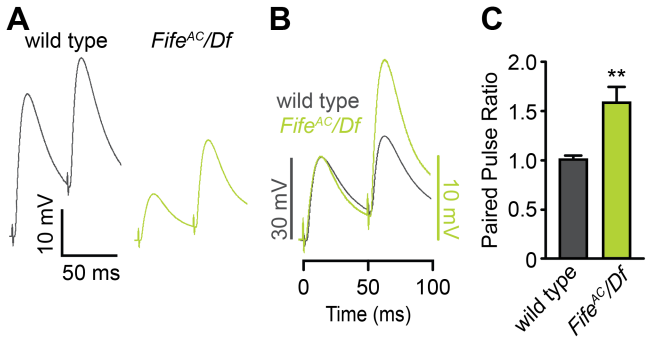

Figure 2. Fife regulates probability of neurotransmitter release. (A) Representative traces of paired EJPs in wild type and *Fife^{AC}/Fife Df* recorded in 0.6 mM Ca^{2+}. **(B-C)** In 0.6 mM Ca^{2+}*Fife* NMJs facilitate significantly in response to paired pulses delivered at 20 Hz, as illustrated by representative traces scaled to the amplitude of the first wild-type pulse (B) and average ratio of the amplitude of the first and second responses (C). The scale of EJPs in (B) is indicated by the adjacent bar of the corresponding color (gray for wild type and green for *Fife^{AC}/Fife Df*). PPF is calculated as the ratio of the amplitude of the second and first pulses. wild type, n = 10 NMJs, *Fife^{AC}/Fife Df*: n = 14 NMJs. **p<0.01. Error bars represent s.e.m.

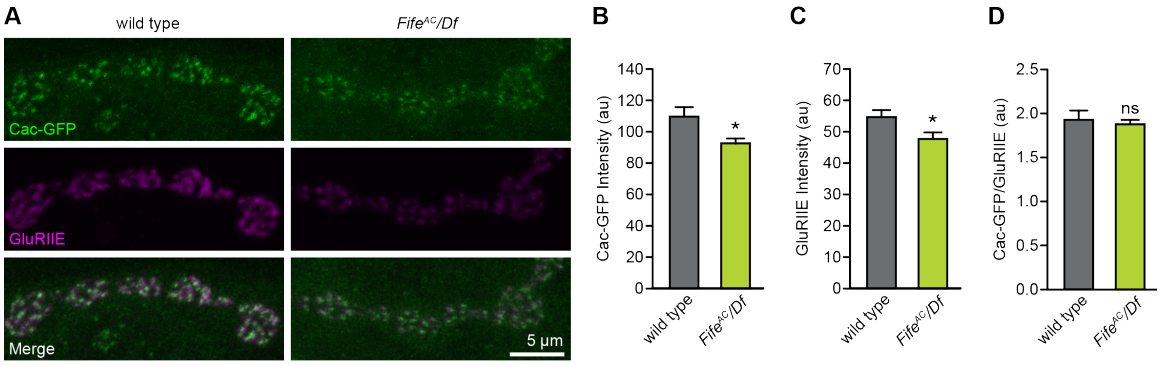

Figure 3. Fife is required for synaptic organization of Ca^{2+} channels and glutamate receptors but not for Ca^{2+} influx. (A) Confocal z-projections of Cacophony Ca^{2+} channel localization at wild type (*elav^{C155} Gal4*; UAS-Cac-GFP/+) and *Fife^{AC}/Fife Df* (*elav^{C155} Gal4*; UAS-Cac-GFP/+; *Fife^{AC}/Fife Df*) NMJ 4s colabeled with antibodies to GFP and GluRIIE. (B-D) Quantifications of raw Cac-GFP (B), raw GluRIIE (C) and normalized Cac-GFP (D) fluorescence intensity in wild-type (n = 6 NMJs) and *Fife^{AC}/Df* (n = 7 NMJs) larvae. (B) Raw Cac-GFP levels are decreased 15% at *Fife^{AC}/Df* NMJs ($P<0.05$). (C) Similarly, raw GluRIIE levels are reduced 13% at *Fife^{AC}/Df* NMJs ($P<0.05$). (D) Because both proteins are decreased to a similar degree, there is no significant difference in Cac-GFP levels when Cac-GFP fluorescence is normalized to GluRIIE fluorescence. Error bars represent s.e.m.

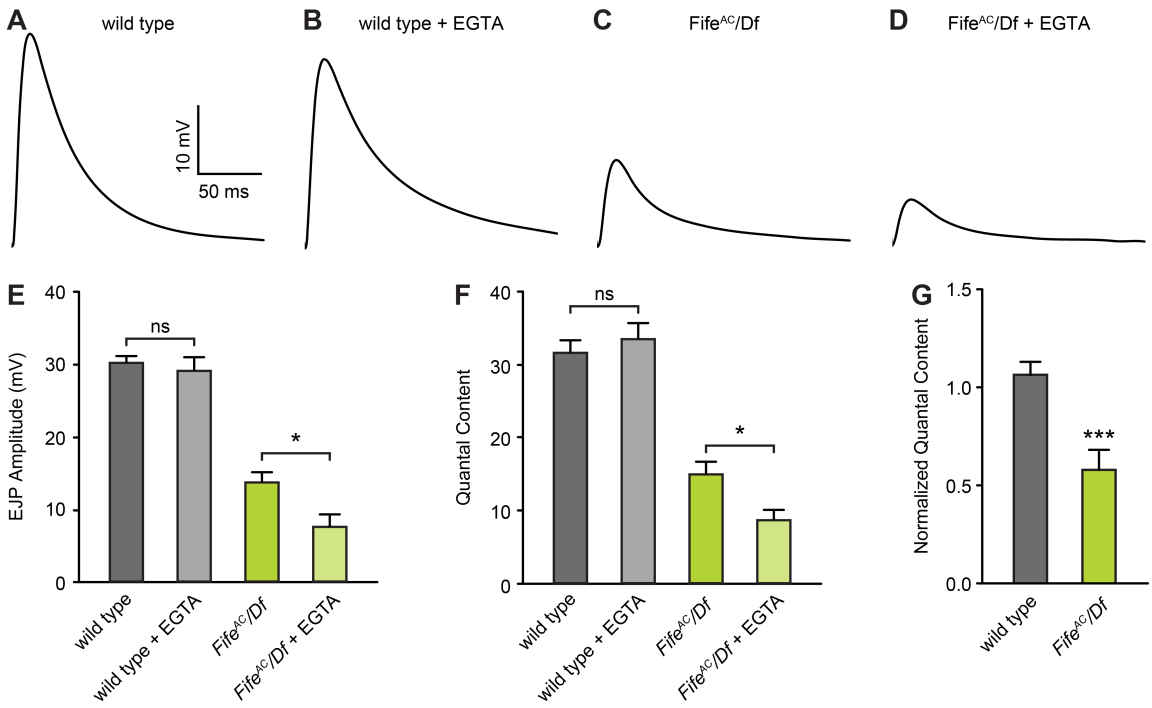

Figure 4 . Fife promotes coupling between calcium channels and release sites. (A-D) Representative traces of EJPs in wild type (A, B) and *Fife^{AC}/Fife Df* (C, D) in the absence of EGTA-AM (A, C) and with 10 min. pretreatment in 100 μM EGTA-AM (B, D) recorded in HL3 containing 0.6 mM extracellular Ca^{2+}. (E-F) Average EJP amplitude (E) and quantal content (F) are unaffected by exposure to EGTA-AM in wild type, but significantly affected in *Fife^{AC}/Fife Df*. (G) Quantal content of EGTA-treated NMJs normalized to untreated controls reveals a significant difference in quantal content in *Fife^{AC}/Fife Df* following EGTA exposure compared to wild type. wild type, n = 16 NMJs, wild type + EGTA, n = 12 NMJs, *Fife^{AC}/Fife Df,* n = 16 NMJs, *Fife^{AC}/Fife Df* + EGTA, n = 11 NMJs. Statistical significance represents comparison to control. ns, not significant; *p<0.05, ***p<0.001. Error bars represent s.e.m.

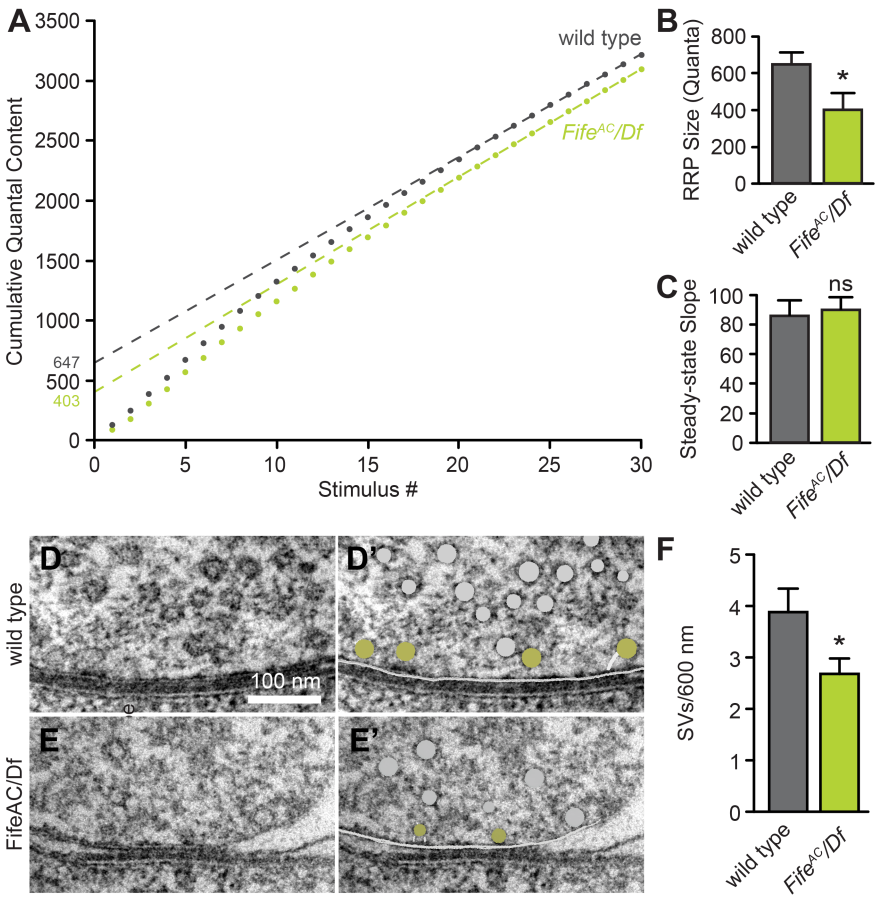

Figure 5. Fife regulates the number of readily releasable synaptic vesicles. **(A)** Average cumulative quantal content in wild type (grey) and *Fife^{AC}/Fife Df* (green). A 60-Hz stimulus train was applied in 1 mM Ca^{2+}. Back extrapolation to time zero of a line fit to the last 10 of 30 quantal content measurements, when neurotransmitter release has reached a steady state, yields an estimate of the initial size of the readily releasable pool. (B) Average RRP size is significantly decreased in *Fife^{AC}/Fife Df*). (C) The slope of the line fit to the steady-state phase is similar in wild type and *Fife^{AC}/Fife Df*, indicating normal vesicle recycling. wild type, n = 9 NMJs, *Fife^{AC}/Df*: n = 8 NMJs. (D-F) Morphological estimation of RRP size in electron micrographs of ultrathin sections of NMJs prepared by HPF/FS. Representative micrographs of wild type (D-D') and *Fife^{AC}/Fife Df* (E-E') active zones illustrate the population of active zone-associated synaptic vesicles (D', E', grey) that are in direct contact with or tethered within 5 nm of the presynaptic membrane (white) and represent the RRP (D', E', green). The number of RRP synaptic vesicles at each active zone was normalized to active zone length and is significantly reduced in *Fife^{AC}/Fife Df* (n = 25 active zones) compared to wild type (F, n = 30 active zones). Scale bar = 100 nm. ns, not significant; *p<0.05. Error bars represent s.e.m.

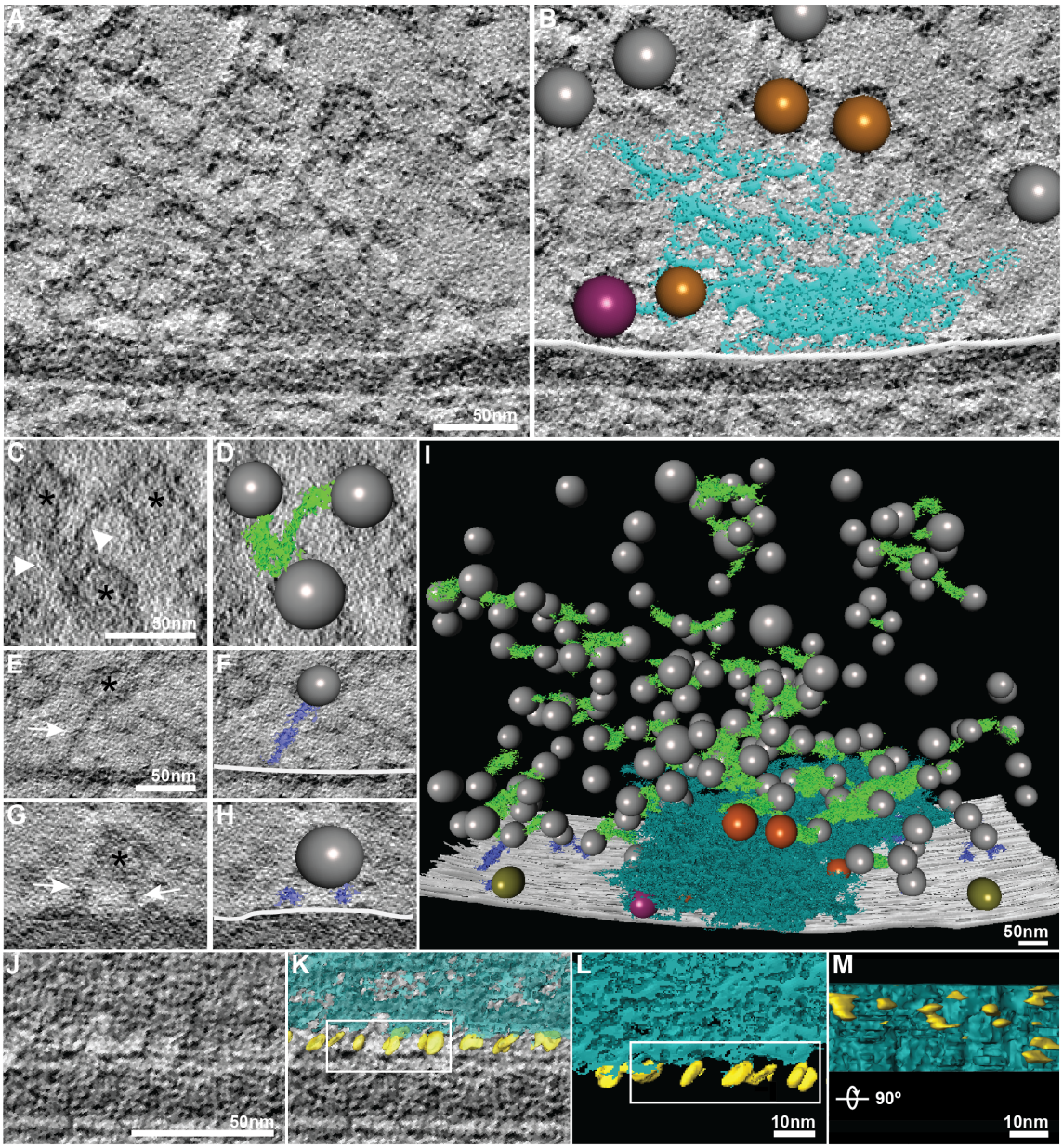

Figure 6. High-resolution architecture of the Drosophila NMJ active zone. **(A)** Single, central virtual slice size? through an EM tomogram of a wild type NMJ synapse prepared using HPF/FS techniques. (B) Segmented synaptic features, including the presynaptic dense projection (cyan), a single morphologically docked RRP synaptic vesicle (magenta), and clustered synaptic vesicles (gray) some of which make contact with the dense projection (orange). Scale bar: 50 nm. (C-D) Representative virtual slice (C) and segmentation (D) of the thin filaments (green) that link synaptic vesicles (gray) clustered at the active zone. Scale bar: 50 nm. (E-H) Representative virtual slices (E,G) and segmentation (F,H) of tethers of variable length (purple) between synaptic vesicles (gray) and the presynaptic membrane (white). Scale bar: 50 nm. (I) Representative fully segmented active zone, illustrating the dense projection (cyan), RRP (olive, or magenta when contacting the dense projection), dense projection-associated synaptic vesicles

(orange), synaptic vesicle pool (grey), filaments (green), and tethers (purple). Scale bar: 50 nm. (J-M) Representative virtual slice (J) and segmentation (K-M) of minute pegs (yellow) linking the dense projection (cyan) and presynaptic membrane. Pegs are regularly spaced (L) and vary in size (M). Scale bar in J: 50 nm. Scale bar in L and M: 10 nm.

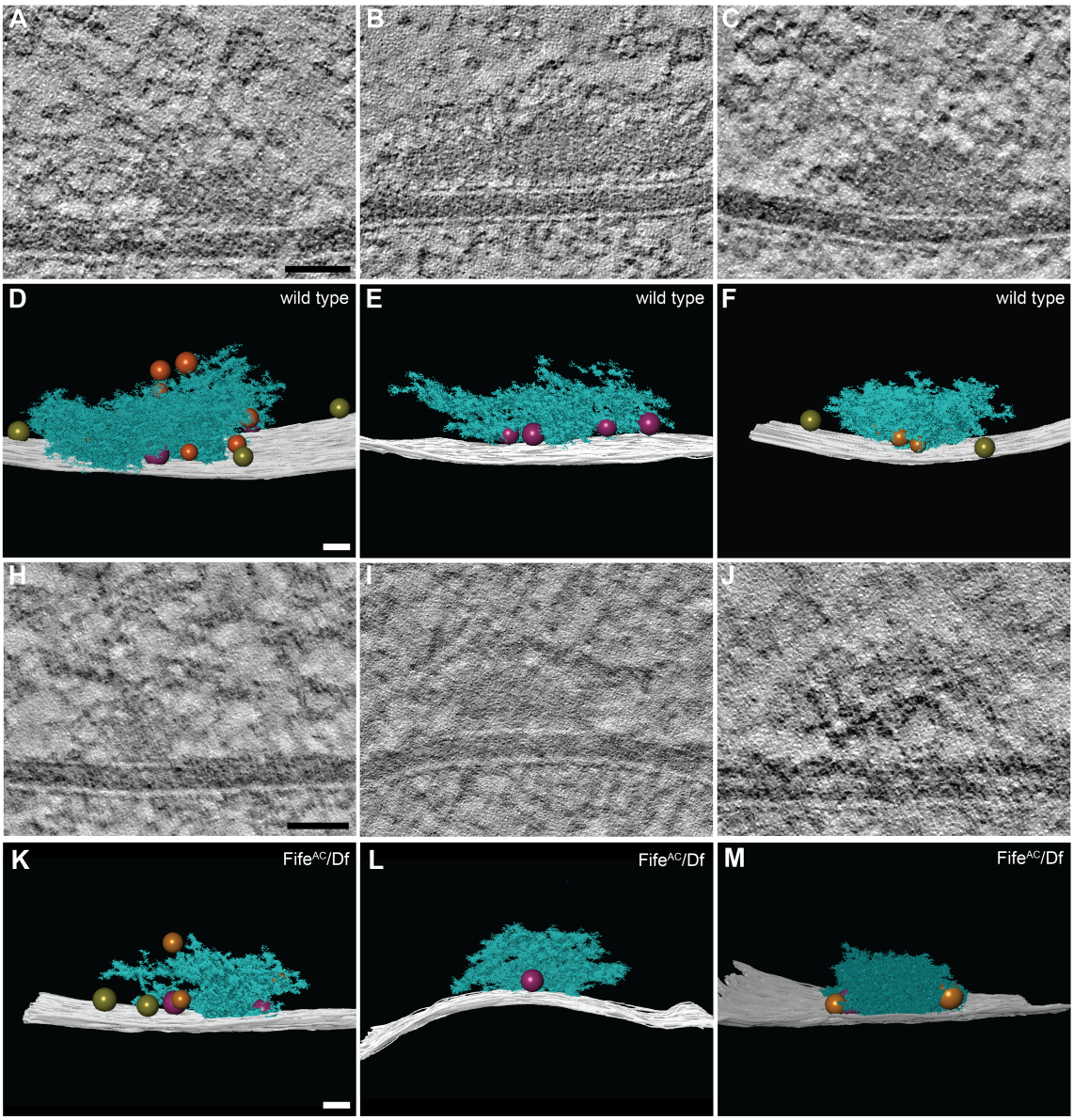

Figure 7. Electron tomographic reconstruction of *Fife* active zones.
Visualization of the RRP and dense projections of wild type and *Fife^{AC}/Fife Df* active zones in high-resolution 3D supports quantification of these features in single ultrathin slices. (A-F) Representative virtual slices (A-C) and segmentations (D-F) of dense projections, RRP synaptic vesicles (olive, magenta) and dense projection-associated synaptic vesicles (orange, magenta) at three wild type NMJ active zones from three distinct animals. (H-M) Three tomograms from three distinct *Fife^{AC}/Fife Df* animals were segmented in a similar manner and compared to wild type. Scale bar: 50 nm.

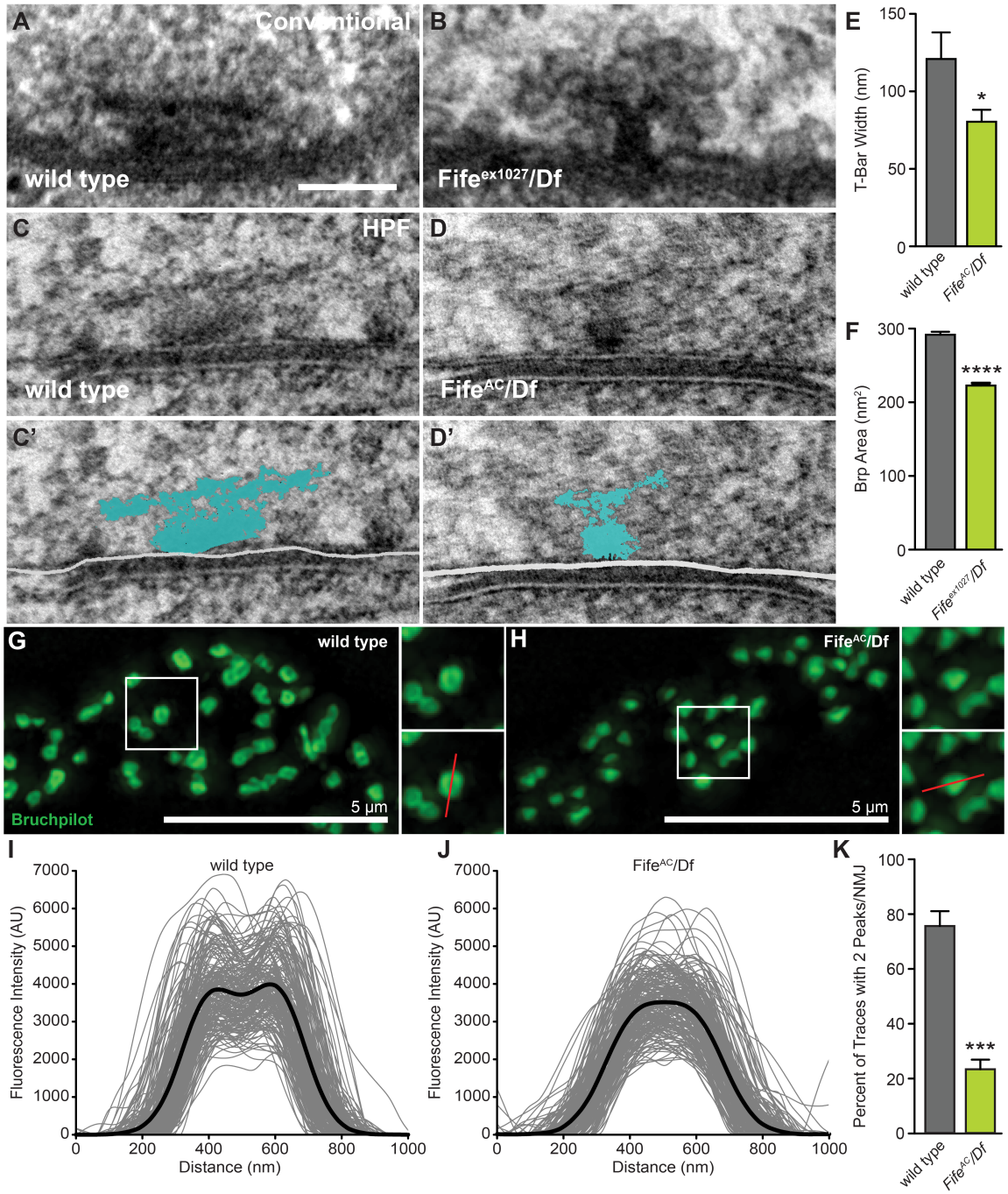

Figure 8. Fife regulates organization of the active zone cytomatrix. (A-E) Disruption of active zone ultrastructure is apparent in EM micrographs of ultrathin sections of wild type (A,C,C') and *Fife*^{AC}*/Fife Df* (B) or *Fife*^{ex1027}*/Fife Df* (D-D') active zones prepared by conventional preservation (A-B) or HPF/FS techniques (C-D'). The width of the T-bar in conventionally prepared samples is significantly reduced in *Fife*^{AC}*/Fife Df* (n = 17 active zones) compared to wild type (n = 13 active zones). Scale bar: 100 nm. (F) Furthermore, the area of Nc82 labelled Brp punctae is significantly reduced in *Fife*^{ex1027}*/Fife Df* (n = 1410 punctae from 4 NMJs) compared to wild type (n = 1479 punctae from 5 NMJs). (G-H) Superresolution structured illumination micrographs

of wild-type (G) and *Fife^{AC}/Fife Df* (H) NMJs stained with anti-Nc82 antibody. Insets show representative Brp punctae and an example line scan used to generate intensity profiles (red line). (I-J) Fluorescence intensity profiles along a line bisecting individual Brp punctae (grey lines) and the average plot profile (black line) are shown for wild-type (I, *n* =) and *Fife^{AC}/Fife Df* (J, *n* =) brp punctae. (K) Quantification of the percent of plot profiles indicating two peaks in wild-type (n = 10 NMJs) and *Fife^{AC}/Df* (n = 10 NMJs) NMJs. *p<0.05, ***p<0.001, ****p<0.0001. Error bars represent s.e.m.

References

Ackermann, F., Waites, C.L. & Garner, C.C., 2015. Presynaptic active zones in invertebrates and vertebrates. *EMBO reports*.

Aravamudan, B. et al., 1999. Drosophila UNC-13 is essential for synaptic transmission. *Nature neuroscience*, 2(11), pp.965–971.

Bailey, C.H., Kandel, E.R. & Harris, K.M., 2015. Structural Components of Synaptic Plasticity and Memory Consolidation. *Cold Spring Harbor perspectives in biology*, 7(7), p.a021758.

Betz, A. et al., 2001. Functional interaction of the active zone proteins Munc13-1 and RIM1 in synaptic vesicle priming. *Neuron*, 30(1), pp.183–196.

Bruckner, J.J. et al., 2012. Fife, a Drosophila Piccolo-RIM homolog, promotes active zone organization and neurotransmitter release. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(48), pp.17048–17058.

Bruckner, J.J., Zhan, H. & O'Connor-Giles, K.M., 2015. Advances in imaging ultrastructure yield new insights into presynaptic biology. *Frontiers in cellular neuroscience*, 9, pp.1–16.

Cao, Y.-Q. et al., 2004. Presynaptic Ca2+ channels compete for channel type-preferring slots in altered neurotransmission arising from Ca2+ channelopathy. *Neuron*, 43(3), pp.387–400.

Catterall, W.A. & Few, A.P., 2008. Calcium channel regulation and presynaptic plasticity. *Neuron*, 59(6), pp.882–901.

Chen, Z. et al., 2015. Ca2+ Channel to Synaptic Vesicle Distance Accounts for the Readily Releasable Pool Kinetics at a Functionally Mature Auditory Synapse. *Journal of Neuroscience*, 35(5), pp.2083–2100.

Eggermann, E. et al., 2012. Nanodomain coupling between Ca^{2+} channels and sensors of exocytosis at fast mammalian synapses. *Nature reviews Neuroscience*, 13(1), pp.7–21.

Fernández-Busnadiego, R. et al., 2013. Cryo-electron tomography reveals a critical role of RIM1α in synaptic vesicle tethering. *The Journal of Cell Biology*, 201(5), pp.725–740.

Fernández-Busnadiego, R. et al., 2010. Quantitative analysis of the native presynaptic cytomatrix by cryoelectron tomography. *The Journal of Cell Biology*, 188(1), pp.145–156.

Fouquet, W. et al., 2009. Maturation of active zone assembly by Drosophila Bruchpilot. *The Journal of cell biology*, 186(1), pp.129–145.

Fujimoto, K. et al., 2002. Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of cAMP-GEFII.Rim2. Piccolo complex in cAMP-dependent exocytosis. *The Journal of*

biological chemistry, 277(52), pp.50497–50502.

Gracheva, E.O. et al., 2010. Differential Regulation of Synaptic Vesicle Tethering and Docking by UNC-18 and TOM-1. *Frontiers in synaptic neuroscience*, 2, p.141.

Graf, E.R. et al., 2012. RIM Promotes Calcium Channel Accumulation at Active Zones of the Drosophila Neuromuscular Junction. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(47), pp.16586–16596.

Gratz, S.J. et al., 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, 196(4), pp.961–971.

Gundelfinger, E.D. & Fejtová, A., 2012. Molecular organization and plasticity of the cytomatrix at the active zone. *Current opinion in neurobiology*, 22(3), pp.423–430.

Harlow, M.L. et al., 2001. The architecture of active zone material at the frog's neuromuscular junction. *Nature*, 409(6819), pp.479–484.

Heidelberger, R. et al., 1994. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature*, 371(6497), pp.513–515.

Helmprobst, F., Frank, M. & Stigloher, C., 2015. The presynaptic architecture of the larval zebrafish neuromuscular junction. *The Journal of comparative neurology*, pp.n/a–n/a.

Holderith, N. et al., 2012. Release probability of hippocampal glutamatergic terminals scales with the size of the active zone. *Nature Publishing Group*, 15(7), pp.988–997.
Imig, C. et al., 2014. The morphological and molecular nature of synaptic vesicle priming at presynaptic active zones. *Neuron*, 84(2), pp.416–431.

Kittelmann, M. et al., 2013. Liprin-α/SYD-2 determines the size of dense projections in presynaptic active zones in C. elegans. *The Journal of Cell Biology*, 203(5), pp.849–863.

Kremer, J.R., Mastronarde, D.N. & MCINTOSH, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology*, 116(1), pp.71–76.

Landis, D.M. et al., 1988. The organization of cytoplasm at the presynaptic active zone of a central nervous system synapse. *Neuron*, 1(3), pp.201–209.

Lin, D.M. & Goodman, C.S., 1994. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron*, 13(3), pp.507–523.
Macleod, G.T., 2012. Direct injection of indicators for calcium imaging at the Drosophila larval neuromuscular junction. *Cold Spring Harbor protocols*, 2012(7), pp.797–801.

Liu, Karen S Y, Matthias Siebert, Sara Mertel, Elena Knoche, Stephanie Wegener, Carolin Wichmann, Tanja Matkovic, et al. 2011. “RIM-Binding Protein, a Central Part of the Active Zone, Is Essential for Neurotransmitter Release..” *Science (New York, N.Y.)* 334 (6062): 1565–69. doi:10.1126/science.1212991.

Macleod, G.T. et al., 2002. Fast calcium signals in Drosophila motor neuron terminals. *Journal of neurophysiology*, 88(5), pp.2659–2663.

Marrus, S.B. et al., 2004. Differential localization of glutamate receptor subunits at the Drosophila neuromuscular junction. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(6), pp.1406–1415.

Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology*, 152(1), pp.36–51. Matkovic, T. et al., 2013. The Bruchpilot cytomatrix determines the size of the readily releasable pool of synaptic vesicles. *The Journal of cell biology*, 202(4), pp.667–683.

Matz, J. et al., 2010. Rapid structural alterations of the active zone lead to sustained changes in neurotransmitter release. *Proceedings of the National Academy of Sciences*, 107(19), pp.8836–8841.

Michel, K. et al., 2015. The presynaptic active zone: A dynamic scaffold that regulates synaptic efficacy. *Experimental Cell Research*.

Mittelstaedt, T. & Schoch, S., 2007. Structure and evolution of RIM-BP genes: identification of a novel family member. *Gene*, 403(1-2), pp.70–79.

Müller, M. et al., 2012. RIM Controls Homeostatic Plasticity through Modulation of the Readily-Releasable Vesicle Pool. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(47), pp.16574–16585.

Neher, E. & Sakaba, T., 2008. Multiple roles of calcium ions in the regulation of neurotransmitter release. *Neuron*, 59(6), pp.861–872.

O'Connor-Giles, K.M., Ho, L.L. & Ganetzky, B., 2008. Nervous wreck interacts with thickveins and the endocytic machinery to attenuate retrograde BMP signaling during synaptic growth. *Neuron*, 58(4), pp.507–518.

Pinto, D. et al., 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304), pp.368–372.

Regehr, W.G., 2012. Short-term presynaptic plasticity. *Cold Spring Harbor perspectives in biology*, 4(7), p.a005702.

Rostaing, P. et al., 2006. Analysis of synaptic ultrastructure without fixative using high-pressure freezing and tomography. *The European journal of neuroscience*, 24(12), pp.3463–3474.

Schindelin, J. et al., 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), pp.676–682.

Sheng, J. et al., 2012. Calcium-channel number critically influences synaptic strength and plasticity at the active zone. *Nature Publishing Group*, 15(7), pp.998–1006.

Shibasaki, T. et al., 2004. Interaction of ATP sensor, cAMP sensor, Ca2+ sensor, and voltage-dependent Ca2+ channel in insulin granule exocytosis. *The Journal of biological chemistry*, 279(9), pp.7956–7961.

Siksou, L. et al., 2007. Three-dimensional architecture of presynaptic terminal cytomatrix. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(26), pp.6868–6877.

Wadel, K., Neher, E. & Sakaba, T., 2007. The coupling between synaptic vesicles and Ca2+ channels determines fast neurotransmitter release. *Neuron*, 53(4), pp.563–575.

Wagh, D.A. et al., 2006. Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in Drosophila. *Neuron*, 49(6), pp.833–844.

Wang, X. et al., 2009. A Protein Interaction Node at the Neurotransmitter Release Site: Domains of Aczonin/Piccolo, Bassoon, CAST, and Rim Converge on the N-Terminal Domain of Munc13-1. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(40), pp.12584–12596.

Wang, Y. & Südhof, T.C., 2003. Genomic definition of RIM proteins: evolutionary amplification of a family of synaptic regulatory proteins(small star, filled). *Genomics*, 81(2), pp.126–137.

Wang, Y. et al., 1997. Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion. *Nature*, 388(6642), pp.593–598.

Weimer, R.M. et al., 2006. UNC-13 and UNC-10/rim localize synaptic vesicles to specific membrane domains. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(31), pp.8040–8047.

Yang, Y. & Calakos, N., 2013. Presynaptic long-term plasticity. *Frontiers in synaptic neuroscience*, 5, p.8.

Zhai, R.G. & Bellen, H.J., 2004. The architecture of the active zone in the presynaptic nerve terminal. *Physiology (Bethesda, Md.)*, 19, pp.262–270.

Zhao, S. et al., 2012. Fine structure of hippocampal mossy fiber synapses following rapid high-pressure freezing. *Epilepsia*, 53(s1), pp.4–8.

Section 2

Developing the CRISPR-Cas9 system for rapid and reliable genome engineering in *Drosophila melanogaster*.

As detailed in Chapter 1, *Fife* was uncovered in an RNAi screen. After initial characterization of *Fife* using the RNAi line, we wanted to confirm the identified phenotypes in a classical allele. My subsequent efforts to generate a *Fife* null, a venture that proved much more difficult than I anticipated, sparked my interest in targeted genome engineering. Because we lacked both a readily screenable phenotype for a non-complementation-based mutagenesis strategy and a nearby P element for generating imprecise excisions, I used a homology-directed repair (HDR) strategy designed to generate a targeted deletion in the *Fife* locus. The HDR methods available for targeted mutagenesis in *Drosophila* at the time (2011-2012) relied on donor vectors with very long homology arms that had to be integrated into the genome and subsequently liberated and linearized to generate the double-stranded DNA breaks (DSBs) that could engage the DNA repair pathway to incorporate the desired modification. These methods were laborious and inefficient, likely due to the fact that there are only 2 donor copies per cell and the DSBs are generated in a small extrachromosomal donor molecule that is unlikely to engage the DNA repair system to the same extent a chromosomal DSB might. Ultimately, these attempts failed to generate the desired mutation despite 9 months of considerable effort. This failure piqued my interest in improving methods of genome engineering in *Drosophila*.

DSBs generated in the genomic target site are considerably more efficient at inducing HDR. To that end, zinc- finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) had been successfully employed in *Drosophila* to catalyze HDR. Unfortunately, the use of ZFNs and TALENs requires the generation of two unique sequence-specific nucleases for each application, which is not trivial. While I was attempting HDR, we were also undertaking an excision screen of a Minos element in hopes of generating a large enough deletion to create a null allele. Excision mutagenesis relies on aberrant repair of the DSB generated by the mobilization of a transposable element already inserted in the target region. Realizing that DSBs generated by mobilization of a transposon are functionally analogous to DSBs generated by a targeted nuclease, I started experimenting with the use of transposon mobilization to initiate targeted HDR – an exciting foray into work begun in the laboratory of Bill Engels beginning over 20 years ago. The approach, when combined with HDR techniques developed in the intervening years, appeared to significantly increase the efficiency of HDR, and I successfully deleted the Fife locus. While I was focusing on optimizing the strategy, news of a new method for generating targeted DSBs hit. It was immediately clear that this new approach would far surpass anything we could achieve with a transposon-based approach. Already keen on improving genome-engineering methods, we immediately switched gears and began working to apply the CRISPR-Cas9 system to *Drosophila*.

For both chapters 3 and 4, I participated in the design and analysis of all genome-engineering experiments. I designed and helped generate all the vectors and fly lines necessary for CRISPR-Cas9 genome engineering, screened for candidate mutants and performed molecular characterization experiments. I have also included a review I published about using the CRISPR-Cas9 system in *Drosophila* and a protocol chapter as appendices A and B respectively.

Chapter 3

Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease.

Scott J. Gratz[1], Alexander M. Cummings[2], Jennifer N. Nguyen[3], Danielle C. Hamm[4], Laura K. Donohue[2], Melissa M. Harrison[4,5*], Jill Wildonger[3*] and Kate M. O'Connor-Giles[1,2,6*]

[1] Genetics Training Program, University of Wisconsin-Madison, Madison, WI 53706

[2] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

[3] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[4] Integrated Program in Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[5] Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706

[6] Laboratory of Cell and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706

This chapter is adapted from the publication:

Gratz, Scott J, Alexander M Cummings, Jennifer N Nguyen, Danielle C Hamm, Laura K Donohue, Melissa M Harrison, Jill Wildonger, and Kate M O'Connor-Giles. 2013. "Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease." Genetics 194 (4): 1029–35. doi:10.1534/genetics.113.152710.

Abstract

We have adapted a bacterial CRISPR RNA/Cas9 system to precisely engineer the Drosophila genome and report that Cas9-mediated genomic modifications are efficiently transmitted through the germline. This RNA-guided Cas9 system can be rapidly programmed to generate targeted alleles for probing gene function in Drosophila.

Introduction

Genetic and molecular techniques to manipulate the genomes of model organisms are invaluable tools for understanding gene function. In Drosophila, chemical and insertional mutagenesis are powerful and widely utilized methods for disrupting gene function (ST JOHNSTON 2002; VENKEN and BELLEN 2005). Imprecise excision of a transposable element inserted near a gene of interest can result in deletion of all or part of the locus. More recently, techniques that stimulate homologous recombination (HR) using an exogenous template have made precisely targeted genome modifications possible in Drosophila (GLOOR *et al.* 1991; BANGA and BOYD 1992; NASSIF *et al.* 1994; RONG and GOLIC 2000; GONG and GOLIC 2003; HUANG *et al.* 2009). The most widely used methods rely on double-strand breaks (DSBs) in the donor template to trigger HR, but these can be time consuming and labor intensive (GAO *et al.* 2009; HUANG *et al.* 2009). Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been successfully employed in Drosophila to generate DSBs in genomic targets that trigger repair by either HR or error-prone non-homologous end joining (NHEJ) (BIBIKOVA *et al.* 2002; LIU *et al.* 2012). However, ZFNs and TALENs, which both comprise a nuclease joined to a site-specific DNA: binding domain, require the generation of a unique protein for each genomic manipulation (WOOD *et al.* 2011). In the past few months, the simplified CRISPR RNA-guided Cas9 nuclease has shown broad potential for genome engineering in metazoans.

CRISPRs (clustered regularly interspaced short palindromic repeats) and the CRISPR-associated nuclease Cas9 were first identified in *E. coli* in 1987 and later shown to function as part of an adaptive immune system in bacteria and archaea (ISHINO *et al.* 1987; MAKAROVA *et al.* 2006; BARRANGOU *et al.* 2007). In type II CRISPR systems, a CRISPR RNA (crRNA), which contains sequence complementary to invading virus or

plasmid DNA, and a trans-activating CRISPR RNA (tracrRNA) interact with Cas9 to direct sequence-specific cleavage of exogenous DNA. Recognizing the potential of harnessing this system for precise genome engineering in other organisms, Jinek and colleagues identified a minimal two-component system required for the site-specific cleavage of DNA – Cas9 and a chimeric RNA (chiRNA) comprising the crRNA and tracrRNA from *S. pyogenes* (JINEK *et al.* 2012). Thus, in contrast to ZFNs and TALENs, this modified CRISPR RNA/Cas9 system directs a common nuclease to specific DNA sequences by a short, readily generated RNA. This obviates the need to create a unique chimeric nuclease for each genome manipulation and raises the possibility and promise of routine genome engineering.

In recent months, a number of laboratories have demonstrated that this heterologously expressed two-component CRISPR RNA/Cas9 system can induce site-specific DSBs in eukaryotic genomes, including cultured mammalian cells, human stem cells and yeast (CHO *et al.* 2013; CONG *et al.* 2013; DICARLO *et al.* 2013; DING *et al.* 2013; JINEK *et al.* 2013; MALI *et al.* 2013). In vivo experiments in zebrafish and mice yielded mosaic animals, demonstrating that the system has the capacity to manipulate a variety of eukaryotic genomes (CHANG *et al.* 2013; HWANG *et al.* 2013; SHEN *et al.* 2013). Most recently, mice carrying homozygous mutations in two genes were generated through embryo injections (WANG *et al.* 2013). Nonetheless, to date the CRISPR RNA/Cas9 system has not been adapted for Drosophila, a key genetic model, and germline transmission of Cas9-induced changes has not been demonstrated in any organism. Here we demonstrate that the CRISPR RNA/Cas9 system can (1) mediate efficient genome engineering in Drosophila and (2) induce targeted genome modifications in the Drosophila germline that are transmitted to progeny.

Materials and methods

Molecular reagents

phsp70-Cas9: Sequence encoding 3X Flag-NLS-Cas9-NLS was amplified from pX330 (CONG *et al.* 2013) and cloned as a ClaI/XbaI fragment between the Drosophila *hsp70* promoter and 3′ UTR in pHSS6hsILMi20 (a gift from Anastasios Pavlopoulos). For annotated sequence, see Supporting Information Figure S2A. Expression of Cas9 in embryos was confirmed by immunoblotting.

pU6-BbsI-chiRNA: The chiRNA sequence was placed under control of the Drosophila snRNA:U6:96Ab promoter for in vivo transcription (WAKIYAMA *et al.* 2005). The U6-chiRNA backbone was synthesized as a gBlock gene fragment (Integrated DNA Technologies) and blunt-end ligated into the EcoRV site of pBluescript SK(-). The resulting vector contains two BbsI cut sites to facilitate insertion of target-specific sequences (for annotated sequence, see Supporting Information Figure S2B). Target-specific sequences for *yellow* and *rosy* were synthesized as 5′-phosphorylated oligonucleotides, annealed and ligated into the BbsI sites of pU6-BbsI-chimericRNA.

Donor template. The single-stranded DNA oligonucleotide (ssODN) donor template for homologous recombination was designed to contain 60-nt of homology directly adjacent to each Cas9-mediated DSB in the target locus and a 50-nt attP docking site (for annotated sequence, see Supporting Information, Figure S2C). The ssODN was synthesized by Integrated DNA Technologies.

Embryo injections

w^{1118} preblastoderm embryos were injected through the chorion membrane using standard protocols. phsp70-Cas9 was injected at a concentration of 500 ng/μL. The

pU6-chiRNA targeting constructs were injected at 500 ng/μL for single chiRNAs and 250 ng/μL each when two chiRNAs were injected. The donor template ssODN was diluted according to the manufacturer's recommendations and injected at a concentration of 100 ng/μL. All injection mixtures were prepared in water. Average embryonic survival following injection with Cas9 and a single chiRNA was 50%. Embryonic survival rates following multiplex injections of 2 chiRNAs with and without the ssODN donor were 68 and 69%, respectively. While this difference likely reflects improved quality of injections rather than the differences in the injection components, these rates indicate that expression of the components of the CRISPR RNA/Cas9 system does not significantly impair development.

Surveyor assay

Analysis of NHEJ products resulting from single chiRNA targeting was performed using the Surveyor Mutation Detection kit (Transgenomic). Briefly, genomic DNA was isolated from individual embryos 24 hours after injection. Approximately 500-bp flanking the target Cas9 cut sites in *yellow* and *rosy* were amplified using Herculase DNA polymerase (Agilent Technologies) according to the manufacturer's protocol and the resulting product was purified using the Wizard SV Gel and PCR Clean-Up System (Promega). A total of 500 ng of purified PCR product was diluted in 1X Herculase reaction buffer to a final volume of 20 uL. Heteroduplexes were formed using the following parameters: 95°C for 10 min, 95°C to 85°C ramping at rate of -2.0°C/sec, 85°C for 1 min, 85°C to 25°C at a rate of -0.3°C/second with 1 min holds at 75°C, 65°C, 55°C, 45°C, 35°C, and 25°C. After this duplex annealing step, 16 μL of each sample (400 ng annealed duplexes) was mixed with 2 μL 0.15 M $MgCl_2$, 1 μL Surveyor Enhancer S, 1 μL Surveyor Nuclease S and incubated at 42°C for 60 min. The Surveyor reaction was stopped with 2 μL of Stop solution, and the products were separated by polyacrylamide gel electrophoresis. The gel

was stained with SYBR Gold (Invitrogen), and visualized on a GE ImageQuant.

Molecular characterization of engineered loci

Genomic DNA was isolated from individual embryos 24 hours after injection. PCR was performed using primers flanking the target lesion (Figure 1A, open arrows). Amplified products were purified and analyzed by Sanger sequencing or subcloned into pCR4-TOPO or pCRBluntII-TOPO (Invitrogen) prior to Sanger sequencing.

Screening

To assess germline transmission of targeted modifications, adults that developed from injected embryos were individually crossed to y^1, w^1. The offspring of crosses were screened for 10 days after the first flies emerged for progeny with yellow cuticles. Transmission rates were calculated both as a percent of parental crosses that produced one or more yellow progeny and as a percent of total progeny. The total number of progeny screened was calculated by weight. Specifically, for each genome manipulation we pooled progeny daily and weighed 100 flies. The remaining flies were weighed and their number calculated based on the weight of 100 flies. Transmission of expected events was confirmed by sequence analysis.

Results

Site-directed cleavage of Drosophila genes

To test whether the CRISPR RNA/Cas9 system could induce site-specific DSBs in Drosophila, we targeted the *yellow* gene, which is on the X chromosome and commonly used for genome engineering studies (RONG and GOLIC 2000; BIBIKOVA *et al.* 2002; LIU *et al.* 2012). Efficient target recognition by the CRISPR RNA/Cas9 system requires 20 nucleotides (nt) of homology between the chiRNA and its genomic target. Cleavage also requires that the 3′ end of the genomic target sequence contains a 3-basepair (bp) proto-spacer adjacent motif (PAM) sequence, NGG, which differentiates self from invading DNA in the endogenous system (JINEK *et al.* 2012). Thus, selection of a 20-nt target sequence is limited only by the requirement for an adjacent PAM sequence. In our plasmid-based system, the ideal target sequence also begins with a G that allows for precise transcriptional initiation of the chiRNA from the U6 promoter (WAKIYAMA *et al.* 2005)(see Materials and Methods).

Plasmids encoding Cas9 and a chiRNA targeting the first exon of *yellow* (YE1), disruption of which is expected to abolish *yellow* function, were co-injected into preblastoderm embryos (Figures 1A, S1A,B). To assess Cas9-mediated cleavage, we isolated DNA from individual embryos 24-hours post injection. Following PCR amplification of the targeted genomic site, we used the Surveyor Mutation Detection assay, which employs an endonuclease that recognizes and cuts at the site of mismatches, to detect small insertions or deletions (indels) that result when DSBs are imperfectly repaired by NHEJ. DSBs that are repaired by HR using either the sister chromatid or homologous chromosome as a template do not result in an altered locus and are not detectable. The Surveyor assay indicated site-specific indels in one of 13 embryos injected with Cas9 and the YE1 chiRNA (Figure 1B). Consistent with the loss of *yellow* in

a subset of somatic cells, we observed mosaic yellow patches in adult male cuticles after YE1 injection (6%, n=67). Based on the size of the cleavage products produced in the Surveyor assay, the induced cleavage occurred at the YE1 target site indicating that the chiRNA successfully directed Cas9 to the *yellow* locus where it induced DSBs in the targeted DNA.

To determine if co-injection of Cas9 and a chiRNA could generate indels at other loci, we designed two chiRNAs to target the *rosy* gene (Figure 1C, R1 and R2). Again, the Surveyor assay identified Cas9-induced site-specific cleavage at the targeted sites in a subset of embryos (Figure 1D, 3/4 embryos for each chiRNA). Thus, the RNA-guided Cas9 system can be readily programmed to induce DSBs in multiple Drosophila genomic targets.

Engineering defined deletions

Encouraged that the CRISPR RNA/Cas9 system functioned in Drosophila, we next asked whether Cas9-mediated DSBs could be harnessed to generate defined deletions. We reasoned that introducing two chiRNAs, Y5′ targeting the 5′ end of *yellow* and Y3′ targeting the 3′ end, might guide Cas9 to precisely delete the 4.6-kilobase (kb) *yellow* locus (Figure 1A). We isolated DNA from embryos 24 hours after co-injection of Cas9, Y5′ chiRNA and Y3′ chiRNA, and performed PCR to determine whether *yellow* was deleted. Using primers flanking the targeted sites (Figure 1A, open arrows), a 650-bp product was amplified as expected if targeted deletion occurred in 13/16 embryos (Figure 1E). Moreover, as adults, 66% of injected males (n= 52) displayed yellow mosaicism, consistent with a high level of targeted deletion in somatic cells.

To investigate individual repair events, the DNA amplified from the targeted locus was pooled, subcloned and sequenced. In nine of nine clones analyzed, the 4.6-kb *yellow* locus was deleted and the junction created by repair of the two cut sites showed the small indels typical of NHEJ (Figure 1F). These results extend findings from mammalian cell culture in which 19- and 118-bp sequences were deleted by co-transfection of Cas9 and two chiRNAs (CONG *et al.* 2013; MALI *et al.* 2013), and demonstrate that such "multiplex" genome editing is an effective tool for precisely deleting large genomic regions in organisms.

Cas9-mediated homologous recombination

The ability to edit genomic sequence by introducing specific mutations or exogenous sequences to monitor or manipulate protein activity provides a critical tool for dissecting gene function. One means of achieving this goal is to replace a gene with a sequence that enables site-specific integration of engineered genes into the targeted locus, such as the attP ΦC31 phage recombination site. Such gene replacement can be accomplished by stimulating HR in the presence of a donor template that includes the site-specific integration sequence (GAO *et al.* 2008; HUANG *et al.* 2009). To this end, we designed a single-stranded oligodeoxynucleotide (ssODN) repair template that includes a 50-nt attP site flanked by 60-nt homology arms corresponding to sequences 5′ and 3′ of *yellow* (Figure 1G and S1C). We co-injected Cas9 and the Y5′ and Y3′ chiRNAs with this ssODN, which we predicted would serve as a template for repair of the deleted locus by HR resulting in the replacement of *yellow* with attP. Using the same primers as above (Figure 1A, open arrows), we performed PCR on DNA extracted from single injected embryos. In 13/16 embryos, we detected the 700-bp product that is expected to result from replacement of *yellow* with attP (Figure 1H). In agreement with highly efficient disruption of *yellow*, 62% of injected males exhibited a mosaic cuticle as adults (n=35).

To confirm that the ssODN donor served as a template for repair by HR, the PCR products amplified from injected embryos were pooled and subcloned. Sequence analysis of individual repair events revealed that the *yellow* gene was replaced by attP from the ssODN template in five of eight clones. Two clones contained the 4.6-kb deletion but lacked the attP sequence, indicating site-specific induction of DSBs repaired by NHEJ rather than HR. This mirrors our PCR results, in which doublets of approximately 700 and 650-bp were detected, likely representing the inclusion or lack of the 50-bp attP sequence (Figure 1H, lanes 3, 10 and 11). The higher molecular weight bands faintly visible in some lanes may reflect aberrant repair events (Figure 1H, lanes 11 and 15). We observed such an event in one of eight clones. In this clone, the 4.6-kb *yellow* locus was deleted and part of the attP sequence was incorporated along with 40-bp of novel sequence. This unusual event may reflect repair via a synthesis-dependent strand annealing (SDSA) mechanism in which the two broken DNA strands independently select repair templates (NASSIF *et al.* 1994). Previous studies of HR induced by DSBs in genomic targets have also identified SDSA as a primary mechanism of repair and reported similar rare events (NASSIF *et al.* 1994; BOZAS *et al.* 2009).

Germline transmission of Cas9-induced genome modifications

Our results demonstrate that the readily programmed CRISPR RNA/Cas9 system can be employed to precisely mutate, delete or replace genes in vivo, so we next sought to determine whether Cas9-induced mutations could be transmitted through the germline to create stable mutant strains. We crossed adult flies injected as embryos with Cas9 and YE1 chiRNA to flies carrying a null mutation in *yellow* (y^1), allowing us to phenotypically screen female progeny of injected males and all progeny of injected females. 6.4% of injected males and 4.8% of injected females yielded at least one offspring exhibiting the

yellow null phenotype (Table 1, injected flies that yield yellow offspring are referred to as 'founders'). To determine if yellow progeny resulted from Cas9-induced mutations, the target locus was amplified and sequenced from at least one yellow progeny per founder. Indels were present at the predicted Cas9 cut site in each case (Figure 2A). Interestingly, we recovered a 4-bp deletion in the progeny of multiple founders. This is likely due to preferential repair by microhomology-mediated end joining facilitated by the presence of a short repeat adjacent to the YE1 cut site (Figure 2A, underlined in wild-type sequence) (MCVEY and LEE 2008). Thus, the rate of mutations induced by Cas9 and a single chiRNA is comparable to that obtained with a pair of ZFNs targeting the *yellow* locus (6% transmission for males and 0% for females) and lower than that reported for two TALENs (17% for males and 45% for females) (BIBIKOVA *et al.* 2002; LIU *et al.* 2012).

We also investigated whether multiplex-mediated genome editing was transmitted through the germline. We first analyzed the progeny of flies that had been injected with Cas9, Y5′ chiRNA and Y3′ chiRNA to precisely delete *yellow*. 25% of injected males and 14.3% of injected females produced yellow progeny (Table 1, founders). PCR and sequence analysis of mutant progeny identified the precise deletion of *yellow* in the progeny of one founder (Figure 2B and Table 1). Partial deletions of *yellow*, up to ~3.7-kb in size, were also detected by PCR and sequence analysis, consistent with imperfect repair of Cas9-induced deletions (Figure 2B).

We next analyzed flies that had been injected with Cas9, Y5′ chiRNA, Y3′ chiRNA and the ssODN donor template to determine if the replacement of *yellow* with attP was also transmitted to progeny. 8.6% of injected males and 23.1% of injected females produced one or more yellow offspring (Table 1, founders). Sequence analysis showed that two (22%) of the founders transmitted the precisely engineered replacement of *yellow* with

attP (Figure 2C and Table 1). By comparison, depending on the specific donor, 1-19% of mutant progeny of flies injected with ZFNs and ssODN donor templates designed to introduce stop mutations in a *rosy* exon reflected HR-mediated repair (BEUMER *et al.* 2013). These data demonstrate that Cas9-mediated HR events are effectively transmitted through the germline.

A common concern with sequence-guided nucleases is the potential for off-target cleavage. Cong et al. (2013) demonstrated that cleavage by Cas9 requires a PAM and a perfect match of at least 12-nt at the 3′ end of the targeting sequence, while single mismatches in the remaining 8-nt of targeting sequence are tolerated. Given the high frequency at which PAM sites occur in the Drosophila genome, we were able to readily select targets with minimal potential for off-target cleavage. Of the three chiRNAs used to generate germline transformants, only the Y5′ chiRNA has a 12-nt match to sequence adjacent to a PAM elsewhere in the genome. PCR and sequence analysis of this site in flies carrying Y5′-mediated deletion or replacement of *yellow* demonstrated the presence of wild-type sequence, consistent with the absence of off-target cleavage (Figure S3). Moreover, engineered flies developed into viable and fertile adults. The lack of detectable off-target effects is in accordance with findings in other systems and suggests high specificity of targeting by the CRISPR RNA/Cas9 system (WANG *et al.* 2013).

In summary, overall germline transmission rates, defined as the percent of injected flies that yielded at least one offspring carrying the targeted event, ranged from 1.1% for multiplex-mediated deletion of *yellow* to 5.9% for single chiRNA-induced NHEJ. Perfect replacement of *yellow* with attP was transmitted through the germline of 3.3% of injected flies (2/61). Importantly, transformed flies were healthy and fertile and showed no evidence of off-target cleavage.

Discussion

Our results demonstrate the efficient generation of Cas9-mediated genome modifications in Drosophila, and for the first time show germline transmission of genome engineering via the CRISPR RNA/Cas9 system. A single chiRNA can guide Cas9 to a specific genomic sequence to induce DSBs that are imperfectly repaired by NHEJ, while multiplex targeting can be used to generate large defined deletions. By combining multiplex targeting of *yellow* with the introduction of an ssODN donor template containing exogenous sequence, we successfully replaced *yellow* with an attP docking site. Interestingly, while germline transmission rates for each of these manipulations were similar, injection of two chiRNAs (with or without the ssODN donor) yielded approximately 10 times more mosaic males than a single chiRNA. This may be due to the fact that multiplex injections provide multiple routes for mutating the target locus, including single DSBs induced by either chiRNA and repaired by NHEJ. The lower percent of founders yielding the precise targeted event in the multiplex injection paradigms supports this interpretation. Our results also suggest that the frequency of events in somatic tissue does not directly correlate with the frequency of events in the germline, consistent with previous observations with ZFNs and TALENs in Drosophila (BIBIKOVA *et al.* 2002; LIU *et al.* 2012).

It is likely that the CRISPR RNA/Cas9 system can be further optimized. Studies in vertebrates suggest that expression levels of CRISPR RNAs are a key determinant of efficiency (JINEK *et al.* 2013). Transgenic lines expressing Cas9 and tracrRNA under the control of endogenous germline-specific promoters might yield increased efficiency. Moreover, it may be possible to bias repair to HR over NHEJ by exploiting a Cas9 mutant that induces single-strand DNA breaks (JINEK *et al.* 2012; CONG *et al.* 2013; MALI *et al.* 2013) or conducting Cas9-mediated genome engineering in a genetic background

that reduces NHEJ (BEUMER *et al.* 2008). Because the components of the CRISPR RNA/Cas9 system can be injected into embryos of any genotype, it should also be possible to use a marked transposable element residing in the targeted locus to identify defined deletions or replacements by loss of the positive marker, obviating the need for molecular screening when screening by mutant phenotype is not possible. Finally, combining the CRISPR RNA/Cas9 system with dsDNA donors containing larger homology regions may also increase efficiency (NASSIF *et al.* 1994; BEUMER *et al.* 2013). While the generation of such templates is more labor intensive, the ability to incorporate a selectable marker would add to the utility of the CRISPR RNA/Cas9 system for some applications.

The ease of producing sequence-specific chiRNAs makes the CRISPR RNA/Cas9 system an appealing method for genome editing. From the initial cloning steps to identifying transformants, stable lines with targeted genome alterations can be generated within a month. Our demonstration that the CRISPR RNA/Cas9 system can be used to create and transmit precise genomic modifications in Drosophila opens the door to rapid engineering of the genome to investigate gene function and regulation.

Acknowledgements

We thank the Bloomington Drosophila Stock Center for fly stocks and Michael Cox, Carlos Flores, Judith Kimble and Andrew Mehle for invaluable discussions and critical comments on the manuscript. We are grateful to Feng Zhang for sharing reagents and an unpublished manuscript with us, Kevin Eliceiri for sharing injection equipment and Kathy Vaccaro for sharing her expertise on embryo injections. This work was funded by startup funds from the University of Wisconsin to MMH, JW and KOCG and grants from the National Institute of Neurological Disorders and Stroke, National Institutes of

Health to JW (R00 NS072252) and KOCG (R00 NS060985). All plasmids used in this study will be made available to academic laboratories through the non-profit distributor Addgene along with detailed guidance at FlyCRISPR.molbio.wisc.edu.

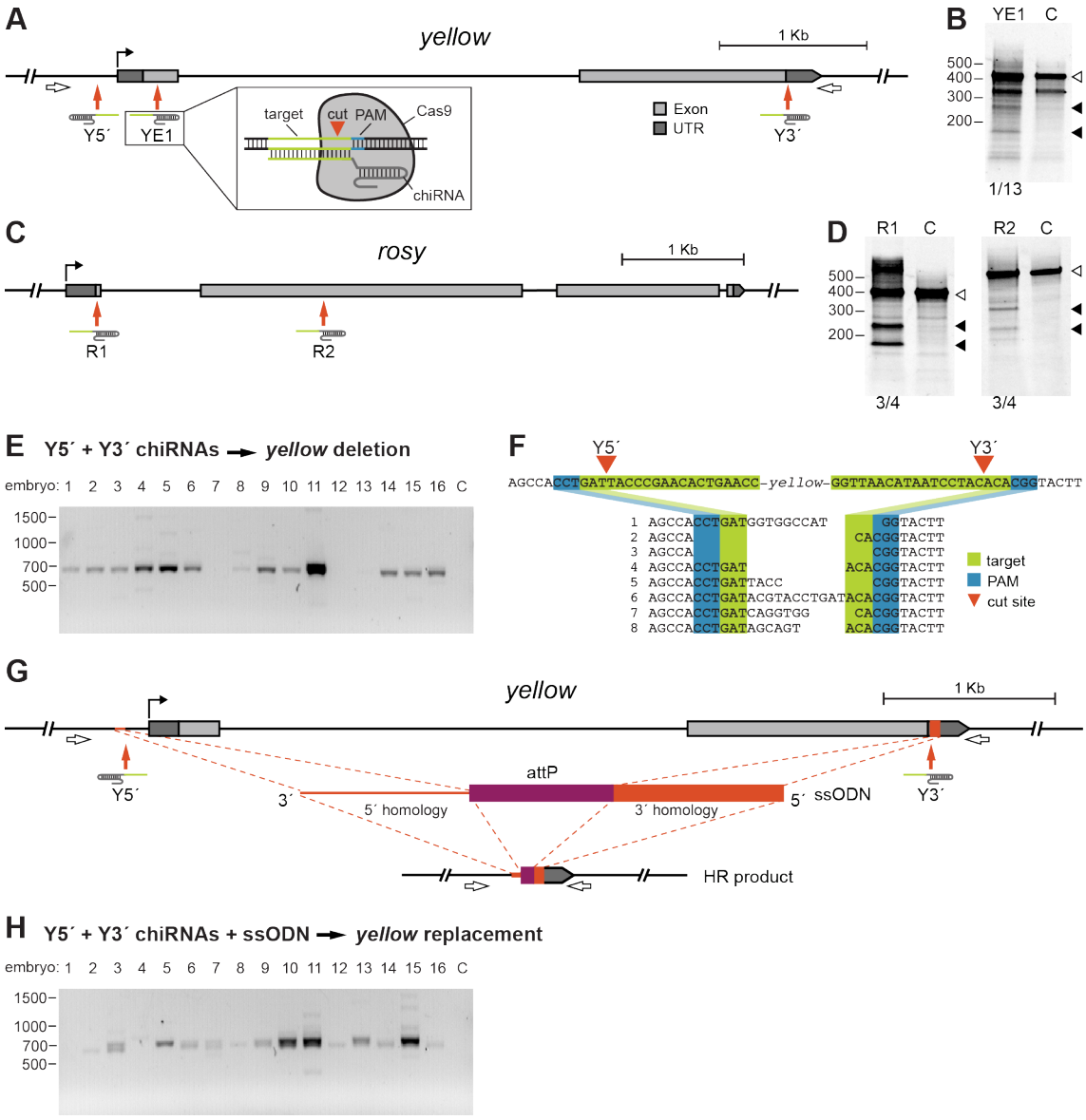

Figure 1. Cas9-generated DSBs are repaired by NHEJ and HR in Drosophila embryos. (A) Schematic of the *yellow* locus indicating Y1E, Y5′ and Y3′ chiRNA target sites. The predicted cut site for each chiRNA is indicated (red arrows). Inset shows schematic of CRISPR RNA/Cas9 system components interacting with target DNA. For additional detail, see Supporting Information, Figure S1A. (B) Results of the Surveyor assay indicate targeted cleavage occurred in *yellow* at the predicted YE1 cut site (left: embryos injected with YE1 chiRNA, right: control, C, embryos). Duplexes lacking indels were uncut (open arrowhead), whereas duplexes containing indels were cut asymmetrically at the targeted site (closed arrowheads). 1/13 embryos injected with YE1 chiRNA produced duplexes with indels. (C) Schematic of the *rosy* locus indicating R1 and R2 chiRNA target sites (red arrows). (D) Results from Surveyor assays indicate R1 and R2 chiRNAs generated indels at the targeted cut sites (embryos injected with R1 chiRNA, R2 chiRNA or control, C, are as indicated). Uncut duplexes (open arrowhead) and cut duplexes (closed arrowhead) as indicated. 3/4 embryos injected with either R1 chiRNA or R2 chiRNA produced duplexes with indels. (E) The *yellow* locus was

amplified from embryos injected with Cas9, Y5′ chiRNA and Y3′ chiRNA using the primers indicated in (A) (open arrows). Products from 9 reactions were pooled for sequencing. Targeted deletions result in a 650-bp PCR product (C = uninjected control embryo). (F) Sequence alignments of the targeted deletion PCR products reveal breakpoints and imprecise repair at predicted cut sites (red arrowheads). One repair event that resulted in the 4.6-kb deletion of *yellow* plus an indel deleting an additional 27-bp and inserting 93-bp is not shown. (G) Schematic of the *yellow* ssODN design. The ssODN includes 60-nt homology arms flanking a 50-nt attP docking site (purple). The homology arms correspond to the sequences immediately adjacent to the predicted cut sites (red arrows). For additional detail see Supplemental Information, Figure S1C. (H) The *yellow* locus was amplified from embryos injected with Cas9, Y5′ chiRNA, Y3′ chiRNA and ssODN using the primers indicated in (A). HR events in which the locus is deleted and replaced with attP will yield a 700-bp product. Products from 3 reactions were pooled for sequence analysis, which confirmed the replacement of *yellow* with attP in 5/8 clones.

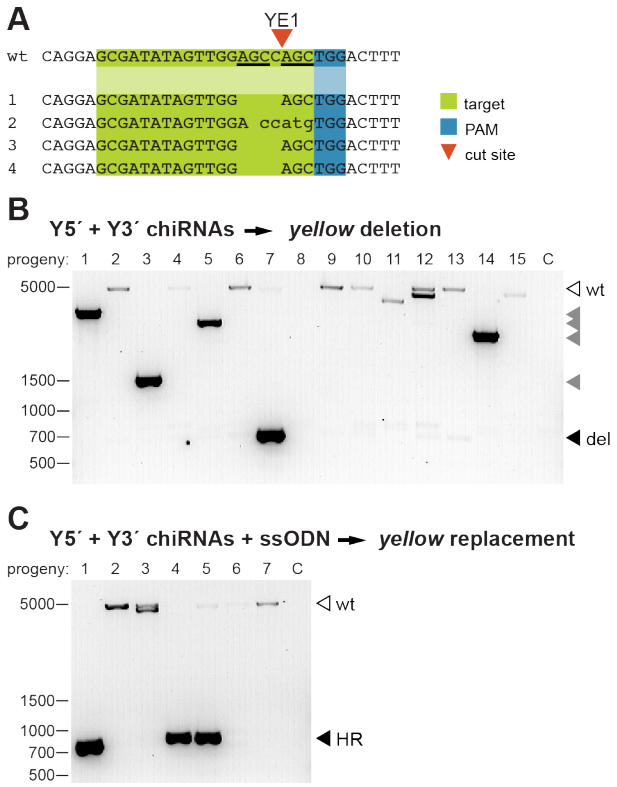

Figure 2. CRISPR-induced modifications are efficiently transmitted through the germline. (A) Representative sequences of independent *yellow* mutants generated by Y1E-guided Cas9-induced DSBs. The predicted YE1 cleavage site is indicated with a red arrowhead. Short sequence repeats flanking the cleavage site (underlined in the wild-type sequence) may have facilitated the preferential generation of the same 4-bp deletion in three independent founders by MMEJ. (B) The *yellow* locus was amplified from yellow progeny of males and females injected with Cas9, Y5′ chiRNA and Y3′ chiRNA using the primers indicated in Figure 1A. 15 independent progeny are represented (C = uninjected control). A 650-bp PCR product is expected (black arrowhead) if a targeted deletion of *yellow* has been transmitted. PCR products corresponding to partial deletions (grey arrowheads) and wild-type (open arrowhead) *yellow* are indicated. (C) The *yellow* locus was amplified from embryos injected with Cas9, Y5′ chiRNA, Y3′ chiRNA and ssODN using the primers indicated in Figure 1A. Seven independent progeny are represented (C= uninjected control). A 700-bp product is expected following the targeted replacement of *yellow* with attP. PCR products corresponding to correct HR events are indicated (black arrowheads) and were confirmed by Sanger sequencing.

Table 1. Germline transmission rates

chiRNA(s)	ssODN donor	Injected males		Injected females		% (#) founders yielding targeted event	% (#) overall germline transmission
		% (#) founders	% (#) progeny	% (#) founders	% (#) progeny		
YE1	-	6.4 (3/47)	0.23 (5/2128)	4.8 (1/21)	0.26 (4/1525)	100 (4/4)	5.9 (4/68)
Y5', Y3'	-	25 (13/52)	1.3 (61/4608)	14.3 (5/35)	1.5 (35/2421)	5.6 (1/18)	1.1 (1/87)
Y5', Y3'	+	8.6 (3/35)	1.0 (24/2336)	23.1 (6/26)	1.3 (34/2655)	22.2 (2/9)	3.3 (2/61)

Flies injected with Cas9 and the indicated chiRNAs and ssODN template were crossed to *yellow* (y^1) and progeny screened for yellow cuticles. The percent of injected flies producing one or more yellow progeny (founders) is indicated along with the percent of total progeny exhibiting yellow cuticle. At least one progeny per founder was sequenced to determine if the targeted event had occurred. The percent of founders in which the expected event occurred in one or more progeny is reported, as is the overall germline transmission rate (% injected flies yielding expected event)

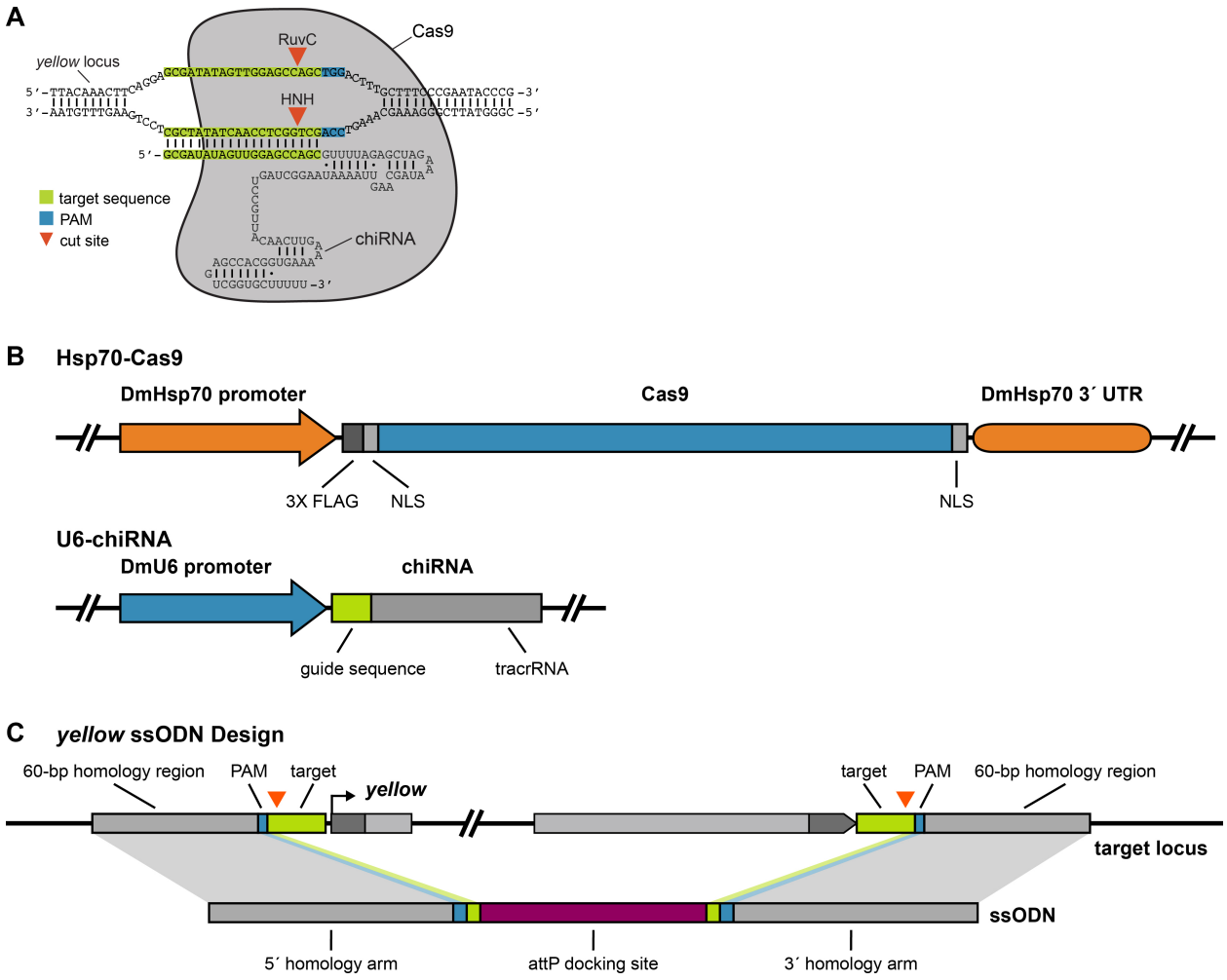

Figure S1. CRISPR RNA/Cas9 molecular reagents used in this study. (A) Schematic of the modified CRISPR RNA/Cas9 system (based on Jinek et al., 2012) used to target *yellow* in this study (YE1 is shown as an example). A chiRNA guides Cas9 to complementary target DNA. Target recognition requires 20-nt of homology and a 3-bp PAM sequence, NGG, at the 3′ end of the genomic target sequence. Cas9 cleaves the complementary DNA strand via its HNH endonuclease domain, while its RuvC-like endonuclease domain cleaves the noncomplementary strand (red arrowheads), resulting in a DSB at the targeted site. (B) Cas9, derived from *Streptococcus pyogenes* and codon optimized for eukaryotic expression (CONG *et al.* 2013), was expressed under the control of the Drosophila *hsp70* promoter and 3′ UTR. The chiRNAs utilized in this study consist of 20-nt of target-specific guide sequence followed by 76-nt of tracrRNA sequence as this chiRNA exhibited maximal activity in mammalian cells (Patrick Hsu and Feng Zhang, personal communication). The chiRNA is transcribed from the Drosophila snRNA:U6:96Ab promoter (WAKIYAMA *et al.* 2005). (C) ssODN donor templates contain 60-nt homology arms immediately adjacent to each site-specific cut flanking a 50-nt attP docking site for efficient subsequent manipulation of the targeted locus. Red arrowheads mark the predicted Cas9 cleavage sites. Note that the ssODN is designed such that neither it nor the modified locus will be CRISPR targets. Not to scale. More detailed information and protocols can be found at FlyCRISPR.molbio.wisc.edu.

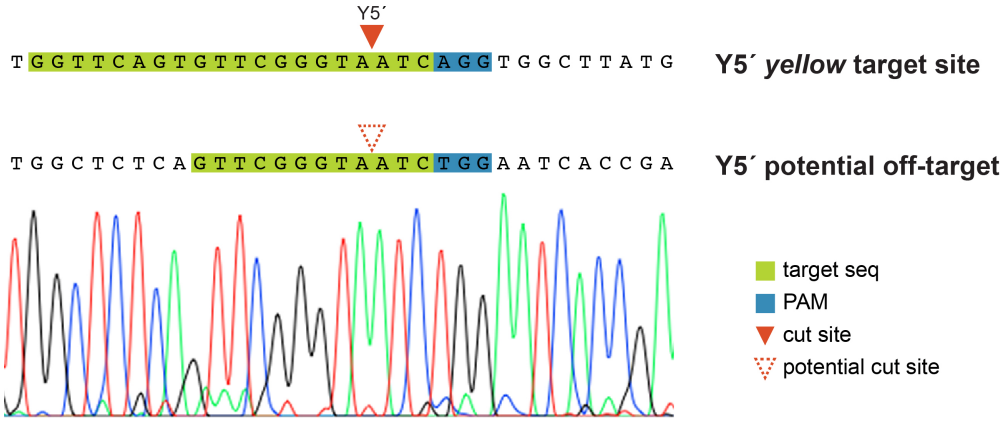

Figure S3. Off-target analysis of germline transformants. The potential off-target site of Y5′, based on the criteria established by Cong et al., 2013, is shown with a representative sequence trace indicating the presence of wild-type sequence at this site in germline transformants generated with the Y5′ chiRNA. Note that Y3′ and YE1 have no potential off-target sites that meet these criteria.

References

Banga, S. S., and J. B. Boyd, 1992 Oligonucleotide-directed site-specific mutagenesis in Drosophila melanogaster. Proc Natl Acad Sci U S A 89**:** 1735-1739.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval *et al.*, 2007 CRISPR provides acquired resistance against viruses in prokaryotes. Science 315**:** 1709-1712.

Beumer, K. J., J. K. Trautman, A. Bozas, J. L. Liu, J. Rutter *et al.*, 2008 Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105**:** 19821-19826.

Beumer, K. J., J. K. Trautman, K. Mukherjee and D. Carroll, 2013 Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. G3 (Bethesda).

Bibikova, M., M. Golic, K. G. Golic and D. Carroll, 2002 Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161**:** 1169-1175.

Bozas, A., K. J. Beumer, J. K. Trautman and D. Carroll, 2009 Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics 182**:** 641-651.

Chang, N., C. Sun, L. Gao, D. Zhu, X. Xu *et al.*, 2013 Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. Cell Res 23**:** 465-472.

Cho, S. W., S. Kim, J. M. Kim and J. S. Kim, 2013 Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31**:** 230-232.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science 339**:** 819-823.

Dicarlo, J. E., J. E. Norville, P. Mali, X. Rios, J. Aach *et al.*, 2013 Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41**:** 4336-4343.

Ding, Q., S. N. Regan, Y. Xia, L. A. Oostrom, C. A. Cowan *et al.*, 2013 Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. Cell Stem Cell 12**:** 393-394.

Gao, G., C. McMahon, J. Chen and Y. S. Rong, 2008 A powerful method combining homologous recombination and site-specific recombination for targeted mutagenesis in Drosophila. Proc Natl Acad Sci U S A 105**:** 13999-14004.

Gao, G., N. Wesolowska and Y. S. Rong, 2009 SIRT combines homologous recombination, site-specific integration, and bacterial recombineering for targeted mutagenesis in Drosophila. Cold Spring Harb Protoc 2009**:** pdb prot5236.

Gloor, G. B., N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston and W. R. Engels, 1991 Targeted gene replacement in Drosophila via P element-induced gap repair. Science 253**:** 1110-1117.

Gong, W. J., and K. G. Golic, 2003 Ends-out, or replacement, gene targeting in Drosophila. Proc Natl Acad Sci U S A 100**:** 2556-2561.

Huang, J., W. Zhou, W. Dong, A. M. Watson and Y. Hong, 2009 From the Cover: Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. Proc Natl Acad Sci U S A 106**:** 8284-8289.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai *et al.*, 2013 Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31**:** 227-229.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura and A. Nakata, 1987 Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol 169**:** 5429-5433.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337**:** 816-821.

Jinek, M., A. East, A. Cheng, S. Lin, E. Ma *et al.*, 2013 RNA-programmed genome editing in human cells. Elife 2**:** e00471.

Liu, J., C. Li, Z. Yu, P. Huang, H. Wu *et al.*, 2012 Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics 39**:** 209-215.

Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf and E. V. Koonin, 2006 A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1**:** 7.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell *et al.*, 2013 RNA-guided human genome engineering via Cas9. Science 339**:** 823-826.

McVey, M., and S. E. Lee, 2008 MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24**:** 529-538.

Nassif, N., J. Penney, S. Pal, W. R. Engels and G. B. Gloor, 1994 Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol 14**:** 1613-1625.

Rong, Y. S., and K. G. Golic, 2000 Gene targeting by homologous recombination in Drosophila. Science 288**:** 2013-2018.

Shen, B., J. Zhang, H. Wu, J. Wang, K. Ma *et al.*, 2013 Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res.

St Johnston, D., 2002 The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet 3**:** 176-188.

Venken, K. J., and H. J. Bellen, 2005 Emerging technologies for gene manipulation in Drosophila melanogaster. Nat Rev Genet 6**:** 167-178.

Wakiyama, M., T. Matsumoto and S. Yokoyama, 2005 Drosophila U6 promoter-driven short hairpin RNAs effectively induce RNA interference in Schneider 2 cells. Biochem Biophys Res Commun 331**:** 1163-1170.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng *et al.*, 2013 One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell 153**:** 910-918.

Wood, A. J., T. W. Lo, B. Zeitler, C. S. Pickle, E. J. Ralston *et al.*, 2011 Targeted genome editing across species using ZFNs and TALENs. Science 333**:** 307.

Chapter 4

Highly specific and efficient CRISPR-Cas9-catalyzed homology-directed repair in Drosophila

Scott J. Gratz[1§], Fiona P. Ukken[2§], C. Dustin Rubinstein[3], Gene Thiede[3], Laura K. Donohue[3], Alexander M. Cummings[3], and Kate M. O'Connor-Giles[1,2,3*]

[1] Genetics Training Program, University of Wisconsin-Madison, Madison, WI 53706

[2] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

[3] Laboratory of Cell and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706

[§] These authors contributed equally to this work.

This chapter is adapted from the publication:

Gratz, S J, F P Ukken, C D Rubinstein, G Thiede, and et al. 2014. "Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila." *Genetics* 196 (4): 961–71.

Abstract

We and others recently demonstrated that the readily programmable CRISPR/Cas9 system can be used to edit the Drosophila genome. However, most applications to date have relied on aberrant DNA repair to stochastically generate frame-shifting indels and adoption has been limited by a lack of tools for efficient identification of targeted events. Here we report optimized tools and techniques for expanded application of the CRISPR/Cas9 system in Drosophila through homology-directed repair (HDR) with double-stranded DNA (dsDNA) donor templates that facilitate complex genome engineering through the precise incorporation of large DNA sequences including screenable markers. Using these donors, we demonstrate the replacement of a gene with exogenous sequences and the generation of a conditional allele. To optimize efficiency and specificity, we generated transgenic flies that express Cas9 in the germline, and directly compared HDR and off-target cleavage rates of different approaches for delivering CRISPR components. We also investigated HDR efficiency in a mutant background previously demonstrated to bias DNA repair towards HDR. Finally, we developed a web-based tool that identifies CRISPR target sites and evaluates their potential for off-target cleavage using empirically rooted rules. Overall, we have found that injection of a dsDNA donor and guide RNA-encoding plasmids into *vasa-Cas9* flies yields the highest efficiency HDR, and that target sites can be selected to avoid off-target mutations. Efficient and specific CRISPR/Cas9-mediated HDR opens the door to a broad array of complex genome modifications and greatly expands the utility of CRISPR technology for Drosophila research.

Introduction

Genome engineering is a powerful tool for dissecting biological mechanisms. The ability to precisely edit genomes has expanded significantly in recent years with the development of sequence-directed nucleases that can generate targeted double-strand breaks (DSBs) in DNA (BIBIKOVA *et al.* 2002; LIU *et al.* 2012; BASSETT *et al.* 2013; GRATZ *et al.* 2013a; YU *et al.* 2013). Chromosomal DSBs trigger DNA repair via two cellular pathways that can be harnessed for genome editing. Non-homologous end-joining (NHEJ) is an error-prone ligation process that can result in small insertions and deletions (indels) at cleavage sites. By targeting open-reading frames, this pathway can be used to disrupt genes through frame-shifting mutations. Homologous recombination or homology-directed repair (HDR) employs homologous DNA sequences as templates for precise repair. By supplying a donor repair template, this pathway can be exploited to precisely edit genomic sequence or insert exogenous DNA. Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are sequence-directed nucleases that, in pairs, can generate targeted DSBs and have shown promise as genome engineering tools in Drosophila (Bibikova et al., 2002; Liu et al., 2012). However, ZFN and TALEN proteins must be custom designed for each genome modification – a costly and uncertain process. In contrast, in the CRISPR/Cas9 system, a common nuclease, Cas9, is directed to targeted loci through simple base pairing by a small RNA molecule. With the ease of producing a targeting RNA, the CRISPR/Cas9 system promises to transform genome engineering.

Endogenous prokaryotic type II CRIPSR immune systems comprise a single polypeptide nuclease, Cas9, that is guided to target sites by a complex of two small RNAs (GASIUNAS *et al.* 2012; JINEK *et al.* 2012). For use in genome engineering, the *Streptococcus pyogenes* system was simplified to two-components: the Cas9 nuclease and a single

chimeric guide RNA (gRNA). The gRNA recognizes a 20-nt target sequence next to a trinucleotide NGG protospacer adjacent motif (PAM) in the genome to direct Cas9-mediated cleavage of both DNA strands within the target sequence (JINEK *et al.* 2012). We and others have recently demonstrated that the CRISPR/Cas9 system can be employed to make targeted modifications to the Drosophila genome that are efficiently transmitted through the Drosophila germline (BASSETT *et al.* 2013; GRATZ *et al.* 2013a; KONDO and UEDA 2013; REN *et al.* 2013; SEBO *et al.* 2013; YU *et al.* 2013). Nearly all the work conducted to date in Drosophila has taken advantage of aberrant NHEJ to generate frame-shifting mutations in open reading frames. While this is an effective approach for disrupting gene function, it is stochastic and limited to the generation of indels. NHEJ does not permit the precise incorporation of exogenous DNA, including visual markers to aid in screening, limiting its application. Further, more complex genome engineering applications such as the incorporation of recombination sites for making conditional alleles, endogenous protein tagging, or precise editing of genomic sequences require HDR (GOLIC 2013; GRATZ *et al.* 2013b; BASSETT and LIU 2014; BEUMER and CARROLL 2014). Here we present tools and techniques that overcome these limitations. We demonstrate efficient methods for CRISPR/Cas9-mediated HDR using dsDNA donors that facilitate the incorporation of large exogenous sequences including easily screened and removable visible markers. We also present a web-based tool for identifying and evaluating CRISPR target sites, and demonstrate that gRNAs designed with this tool are highly specific. These advances enable broad application of the CRISPR/Cas9 system for complex genome engineering in Drosophila.

Materials and Methods

Generation of *vasa-Cas9* lines

To generate transgenic flies that express Cas9 in the germline, Cas9 was placed under the regulatory control of *vasa*. We began with the p3xP3-EGFP/vasa- ΦC31^{NLS}attB plasmid (gift from Konrad Basler). The coding sequence for ΦC31 was replaced with an XbaI-containing multiple cloning site through amplification of the *vasa* regulatory elements and plasmid backbone. Sequence encoding 3xFLAG-NLS-Cas9-NLS was amplified as an XbaI-flanked fragment from pX330 (CONG *et al.* 2013) and ligated into the XbaI site of the modified p3xP3-EGFP/*vasa*- ΦC31^{NLS}attB plasmid. The resulting p3xP3-EGFP/*vasa*-3xFLAG-NLS-Cas9-NLS plasmid was used for ΦC31-mediated integration (Supporting Information, Figure S1 and Table S1).

Fly stocks

hs-Cre, *hs-flp DSH3PX1^{EY08084}* and *lig4^{169}* lines are available at the Bloomington Drosophila Stock center (AHMAD and GOLIC 1996; SIEGAL and HARTL 1996; BELLEN *et al.* 2004; MCVEY *et al.* 2004).

Molecular reagents

pBS-hsp70-Cas9: To generate a high-copy number version of the Cas9 vector for injection, a NotI fragment containing Cas9 and the hsp70 regulatory regions was excised from phsp70-Cas9 (GRATZ *et al.* 2013a). This fragment was ligated into the NotI site of pBluescript SK+ (Agilent Technologies).

pU6-gRNAs: Target-specific sequences for *rosy* and *DSH3PX1* gRNAs were synthesized as 5′-phosphorylated oligonucleotides, annealed, and ligated into the BbsI sites of pU6-BbsI-chiRNA (GRATZ *et al.* 2013a).

In vitro transcribed gRNAs: Following the protocol of Basset et al., 2013, *DSH3PX1* gRNAs S1 and S2 were generated by PCR (Phusion polymerase, NEB) amplification of two oligonucleotides that anneal to create a template for *in vitro* transcription by T7 polymerase. PCR products were purified using the QiaexII agarose gel extraction kit (Qiagen) and *in vitro* transcription was performed with the Megascript T7 Kit (Ambion). Transcripts were purified by phenol-choloroform extraction and isopropanol precipitation.

Donor templates: The pHD-DsRed-attP vector was designed for rapid generation of gene-specific donor templates and custom synthesized (DNA2.0; Supporting Information, Figure S1). The vector contains an attP ΦC31 docking site for subsequent modification of the targeted locus and a 3xP3-DsRed marker that drives expression of DsRed in the eye, flanked by loxP recombination sites for its removal. Multiple cloning sites on either side of the replacement/marker cassette are flanked by type IIS restriction sites, AarI and SapI respectively to facilitate seamless incorporation of homology arms. The multiple cloning sites of pHD-DsRed-attP were designed to conserve as many restriction sites as possible from pGX-attP to make the vector compatible with homology arms previously generated for ends-out homologous recombination (HUANG *et al.* 2009).

To generate the *DSH3PX1* replacement donor, 1.1-kb homology arms flanking the S1 and S2 cleavage sites were amplified and incorporated via AarI and SapI, respectively (Supporting Information, Figure S1 and Table S2). The *DSH3PX1* conditional allele donor was constructed by assembling three DNA fragments into a pUC19 vector. A 1.5 kb 5′-homology arm leading up to the S1 cleavage site was amplified (Phusion polymerase, NEB) with a forward primer containing a SpeI site and a reverse primer incorporating a

partial (up to the XbaI site) FRT site. This fragment was cloned into the XbaI site of pUC19. A 2.5-kb region beginning downstream of S1 cleavage site and extending to the second intron of *DSH3PX1* was amplified with a forward primer incorporating a partial (beginning at the XbaI) FRT sequence and a reverse primer containing a HindIII site. This fragment was cloned into the XbaI and HindIII sites of the pUC19-fragment 1 plasmid. A third fragment consisting of the 3xP3-DsRed-FRT sequence was amplified from a transgenic fly line with a pBac{SAstopHD-DsRed} insertion using primers containing NgoMIV site overhangs. The PCR product was ligated into the pUC19-fragments 1 and 2 plasmid at the NgoMIV site present in the previously cloned intronic sequence of *DSH3PX*1 (Supporting Information, Figure S1 and Table S2). We have since generated an additional donor cloning vector containing 3xP3-DsRed but no attP site (pHD-DsRed) for rapid cloning of similar dsDNA donors that will be available through Addgene.

Embryo injections

Preblastoderm embryos were injected through the chorion membrane using standard protocols. Injections were carried out at 18°C and embryos were shifted to 25°C immediately following injection. pBS-hsp70-Cas9 was injected at a concentration of 250 ng/μL. The pU6-gRNA targeting constructs were injected at 250 ng/μL for single gRNAs and 100 ng/μL each when two gRNAs were injected. *In vitro* transcribed gRNAs were injected at 25 or 50 ng/μL. dsDNA donor templates were injected at a concentration of 500 ng/μL. All injection mixtures were prepared in water. *lig4* injections were performed by Genetic Services, Inc.

Screening

To assess germline transmission of targeted modifications, adults that developed from injected embryos were outcrossed to $y^1; ry^1$ or w^{1118}. Offspring were screened for 10 days after the first flies emerged for progeny with rosy or RFP+ eyes, indicating transmission of the dsDNA donor. Transmission rates were calculated both as a percent of fertile crosses that produced one or more transgenic progeny and as a percent of total progeny. Note that in nearly every case, crosses that yielded transformed progeny did so within the first few days of screening, suggesting that screening effort can be minimized by limiting screening to 3-5 days with little risk of missing independent transformants. Biallelic events were recognized by either transmission of modified loci to greater than 50% of progeny or modification of two discernable chromosomes (e.g., the *vasa-Cas9* and balancer chromosomes).

Molecular characterization of engineered and off-target loci

Genomic DNA was isolated from individual F1 flies. PCR was performed using primers flanking the targeted modifications or potential off-target cleavage site. Amplified products were purified and sequenced.

Flp and Cre induction

To remove the DsRed visible marker, $SH3PX1^{DsRed\text{-}attP}$ males were crossed to *hs-Cre* virgin females. Progeny were either reared at 25°C without heat shock or at 18°C with a single heat shock of 1 hour at 37°C. To remove the first exon of *SH3PX1*, $SH3PX1^{conditional}$ flies were crossed to *hs-FLP*. Progeny were heat shocked every 24 hours until pupation at 37°C for 1h.

Development of CRISPR Optimal Target Finder

Euchromatic regions of the dm3/BDGP release 5 *Drosophila melanogaster* genome were indexed as in Iseli et al., 2007. PHP code was developed to (1) parse user-inputted DNA sequence to detect CRISPR targets on both strands, (2) execute fetchGWI (ISELI *et al.* 2007) to identify similar sequences elsewhere in the genome, (3) employ algorithms based on empirical rules and user-selected parameters to identify potential off-target cleavage sites, and (4) return CRISPR target sites ranked by specificity along with location information and a Gbrowse link for each potential off-target site. The following invertebrate genomes were processed identically: *D. simulans* (annotation DroSim1), *D. yakuba* (DroYak2), *D. sechellia* (DroSec1), *D. virilis* (DroVir3), two strains of *Anopheles gambiae* (AgamM1 and AgamS1)*, Aedes aegypti* (AaegL1), *Apis mellifera* (apiMel3), *Tribolium castaneum* (TriCas2), and *Caenorhabditis elegans* (ce10). A detailed user manual is available at http://tools.flycrispr.molbio.wisc.edu/targetFinder/CRISPRTargetFinderManual.pdf.

Results

Germline expression of Cas9 increases targeting efficiency

Previous work has shown that germline expression of ΦC31 phage recombinase yields increased transformation efficiency over injection protocols (BISCHOF *et al.* 2007). To determine if this is the case for Cas9, we generated transgenic fly lines that express Cas9 in the germline under the control of *vasa* regulatory regions (Figure 1A). These lines are homozygous viable, fertile and healthy suggesting low toxicity from germline expression of Cas9.

CRISPR/Cas9 targeting efficiency varies significantly between loci and even between target sites within the same locus (BASSETT *et al.* 2013; GRATZ *et al.* 2013a; REN *et al.* 2013; YU *et al.* 2013). Therefore, to directly compare the efficiency of targeting in *vasa-Cas9* flies to injection of *Oregon-R* flies with *hsp70-Cas9*, a method we previously employed in targeting *yellow*, we assessed two separate gRNAs targeting the *rosy* locus (Figure 1B) (GRATZ *et al.* 2013a). We found that germline expression of Cas9 yielded the highest efficiencies (Table 1). This effect was most pronounced for the R3 guide, which was moderately efficient in the injection paradigm with 8% of fertile injectees (founders) transmitting *rosy* mutations to 1.4% of all progeny and highly efficient in *vasa-Cas9* flies with 53% of fertile injectees transmitting *rosy* mutations to 15% of progeny. Interestingly, we did not observe a difference in the rate of biallelic indel formation, with both injected and integrated *Cas9* approaches yielding founders in which both alleles were mutated 29% of the time. Overall, our results demonstrate high levels of NHEJ with germline-expressed Cas9, and are in agreement with recent investigations of *vasa-Cas9* in the targeting of integrated GFP and RFP plasmids; *nanos-Cas9* in the targeting of *white;* and, as reported at by Hui-Min Chen and Tzumin Lee at CRISPRflydesign.org, *UAS-Cas9* under the control of a germline Gal4 driver (REN et al. 2013; SEBO et al. 2013).

To assess the nature of the *rosy* mutations, we amplified and sequenced the target sites in F1 flies. We observed similar indels in both cases, indicating that repair of targeted Cas9-induced cleavage by NHEJ occurred in both the injection and germline expression approaches (Figure 1C).

CRISPR/Cas9-mediated HDR with dsDNA donors

Most of the initial experiments employing the CRISPR/Cas9 system in Drosophila have depended on easily scored phenotypes for identification of targeted events (BASSETT *et al.* 2013; GRATZ *et al.* 2013a; KONDO and UEDA 2013; REN *et al.* 2013; SEBO *et al.* 2013; YU *et al.* 2013). To develop a more universal method to screen for targeted events, we optimized CRISPR/Cas9-mediated HDR using dsDNA donors that enable the incorporation of larger DNA sequences, including visible markers. We generated a donor vector that contains multiple cloning sites for rapid incorporation of flanking homology arms, an attP ΦC31 phage recombination site for subsequent access to the locus, and the visible marker 3xP3-DsRed flanked by loxP sites for CRE-mediated removal (Figure 2A).

To assess dsDNA-mediated HDR, we targeted the *DSH3PX1* locus for replacement with an attP docking site (Figure 2B). Our general strategy was to employ two gRNAs targeting sites 5′ and 3′ of the gene and our dsDNA donor with homology arms of approximately 1 kb immediately flanking the targeted cleavages sites (Figure 2B). Homology arm length was selected to balance efficiency and ease of cloning based on a detailed study of ZFN-mediated HDR (BEUMER *et al.* 2013). To optimize efficiency, we directly compared three approaches for delivery of CRISPR components: (1) injection of dsDNA donor and DNA plasmids encoding gRNAs and Cas9 into w^{1118} embryos, (2) injection of dsDNA donor and DNA plasmids encoding gRNAs into *vasa-Cas9* embryos,

and (3) injection of dsDNA donor and *in vitro* transcribed gRNAs into *vasa-Cas9*. Based on the incorporation of the DsRed marker, the overall efficiency of HDR was very high (Table 2). Injection of dsDNA donor and DNA plasmids encoding gRNAs into *vasa-Cas9* flies yielded the highest efficiency of germline transmission (18% founder rate), while gRNA injection as RNA into *vasa-Cas9* yielded the lowest efficiency (3% founder rate). Furthermore, in *vasa-Cas9* targeting experiments 70% of founders exhibited biallelic HDR versus 0% for injections into w^{1118} embryos, consistent with high targeting efficiency with germline expression of Cas9.

To confirm that the precisely targeted events had occurred in the DsRed-positive flies, we amplified and sequenced the engineered region in at least one F1 fly per founder. In 11/12 cases, HDR occurred as anticipated. Thus, with DNA injection into *vasa-Cas9*, 16% of crosses yielded precise HDR. In one of nine independent cases where DsRed-positive flies were recovered from *vasa-Cas9* flies injected with CRISPR components as DNA, the donor was not incorporated at the targeted locus but rather elsewhere in the genome. A previous study in zebrafish demonstrated that homology-independent integration of DNA occurs preferentially at CRISPR-induced DSBs, so we assayed predicted off-target cleavage sites for donor integration by PCR (AUER *et al.* 2014). We found no evidence of integration at any of the eight predicted off-target sites (see Figure 5), suggesting either a random integration event or a complex rearrangement of the target site, both of which have been observed with traditional ends-in and ends-out homologous recombination methods.

In previous studies employing ZFNs, mutation of *lig4*, a DNA ligase in the NHEJ pathway, biased endogenous DNA repair towards HDR (BEUMER *et al.* 2008). To determine the effect on CRISPR/Cas9-mediated genome engineering, we injected the

same dsDNA donor and gRNA plasmids as above along with Cas9 into *lig4* mutants. Of 28 fertile crosses, none yielded DsRed positive progeny. However, we did find that injection into a *lig4* background increased HDR efficiency over injection into wild type in analogous experiments targeting a 25-kb gene (4% vs. 0% founder rate, data not shown). Thus, mutations in *lig4* can facilitate CRISPR/Cas9-mediated HDR. Note, however, that in targeting this gene, injection of CRISPR components as DNA into *vasa-Cas9* flies again yielded the highest efficiency HDR (7% founder rate, data not shown).

Finally, we confirmed the CRE-mediated removal of the 3xP3-DsRed marker (data not shown). As previously reported, we found that *hs-Cre* is leaky, and succeeded in removing the marker in germ cells by either rearing crosses at 25°C without heat shock or rearing at 18°C with a single heat shock of 1 hour at 37°C (HAMPEL *et al.* 2011).

Generation of conditional alleles with CRISPR/Cas9-mediated HDR

The replacement of a gene with an attP recombination docking site provides ongoing genetic access to a targeted locus allowing for an unlimited array of subsequent modifications (GAO *et al.* 2009; HUANG *et al.* 2009). However, in many circumstances, including the targeting of genes required for viability or the precise insertion of exogenous DNA sequences, it will be preferable to directly target a locus for specific modifications. To investigate the possibility of using our marked dsDNA donor strategy to mediate such modifications we sought to create a conditional allele of *DSH3PX1* through the site-specific integration of FRT sites flanking the first exon. We injected a dsDNA donor containing the first exon and a visible marker flanked by FRT sites along with DNA plasmids encoding flanking gRNAs into *vasa-Cas9* embryos (Figure 3A). Two out of 16 crosses yielded DsRed-positive flies for a founder rate of 13% (Table 2). Upon molecular characterization, we found that one of the two founder flies transmitted the

intended repair event, yielding an HDR founder rate of 6%. The imperfect repair event reflected accurate incorporation of the donor at the 5′ end through the DsRed marker and an unknown rearrangement at the 3′ end.

Our donor template was designed such that the Flp-mediated removal of exon 1 also results in the removal of the DsRed marker (Figure 3A). We followed a standard heat-shock protocol and observed the loss of DsRed in the majority of the eye indicating successful removal of the FRT-flanked sequences (Figure 3B).

Targeted deletions can be rapidly identified through negative screening

An advantage of the CRISPR/Cas9 system is the fact that CRISPR components can be introduced into any genetic background through injection. We reasoned that we could capitalize on this to facilitate the identification of targeted deletions by injecting into a fly line that contains a marked transposable element in the targeted locus and screening for its loss. To test this, we targeted the *DSH3PX1* locus for deletion in $DSH3PX1^{EY08084}$ flies, which contain a 10.9-kb EP element in the first exon (Figure 4A). We employed flanking gRNAs, which were injected as plasmid DNA along with *hsp70-Cas9* into $DSH3PX1^{EY08084}$ embryos. The *DSH3PX1* locus is 3.3 kb, bringing the total size of the targeted deletion to 14.2 kb. We observed loss of the w^{+mC} marker in progeny of 1/17 crosses (6%). Molecular analysis confirmed that the targeted deletion had occurred in these flies (Figure 4B).

Optimizing CRISPR/Cas9 specificity

To facilitate the identification of highly specific CRISPR targets in Drosophila, we developed a freely available web-based tool to identify and evaluate CRISPR targets within a sequence of genomic DNA (tools.flycrispr.molbio.wisc.edu). First, CRISPR

Optimal Target Finder identifies all possible CRISPR target sites in user-provided genomic sequence. The user can specify whether they would like the program to identify all CRISPR targets in the input sequence, only those that start with a G for efficient expression from the U6 promoter when supplying gRNA as plasmid DNA, or only those that start with GG for *in vitro* transcription gRNAs from the T7 promoter. Once the pool of target sites is identified, the program rapidly assesses the potential for off-target cleavage for each identified CRISPR target using user-defined criteria and rules based on empirical findings. To find all genomic sequences with sequence similarity to each CRISPR target query, CRISPR Optimal Target Finder employs the fetchGWI algorithm, which can exhaustively search complete-genome sized databases for similar sequences with sub-second speed, while allowing a specified number of mismatches (Iseli et al., 2007). Similar sequences are filtered via algorithms based on large-scale analyses of CRISPR/Cas9 specificity in cell lines (CRADICK *et al.* 2013; FU *et al.* 2013; HSU *et al.* 2013; MALI *et al.* 2013b; PATTANAYAK *et al.* 2013; CHO *et al.* 2014) or animals (CHIU *et al.* 2013; YANG *et al.* 2013) to identify potential off-target cleavage sites for each CRISPR target. These and previous studies demonstrate that the PAM-proximal region of the CRISPR target sequence (frequently referred to as the "seed") is more critical for specificity than the distal eight nucleotides, so our search and filtering strategies consider both the number and location of mismatches to evaluate all potential off-target cleavage sites (SEMENOVA *et al.* 2011; WIEDENHEFT *et al.* 2011; CRADICK *et al.* 2013; FU *et al.* 2013; HSU *et al.* 2013; MALI *et al.* 2013b; PATTANAYAK *et al.* 2013; YANG *et al.* 2013; CHO *et al.* 2014). These studies also suggest that off-target cleavage, or it's aberrant repair, may be greater in cell lines than in animals, so the program allows the user to select the stringency of the algorithm employed. Maximum stringency criteria are based on the composition of off-target cleavage sites observed in large-scale analyses carried out in cell lines (CRADICK *et al.* 2013; FU *et al.* 2013; HSU *et al.* 2013; MALI *et al.* 2013b)

and define potential off-target sites elsewhere in the genome as sites with: (i) perfect matches (zero mismatches) to the seed sequence; (ii) one mismatch in the seed sequence and four or fewer mismatches in the distal sequence; and (iii) two mismatches in the seed sequence with a maximum of one mismatch in the distal sequence. High stringency criteria are based on the composition of off-target cleavage sites observed in large-scale analyses conducted in animals (CHIU *et al.* 2013; YANG *et al.* 2013) and define potential off-target sites as those with (i) perfect matches (zero mismatches) to the seed sequence or (ii) one mismatch in the seed sequence when there are only one or zero mismatches in the distal sequence. Finally, the program allows the user to choose whether or not NAG-adjacent sites are considered in the evaluation of CRISPR target sequence specificity. CRISPR sequences adjacent to an NAG PAM sequence can be cleaved at 1/5th the efficiency of those adjacent to a canonical NGG PAM sequence in transformed cell lines, though this has not been observed in animals to date(HSU *et al.* 2013). Once potential off-target cleavage sites are found, the program returns all identified CRISPR targets sorted by the likelihood of off-target nuclease activity along with the sequence and genomic location, including a Gbrowse link, of each predicted off-target site.

CRISPR Optimal Target Finder also provides the ability to search additional invertebrate genomes. To facilitate the adoption of CRISPR in non-model genetic systems, we included models with relevance to agriculture and global health: *D. simulans*, *D. yakuba*, *D. sechellia*, *D. virilis*, *Apis mellifera*, *Tribolium castaneum*, two strains of *Anopheles gambiae, Aedes aegypti*, and *Caenorhabditis elegans*.

With the exception of gRNA S3, all of the guides used in this work were designed to avoid off-target sequences defined by the animal-based, or 'high' stringency, criteria located next to a canonical NGG PAM (Figure 5A). To assess the specificity of our gRNAs, we

amplified and sequenced the 14 potential off-target sites defined by the more stringent cell culture-based, or 'maximum', criteria and adjacent to a canonical PAM for gRNAs R2, R3, S1 and S2 in 23 independent F1s, and observed intact sequence in every case (Figure 5). We also identified and assessed four potential off-target sites with perfect seed matches to gRNAs R2, S1 and S2 adjacent to non-canonical NAG PAMs, and again observed no indels (Figure 5). These results demonstrate that careful selection of target sites using the high stringency criteria enables the generation of lines carrying highly-specific targeted mutations.

To further probe the potential for off-target cleavage with CRISPR/Cas9, gRNA S3 was designed to challenge the system. S3 has six PAM-adjacent perfect matches to its seed sequence elsewhere in the genome (Figure 5A). The overall identity of these off-target sequences ranges from 16/20 to 12/20 (Figure 5B). We amplified and sequenced each potential off-target site in *DSH3PX1*^{conditional} F1s and found no evidence of cleavage. Similarly, no evidence of off-target cleavage was observed at the one NAG-adjacent perfect seed match in the genome (Figure 5). This demonstrates that the CRISPR/Cas9 system can be employed for highly specific genome editing in the Drosophila germline even when perfect matches to the CRISPR target seed sequence are found elsewhere in the genome, and reinforces the role of the distal eight nucleotides in targeting specificity.

Discussion

Our results demonstrate that the CRISPR/Cas9 system can catalyze complex genome engineering with high efficiency and specificity. We present a universal approach for identifying targeted events through HDR with dsDNA donors containing positive markers, and demonstrate that this approach can be used in conjunction with germline expression of Cas9 to efficiently replace a gene or generate a conditional allele. Through our analysis of off-target cleavage, we show that target site selection with our web-based tool facilitates highly specific modification of the genome free from unintended mutations.

The broad application of CRISPR/Cas9 genome engineering requires tools for rapid identification of targeted events. In most cases phenotypic screening will not be an option, necessitating alternative approaches. PCR-based molecular screening methods are a universal option and have been demonstrated with CRISPR/Cas9-mediated NHEJ, but the associated generation and maintenance of candidate fly stocks is time and labor intensive (Yu *et al.* 2013). We demonstrate two alternative approaches: negative screening through the removal of marked elements in targeted lines and positive screening through HDR with dsDNA donors containing visible markers. The simplicity of the negative screening approach combined with the presence of marked elements in most Drosophila genes makes it an appealing option for some applications. On the other hand, HDR with dsDNA donor templates offers both the opportunity for incorporating screenable markers and the capacity to introduce more complex and larger-scale modifications.

To optimize CRISPR/Cas9-mediated HDR approaches in Drosophila, we performed a direct comparison of the efficiency of multiple approaches. We found that injection of

gRNAs as DNA into *vasa-Cas9* flies maximized efficiency over either injection of in vitro transcribed gRNAs into *vasa-Cas9* flies or co-injection of gRNA and Cas9 DNA into w^{1118} flies. Ren et al., 2013 also observed increased efficiency with germline-expressed Cas9 when targeting *white* for mutation via NHEJ. Our survival rate for both NHEJ and HDR in *vasa-Cas9* flies was approximately 10%. This is lower than observed for *nanos-Cas9*, and may indicate that *nanos*-driven *Cas9* is less toxic than *vasa*-driven *Cas9*, possibly due to more restricted expression in preblastoderm embryos. Alternatively, the difference may reflect variability in the gRNAs, targeted locus or injection technique. Note also that our *vasa-Cas9* lines were not homozygous, thus approximately 25% of the embryos were homozygous for the balancer chromosome and genetically inviable. Consistent with high targeting efficiency, we observed a high rate of biallelic HDR in *vasa-Cas9* founders (70%), while this was not observed in Cas9-injected *wild-type* or transposon-expressing founders. It is important to note that for some applications, such as the targeting of genes required in the germline, it will be desirable to avoid higher cleavage rates that frequently yield biallelic breaks and would, thus, decrease the likelihood of recovering a targeted event.

It is somewhat surprising that we observed lower rates of HDR when injecting gRNAs as DNA since previous studies have reported very high rates of CRISPR/Cas9-mediated NHEJ with injection of in vitro transcribed gRNAs (BASSETT *et al.* 2013; YU *et al.* 2013). This could be attributable to differences in injection protocols and handling, or may reflect the different gRNA concentrations employed. We were unable to co-inject dsDNA donor at our standard concentration (500 ng/μL) with in vitro transcribed gRNAs at concentrations greater than 25 ng/μL each without rapidly clogging the injection needle. As a result, nearly all of the flies we analyzed were injected with half as much gRNA as in previous studies. In either case, our experiences highlight a key advantage of DNA and

transgenic approaches: RNA is more difficult to handle, particularly when sending CRISPR components to injection companies as most labs do, as well as significantly more expensive to generate.

We also used CRISPR/Cas9-mediated HDR to generate a conditional allele through the incorporation of FRT sites flanking the first exon on *DSH3PX1*. This experiment demonstrates that donor templates can be designed for the incorporation of discontinuous exogenous DNA sequences – an approach that enables careful placement of the visible marker for minimal disruption. Analogous donors can be designed for the integration of in-frame protein tags or editing of specific basepairs with visible markers placed in adjacent introns.

In all our HDR experiments, we employ circular dsDNA donor plasmids. We did not compare the efficiency of circular to linear plasmids in our paradigm as injected linear donors had previously been shown ineffective, possibly due to degradation (BEUMER *et al.* 2008). A recent report demonstrated that CRISPR-mediated cleavage of a target site with a single gRNA could improve ends-out homologous recombination catalyzed by a linearized dsDNA donor template (BAENA-LOPEZ *et al.* 2013). However, this approach requires the co-introduction of Flp and I-SceI enzymes and the efficiency appears quite low. A second notable feature of our approach is the use of two gRNAs flanking the targeted locus. While we have not conducted a rigorous comparison of one vs. two gRNAs in HDR, the idea that two gRNAs improve efficiency is consistent with studies demonstrating significantly higher rates of NHEJ-mediated mutation with the introduction of a second gRNA (GRATZ *et al.* 2013a; KONDO and UEDA 2013; REN *et al.* 2013).

Off-target cleavage has been raised as a serious concern for CRISPR/Cas9-mediated genome engineering (CRADICK *et al.* 2013; FU *et al.* 2013). Unlike TALENS and ZFNs, which function as obligate dimers, Cas9 generates DSBs with its two nuclease domains, so only 22 nucleotides of sequence determine cleavage specificity. Further, it has been shown that the CRISPR/Cas9 complex can tolerate mismatches particularly at the 5′ end of the targeting sequence (SEMENOVA *et al.* 2011; WIEDENHEFT *et al.* 2011; CRADICK *et al.* 2013; FU *et al.* 2013; HSU *et al.* 2013; MALI *et al.* 2013b; PATTANAYAK *et al.* 2013; YANG *et al.* 2013; CHO *et al.* 2014). It is possible to mutate one nuclease domain rendering Cas9 capable of cleaving only one DNA strand. This nickase version of Cas9 can be targeted in pairs to increase specificity or singly to generate DNA nicks capable of promoting HDR but not NHEJ, though at a cost of reduced flexibility and efficiency (CONG *et al.* 2013; HSU *et al.* 2013; MALI *et al.* 2013a). To balance efficiency and specificity, we developed an online tool for identifying optimal CRISPR targets. While several online resources have recently been developed for identifying CRISPR targets, they are not available for Drosophila, do not make the functionally important distinction between mismatches in seed and distal sequences in their algorithms, and/or do not provide the transparent off-target analysis and reporting of CRISPR Optimal Target Finder (HSU *et al.* 2013; REN *et al.* 2013). Our tool is designed to provide the user with maximum understanding of and control over the evaluation criteria and to supply the information necessary to select optimal CRISPR targets for their specific genome engineering project. When deleting a gene, there are likely to be many suitable CRISPR targets making target selection straightforward. However, for applications such as inserting an in-frame tag at a precise location, employing a target site as close to the desired change as possible is highly advantageous and might warrant, for example, selecting an optimally located target with a potential for off-target cleavage over a less ideally located target with no predicted off-target sites. Through a comprehensive analysis of off-target cleavage, we demonstrate

that CRISPR Optimal Target Finder facilitates complex genome engineering with high specificity.

The development of approaches for highly specific and efficient CRISPR/Cas9-mediated HDR makes a broad range of precise genome modifications possible anywhere in the genome. Coupled with the ease of generating unique targeting gRNAs, these advances significantly expand the prospects for complex genome engineering in Drosophila.

Acknowledgements

We thank Melissa Harrison, Jill Wildonger and members of the O'Connor-Giles lab for their collaboration in reagent development, invaluable discussions and critical comments on the manuscript. We are grateful to Ed O'Connor-Giles for guidance on software development, and appreciate comments from Peter Soba and members of the O'Connor-Giles and Wildonger labs that improved CRISPR Optimal Target Finder. We also thank the Bloomington Drosophila Stock Center for fly stocks. This work was funded by startup funds from the University of Wisconsin and a grant from the National Institute of Neurological Disorders and Stroke, National Institutes of Health (R01 NS078179) to KOCG. Plasmids and transgenic fly lines described here are available from the Harrison, Wildonger and O'Connor-Giles labs through the non-profit distributor Addgene and the Bloomington Drosophila Stock Center, respectively. Detailed protocols and reagent information are available at flyCRISPR.molbio.wisc.edu. All software code is available upon request.

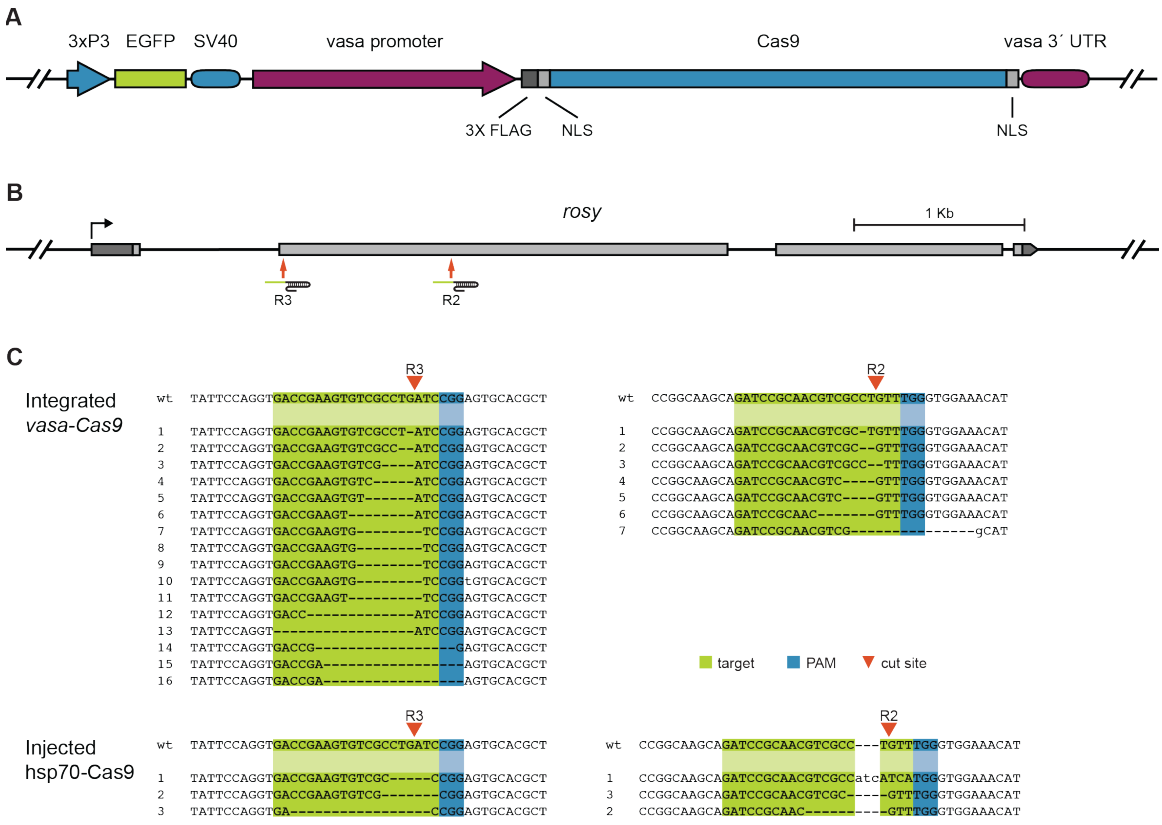

Figure 1. Germline expression of Cas9 efficiently generates DSBs repaired by NHEJ. (A) A *Streptococcus pyogenes*-derived Cas9 codon optimized for eukaryotic expression (Cong *et al.* 2013), was placed under the control of the Drosophila *vasa* promoter and 3′ UTR and incorporated via ΦC31to generate transgenic flies that express Cas9 in the germline. (B) Schematic of the *rosy* locus indicating R2 and R3 gRNA target sites. The predicted cleavage site for each gRNA is indicated (red arrows). (C) Sequences of the independent *rosy* mutants generated by Cas9-induced DSBs. The predicted R2 and R3 cleavage sites are indicated with a red arrowhead. Three independent *rosy* mutants generated using the R3 gRNA have larger indels and for two we did not identify an indel at the cleavage site.

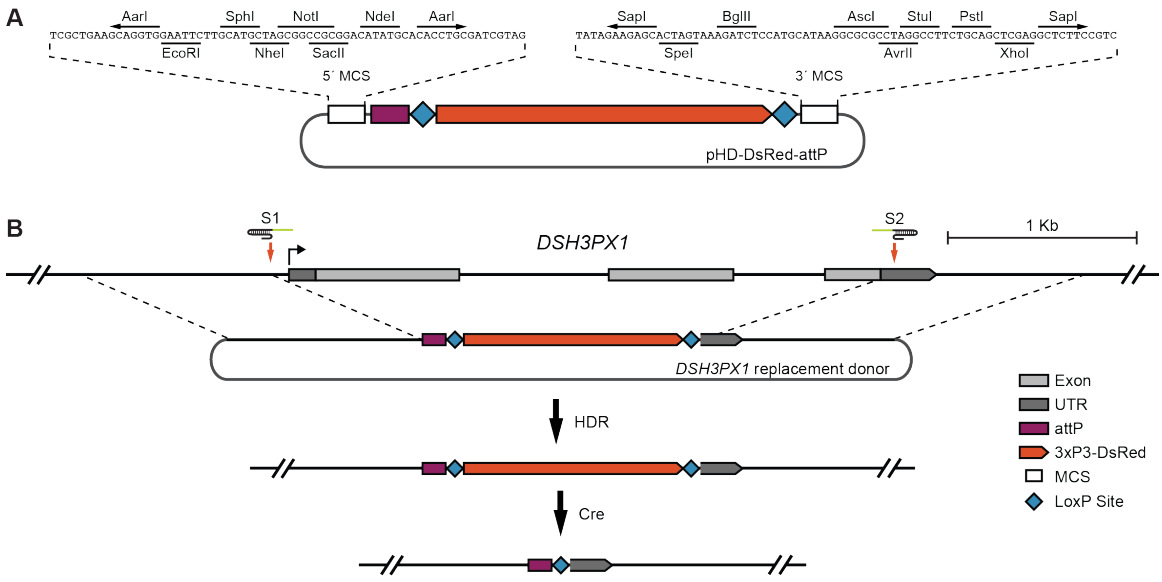

Figure 2. CRISPR-induced HDR with a dsDNA donor vector. (A) Diagram of the pHD-DsRed-attP vector. pHD-DsRed-attP contains an attP ΦC31 docking site for subsequent access to a targeted locus and a 3xP3-DsRed marker, which drives expression of DsRed in the eye, flanked by loxP recombination sites for its removal. Along with traditional restriction sites, the multiple cloning sites are flanked by type IIS restriction sites, AarI and SapI respectively, to facilitate seamless incorporation of homology arms. For compatibility with donors generated for ends-out homologous recombination, the multiple cloning site conserves as many restriction sites as possible from pGX-attP (HUANG *et al.* 2009).(B) Schematic of the *DSH3PX1* locus and the HDR strategy utilized to replace the gene with an attP docking site. The S1 and S2 target sites are indicated (red arrows). Homology arms of approximately 1 kb immediately flanking the cleavage sites were cloned into pHD-DsRed-attP.

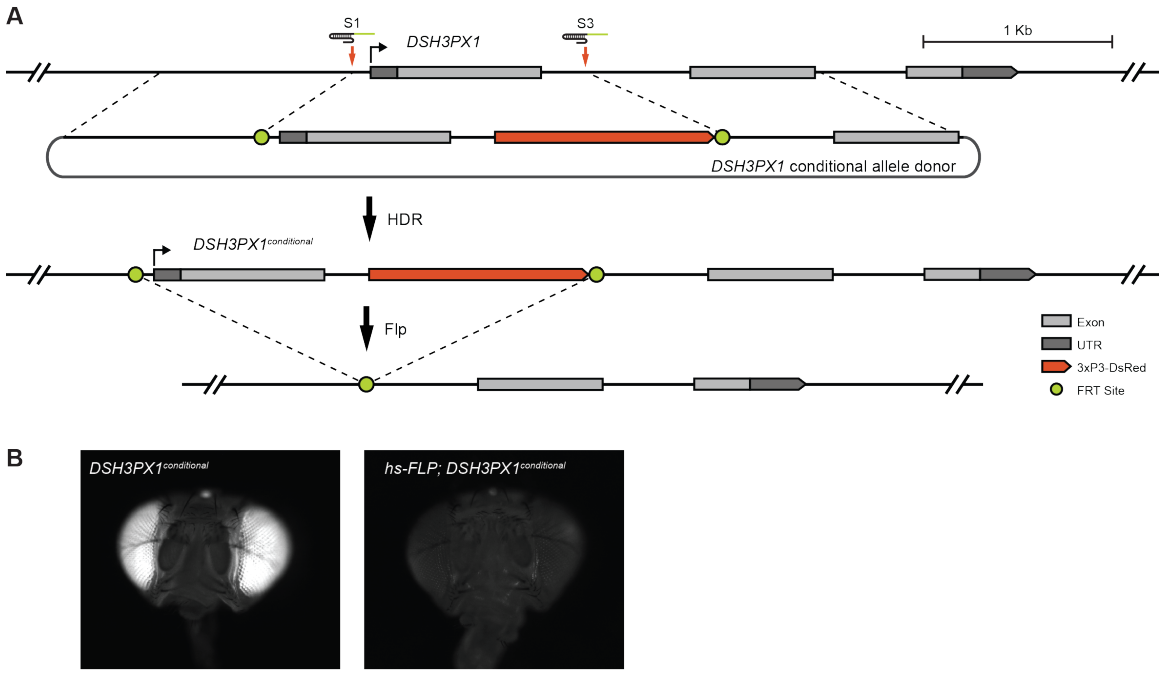

Figure 3. dsDNA donors can be used to generate positively marked conditional alleles. (A) Schematic of the *DSH3PX1* locus including the S1 and S3 target sites flanking the first exon. Homology arms of about 1 kb flank the target sites. The donor vector also contains the first exon and is designed to incorporate an FRT site near the S1 target site as well as the 3xP3-DsRed marker and a second FRT sequence near the S3 target site. In the resulting *DSH3PX1*^{conditional} allele, the first exon and DsRed marker can be removed by Flp recombinase. (B) Flies in which the first exon of *DSH3PX1* has been removed can be identified by the loss of 3xP3-DsRed in the eye.

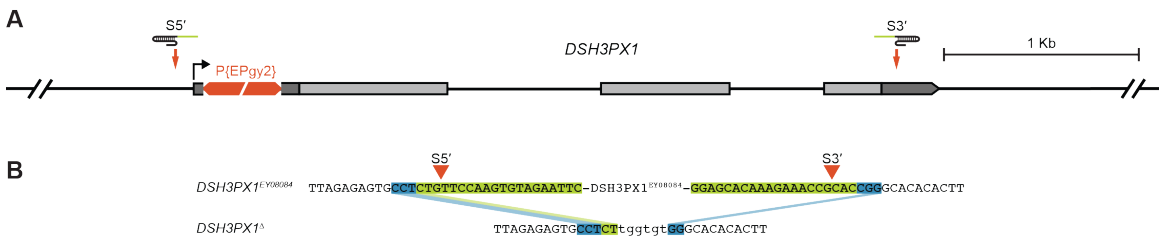

Figure 4. CRISPR-induced deletions can be identified through negative screening. (A) Schematic of the *DSH3PX1*^{EY08084} locus containing a P element marked by w^{+mC}. Target sites for gRNAs S1 and S2 are indicated (red arrows). (B) Sequence alignment of the targeted deletion reveals breakpoints and imprecise repair at the predicted cut sites (red arrowheads) confirming a 14.2 kb deletion that includes the P element.

A

gRNA	High NGG	High NAG	Maximum NGG	Maximum NAG	Independant F1s analyzed
R2	0	1	4	6	3/7
R3	0	0	2	4	6/21
S1	0	2	3	9	14/14
S2	0	1	5	9	12/12
S3	6	1	21	37	2/2

B

gRNA	off target #	off-target site sequence distal | proximal	PAM	# mismatches proximal	# mismatches distal	indel frequency
R2	-	GATCCGCA|ACGTCGCCTGTT	TGG	-	-	-
	1	cATtgtCA|AtGTCGCCTGTT	TGG	1	4	0/3
	2	cAcCgGCg|ACGTgGCCTGTT	TGG	1	4	0/3
	3	cATtCGtc|ACGTCGCaTGTT	TGG	1	4	0/3
	4	GggCCaaA|ACGTCGCCTGTc	GGG	1	4	0/3
	5	acatgGCc|ACGTCGCCTGTT	GAG	0	6	0/3
R3	-	GACCGAAG|TGTCGCCTGATC	CGG	-	-	-
	1	aACtaAtG|TGcCGCCTGATC	CGG	1	4	0/6
	2	cACaaAAc|TGTCGCCTGgTC	TGG	1	4	0/6
S1	-	GAATTCTA|CACTTGGAACAG	AGG	-	-	-
	1	GAcaaCTA|CAaTTGGAACAG	CGG	1	3	0/14
	2	tAtTatTA|CACaTGGAACAG	TGG	1	4	0/14
	3	GAtTgggA|CACTTGGcACAG	TGG	1	4	0/14
	4	aggTgaac|CACTTGGAACAG	TAG	0	7	0/14
	5	tggcgaag|CACTTGGAACAG	TAG	0	8	0/14
S2	-	GGAGCACA|AAGAAACCGCAC	CGG	-	-	-
	1	GGtGaACg|AAaAAACCGCAC	AGG	1	3	0/12
	2	aGAcgACA|AAGcAACCGCAC	AGG	1	3	0/12
	3	atAGaAaA|AAGcAACCGCAC	TGG	1	4	0/12
	4	GGcGgcCc|AAGAAACCGtAC	AGG	1	4	0/12
	5	GGAGCcCA|AAGgAACCGCAg	GGG	2	1	0/12
	6	tacagcat|AAGAAACCGCAC	AAG	0	8	0/12
S3	-	TCAACTCA|GTTGGAAAATGT	TGG	-	-	-
	1	aaActTCA|GTTGGAAAATGT	TGG	0	4	0/2
	2	TgAAaaat|GTTGGAAAATGT	TGG	0	5	0/2
	3	ggAAaatg|GTTGGAAAATGT	TGG	0	6	0/2
	4	ctgAaaat|GTTGGAAAATGT	TGG	0	7	0/2
	5	aacttatt|GTTGGAAAATGT	TGG	0	8	0/2
	6	gagggggt|GTTGGAAAATGT	CGG	0	8	0/2
	7	cggtgggA|GTTGGAAAATGT	CAG	0	7	0/2

Figure 5. Off-target cleavage analysis. (A) Comparison of the off-target site profiles of the gRNAs used in this study. The numbers of NGG and NAG-adjacent off-target cleavage sites identified by CRISPR Optimal Target finder using both high and maximum stringency criteria are represented as are the number of independent F1s investigated out of the total number of founders generated using each guide. The off-target sites analyzed in this study are indicated in green and listed below. (B) All off-target sites surveyed in this study. The CRISPR target sequence is provided for each guide. Potential off-target sequences are listed with mismatches to the target sequence indicated by lowercase red letters. Non-canonical NAG PAM sites are underlined. The location (proximal or distal) and number of mismatches is also indicated. Note that for gRNA S3, a G nucleotide was added to the 5′ end to improve transcription efficiency from the U6 promoter. The frequency of off-target indel formation is indicated as the number of independent events out of the number of independent lines analyzed. Note that for gRNA S3, a G nucleotide was added to the 5′ end to improve transcription efficiency from the U6 promoter.

References

Ahmad, K., and K. G. Golic, 1996 Somatic reversion of chromosomal position effects in Drosophila melanogaster. Genetics 144**:** 657-670.

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

Baena-Lopez, L. A., C. Alexandre, A. Mitchell, L. Pasakarnis and J. P. Vincent, 2013 Accelerated homologous recombination and subsequent genome modification in Drosophila. Development 140**:** 4818-4825.

Bassett, A. R., and J. L. Liu, 2014 CRISPR/Cas9 and genome editing in *Drosophila*. Journal of Genetics and Genomics.

Bassett, A. R., C. Tibbit, C. P. Ponting and J. L. Liu, 2013 Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4**:** 220-228.

Bellen, H. J., R. W. Levis, G. Liao, Y. He, J. W. Carlson *et al.*, 2004 The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics 167**:** 761-781.

Beumer, K. J., and D. Carroll, 2014 Targeted genome engineering techniques in Drosophila. Methods.

Beumer, K. J., J. K. Trautman, A. Bozas, J. L. Liu, J. Rutter *et al.*, 2008 Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105**:** 19821-19826.

Beumer, K. J., J. K. Trautman, K. Mukherjee and D. Carroll, 2013 Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. G3 (Bethesda).

Bibikova, M., M. Golic, K. G. Golic and D. Carroll, 2002 Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161**:** 1169-1175.

Bischof, J., R. K. Maeda, M. Hediger, F. Karch and K. Basler, 2007 An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104**:** 3312-3317.

Chiu, H., H. T. Schwartz, I. Antoshechkin and P. W. Sternberg, 2013 Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. Genetics 195**:** 1167-1171.

Cho, S. W., S. Kim, Y. Kim, J. Kweon, H. S. Kim *et al.*, 2014 Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 24**:** 132-141.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science 339**:** 819-823.

Cradick, T. J., E. J. Fine, C. J. Antico and G. Bao, 2013 CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res 41**:** 9584-9592.

Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31**:** 822-826.

Gao, G., N. Wesolowska and Y. S. Rong, 2009 SIRT combines homologous recombination, site-specific integration, and bacterial recombineering for targeted mutagenesis in Drosophila. Cold Spring Harb Protoc 2009**:** pdb prot5236.

Gasiunas, G., R. Barrangou, P. Horvath and V. Siksnys, 2012 Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109**:** E2579-2586.

Golic, K. G., 2013 RNA-guided nucleases: a new era for engineering the genomes of model and nonmodel organisms. Genetics 195**:** 303-308.

Gratz, S. J., A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue *et al.*, 2013a Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194**:** 1029-1035.

Gratz, S. J., J. Wildonger, M. M. Harrison and K. M. O'Connor-Giles, 2013b CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. Fly (Austin) 7.

Hampel, S., P. Chung, C. E. McKellar, D. Hall, L. L. Looger *et al.*, 2011 Drosophila Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. Nat Methods 8**:** 253-259.

Hsu, P. D., D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann *et al.*, 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31**:** 827-832.

Huang, J., W. Zhou, W. Dong, A. M. Watson and Y. Hong, 2009 From the Cover: Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. Proc Natl Acad Sci U S A 106**:** 8284-8289.

Iseli, C., G. Ambrosini, P. Bucher and C. V. Jongeneel, 2007 Indexing strategies for rapid searches of short words in genome sequences. PLoS One 2**:** e579.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337**:** 816-821.

Kondo, S., and R. Ueda, 2013 Highly improved gene targeting by germline-specific cas9 expression in Drosophila. Genetics 195**:** 715-721.

Liu, J., C. Li, Z. Yu, P. Huang, H. Wu *et al.*, 2012 Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics 39**:** 209-215.

Mali, P., J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner *et al.*, 2013a CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31**:** 833-838.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell *et al.*, 2013b RNA-guided human genome engineering via Cas9. Science 339**:** 823-826.

McVey, M., D. Radut and J. J. Sekelsky, 2004 End-joining repair of double-strand breaks in Drosophila melanogaster is largely DNA ligase IV independent. Genetics 168**:** 2067-2076.

Pattanayak, V., S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna *et al.*, 2013 High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31**:** 839-843.

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

Sebo, Z. L., H. B. Lee, Y. Peng and Y. Guo, 2013 A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly (Austin) 8.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra *et al.*, 2011 Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A 108**:** 10098-10103.

Siegal, M. L., and D. L. Hartl, 1996 Transgene Coplacement and high efficiency site-specific recombination with the Cre/loxP system in Drosophila. Genetics 144**:** 715-726.

Wiedenheft, B., E. van Duijn, J. B. Bultema, S. P. Waghmare, K. Zhou *et al.*, 2011 RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci U S A 108**:** 10092-10097.

Yang, H., H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi *et al.*, 2013 One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154**:** 1370-1379.

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

Section 3

Chapter 5

Employing novel genome engineering technologies to understand the molecular determinants of neurotransmitter release probability.

Abstract

Improved genome engineering methodologies have opened the door for more precise dissection of active zone proteins and their role in regulating neurotransmitter release. The ability to manipulate and tag synaptic genes in their endogenous loci affords us unprecedented opportunities to study synapse formation and function. One highly attractive application of CRISPR/Cas9-mediated genome engineering is the endogenous tagging of target genes with minimal disruption the locus to maintain normal expression. I have focused on tagging the synaptic Ca^{2+} channel *Cacophony* with the goal of understanding how it is localized to active zones and how channel levels relate to release probability. In this chapter I will detail my progress in creating CRISPR-based tools for endogenous tagging. I will describe improvements to my original donor vector designs, and their use for inserting a tag into Cacophony, as well as the use of a PiggyBac transposon-based system for scarless removal of positive screening markers. Together these tools facilitate the rapid generation and recovery of minimally disrupted, endogenously tagged genes. Using these techniques I have generated *GFP-Cacophony*, a tool I will use in future studies of how neurons establish and maintain the distinct neurotransmitter release parameters of individual active zones.

Introduction

The human brain contains an estimated 200 billion neurons each capable of making many thousands of synaptic connections with other neurons. These connections, where information is passed from one neuron to another, are the basic building blocks of a functional nervous system. Throughout the life of a neuron, each of its connections must be independently established and maintained. The neuron must also have the ability to independently modulate each synapse to regulate synaptic strength in response to varying activity and inputs. The mechanisms involved in developing, maintaining and modulating synaptic connections underlie learning, memory and all basic responsibilities of a functioning nervous system. Interest in elucidating these mechanisms has driven my research; sometimes in expected ways, and sometimes in unexpected ways.

Our mechanistic studies of *Fife* point to a role in coupling Ca^{2+} channels and synaptic vesicles (SVs) to promote high probability neurotransmitter release. (Chapters 1 and 2; Bruckner *et al.*, 2012). Reliable neurotransmitter release requires the precise organization of SVs and Ca^{2+} channels in nanodomains. Synaptic strength can be modulated by fine-tuning the coupling between channels and synaptic vesicles to modulate the probability of release. A single *Drosophila* NMJ can contain hundreds of synapses, usually formed by a single motor neuron and its target muscle (for review see Harris and Littleton, 2015). This arrangement provides a platform for studying how release probabilities are generated and modulated at individual active zones (AZs). Interestingly, AZs of *Drosophila* NMJs display highly variable release probabilities even across a single motor neuron. When measured using live imaging with the genetically encoded Ca^{2+} indicator GCaMP, release probabilities at individual AZs can range from 1% to 50% within the same motor neuron terminal (Peled *et al.*, 2011). About 22% of AZs

are not involved in action-potential evoked (synchronous) release at all, but rather participate only in spontaneous (asynchronous) SV fusion (Melom *et al.*, 2013). Intriguingly, AZs with high evoked release probabilities were more likely to have high levels of the AZ marker Bruchpilot, while release sites involved primarily in spontaneous release had lower levels of Bruchpilot (Peled *et al.*, 2014). At mammalian synapses, release probability also scales with the size of the AZ (Holderith *et al.*, 2012). This suggests that modulating the composition of proteins at that AZ is a conserved mechanism for establishing the unique release properties of individual AZs.

Cacophony encodes the pore-forming a1 subunit of a conserved family of voltage-gated Ca^{2+} channels (VGCCs) primarily responsible for synaptic transmission (Smith *et al.*, 1996; Kawasaki *et al.*, 2000). *Cacophony* is the *Drosophila* homolog and lone representative of P/Q- and N-type VGCCs (Smith *et al.*, 1996; Rieckhof *et al.*, 2003). P/Q- and N-type VGCCs interact with a number of AZ proteins to achieve the tight coupling of Ca^{2+} influx with SVs that enables vesicle fusion on the millisecond time scale (Mochida *et al.*, 1996; Rettig *et al.*, 1997; Sheng *et al.*, 1998). Models of Ca^{2+} influx and release machinery suggest that coupling distance is a key factor in regulating release probability. These simulations suggest that coupling distance changes of only 100 nm can alter the release probabilities from nearly 100% to nearly 0% (Chen *et al.*, 2015). Such precise spatial organization of Ca^{2+}channels is likely regulated by the cytomatrix of the AZ (CAZ). Indeed, a number of CAZ proteins including Bruchpilot, RIM and RIM-Binding Protein have been implicated in organizing and clustering Ca^{2+} channels in *Drosophila*, worms and mammals (Kittel *et al.*, 2006; Liu *et al.*, 2011; Graf *et al.*, 2012). Because Ca^{2+} channels are likely a critical determinant of the release characteristics of an AZ, the ability to accurately measure Ca^{2+} channel levels and localization is vital for studying AZ probability of release. Unfortunately, the standard reagent to analyze Ca^{2+}

channel localization is transgenic, C-terminally tagged, *UAS-Cacophony-EGFP*, which is problematic for a couple of reasons. Misexpression of *UAS-Cacophony-EGFP* may cause phenotypic abnormalities, particularly at the detailed level we are investigating. Moreover, the C-terminal GFP disrupts the C-terminal PDZ-binding domain of *Cacophony*, which is likely involved in binding PDZ domain-containing proteins such as Fife and RIM. As such, I sought to develop a better reagent to monitor the clustering of Ca^{2+} channels at AZs by endogenously tagging *Cacophony* using CRISPR-Cas9.

To this point, we had used the CRISPR-Cas9 system to make a number of modifications to the *Drosophila* genome including stochastic indels, defined deletions, gene replacements and conditional alleles (Gratz *et al.*, 2013; Gratz *et al.*, 2014). We had not yet attempted to insert tags into endogenous loci. Endogenous tags have numerous benefits over more traditional transgene-based approaches. Endogenously tagged gene products are more likely to reflect the natural expression pattern and levels of the target gene, especially compared to transgene strategies that rely on binary expression methods such as the GAL4/UAS system that regularly use promoters from completely different genes and often result in significant overexpression of the tagged gene product. Even in cases where the endogenous promoter is used instead of the Gal4/UAS system, it is impossible to know if all of the enhancers and suppressors acting on that locus are included in the promoter region selected or what effect the insertion site of the transgene may have on expression. Phenotypic anomalies that can complicate interpretations of the native function of the gene are an additional concerning caveat of transgene approaches. Based on our previous genome engineering experiments, we determined that a strategy utilizing a dsDNA donor with a screenable marker would be the best approach for endogenous tagging. However, we wanted to develop a genome engineering strategy that enables the generation of specific modifications with minimal additional disruption of

the target locus. This would allow us to edit or insert exogenous sequences while minimizing the chance of altering the expression or regulation of the target gene. One common source of unintended modifications is the marker used for screening or selection of candidate engineering events. Previously, we have used an eye-expressed DsRed marker flanked by loxP sites for removal of the marker after identification (Chapter 4; Gratz *et al.*, 2014). This approach leaves a single 34-bp copy of the loxP site in the target locus after marker removal that could lead to inadvertent disruption of regulatory sequences. An alternative strategy would be to use an unmarked approach. In most cases, this would necessitate molecular screening to recover the targeted events, which is significantly more laborious. I therefore sought to develop a system for scarless removal of marker sequences to be used with our endogenous tagging strategy.

Results

Scarless Genome Engineering with PiggyBac Transposons

The PiggyBac mobile element is a member of a family of short-inverted-repeat type DNA transposons characterized by their extreme target site specificity for the tetranucleotide TTAA (Cary et al., 1989; Fraser et al., 1995). Upon insertion of a PiggyBac element, the target site TTAA sequence is duplicated such that TTAA sites immediately flank each end of the transposon. Upon excision, the duplicated TTAA sites are reduced back to a single TTAA site. The PiggyBac transposon, along with other members of its family, is unique among cut-and-paste transposons in its ability to invariably and precisely excise without leaving behind any residual inverted repeat or filler sequence (Fraser et al., 1996). For the purposes of scarless genome engineering, this means that a PiggyBac-based marker element can be placed in a naturally occurring TTAA site and PiggyBac transposase can be used to precisely excise the marker without leaving behind any residual sequences. I therefore generated a new donor vector with an eye-expressed DsRed marker cassette flanked by PiggyBac inverted repeat ends called pScarlessHD-DsRed (Figure 1).

pScarlessHD-DsRed can be used to generate any number of targeted edits or insertions. To use this donor vector, the desired modifications accompanied by homology arms are cloned into pScarlessHD-DsRed. The donor is designed such that homology-directed repair (HDR) results in the PiggyBac marker being copied into a nearby endogenous TTAA site. The TTAA site is duplicated in the donor vector, flanking each end of the PiggyBac element, mimicking the natural outcome of PiggyBac insertion so that the element can be properly excised by the PiggyBac transposase. After engineered events have been recovered using DsRed expression, the marker can simply be removed by crossing in a transgenic source of PiggyBac transposase, restoring the single TTAA site

present before HDR. To increase the utility of the PiggyBac marker cassette, I created a version with two attP recombination-mediated docking sites flanking the DsRed marker so that exogenous sequences could be inserted into the PiggyBac element through recombination-mediated cassette exchange. If the PiggyBac element is placed in a nearby intron, replacement cassettes harboring splice acceptors could be used to trap the expression pattern of the target gene, for example.

Tagging an Extracellular Loop of Cacophony

We decided to tag an extracellular loop within *Cacophony* using the pScarlessHD-DsRed vector. Extracellular tags provide the ability to resolve membrane-inserted channels at active zones from Ca^{2+} channels localized in intracellular compartments by using pH-dependent genetically encoded fluorophores or antibody staining without membrane permeabilization.

A number of studies have successfully utilized extracellular loop-tagged Ca^{2+} channels in cell culture (Figure 2; Watschinger *et al.*, 2008; Di Biase *et al.*, 2011; Cassidy *et al.*, 2014). While the requirements for function of a tagged gene expressed in cell culture are very different from the requirements in an intact organism, these studies provide insight into which Ca^{2+} channel extracellular loops are amenable to tagging both in terms of accessibility and possibly function. We chose to first target the extracellular loop between the transmembrane domain IIS5 and the pore-forming loop that is orthologous to the loop previously tagged by Di Biase *et al.* (2011) in which they were able to insert both epitope tags and the pH-dependent genetically encoded super-ecliptic pHluorin (SEP).

To edit this loop, we used a strategy analagous to the targeting strategy detailed in Figure 1. Two gRNA target sites were selected in the *Cacophony* locus; one near the insertion

site of the tag and another near a TTAA site in the intron immediately downstream of the targeted exon. Two donor vectors were also generated; one to insert a small tetracysteine (TC) tag of 36 nts and the other to insert SEP. Each donor vector, along with the two gRNA targeting vectors were co-injected into *vasa-Cas9* embryos. Based on incorporation of the PiggyBac-DsRed marker cassette, the efficiency of HDR was 11.3% for the TC donor and 11.0% for the SEP donor (Table 1). To molecularly confirm the candidate progeny, flanking PCRs were performed using primer sites within the PiggyBac-DsRed cassette and genomic primer sites outside of the homology regions contained within the donor vector. Interestingly, of the seven TC donor founders and nine SEP donor founders, only one founder of each produced progeny with the tag inserted (Table 1). Our results suggest that the intervening homology region allows for HDR events that insert the marker without also inserting tag (Figure 3). In fact, if recombination occurs in the intervening region, either the tag or the marker will be inserted into the target region separately. HDR efficiency is partially related to homology region length (Beumer *et al.*, 2013) suggesting that the longer the intervening region is, the more likely it is to be involved in repair. This suggests that the intervening region used for targeting the extracellular loop is too long for reliable insertion of both the marker and the tag and that selecting a marker insertion site closer to the tag insertion site could alleviate this issue.

After molecular confirmation of the TC and SEP insertions, the PiggyBac-DsRed marker was removed by crossing tagged lines to flies with ubiquitous expression of PiggyBac transposase. Progeny of the crosses invariably displayed mosaic DsRed expression in the adult eyes indicating mobilization of the PiggyBac element. DsRed negative flies were then recovered in the F2 generation and molecular tests confirmed the precise removal of PiggyBac-DsRed (data not shown). Both tag insertions were lethal prior to marker

removal; however, after precise removal of the DsRed marker, the TC tag allele was viable while the SEP insertion remained lethal. This suggests that the PiggyBac-DsRed cassette is disruptive even when placed in a nearby intron. These results also suggest that this particular extracellular loop is only tolerant of smaller, epitope-sized insertions. Unfortunately, neither the TC allele nor the SEP allele could be visualized by either live imaging or conventional fixation and staining (data not shown). In the case of the lethal SEP tag, it is possible that the tagged gene product is not folded, transported and/or localized properly. For the TC-tagged allele, it is possible that the tagged loop is inaccessible to the biarsenical fluorescent dyes used to label TC tags. Further complicating analysis of the localization of the TC allele, the biarsenical fluorescent dyes had very high background staining in our NMJ preparations, limiting the utility of these reagents.

Improved Endogenous Tagging Donor Vector Design

Based on our results targeting an extracellular loop of *Cacophony* for tag insertion I developed an improved strategy for endogenous tagging. As noted previously, the strategy of placing the marker in a nearby intron necessitated the addition of intervening homology between the tag and the marker, making it possible to recover HDR events where the marker, but not the tag, had been inserted into the target locus. In fact, in the case of *Cacophony* extracellular loop targeting, alleles with the tag inserted were a minority of the recovered events. We have used the strategy of placing the marker in a nearby intron at other loci for the purposes of endogenous tagging. In these experiments the intervening homology are region has ranges in size from 196 to 876 bp, and in all but one case we recovered alleles without the tag inserted indicating that failure to insert the tag is a general problem with this strategy. Using this strategy I noticed a second type of unintended HDR event in which the entirety of the donor vector, including the backbone,

was incorporated into the target locus. This type of repair event has been noticed by other groups and, in some targeting experiments, alleles with the donor backbone inserted can make up a majority of the recovered HDR events (Yu *et al.*, 2014). While not specifically characterized in the *Cacophony* extracellular loop tag candidates, a small number of individual F2 candidates failed to produce a PCR amplicon when using primers designed to amplify flanking regions, which could be caused by backbone incorporation. Ruling out backbone incorporation is not trivial, as it requires more difficult long amplification reactions spanning the entire region, donor backbone specific reactions relying on negative results, or more laborious experiments like Southern blots.

To circumvent both of these concerns, I designed a new set of vectors for endogenous tagging with a number of advantages over the previous strategy. First, instead of separating the tag sequence and the marker, the PiggyBac-DsRed cassette is placed within the tag sequence. Specifically, a TTAA site was placed in the flexible linker sequence separating the tag from the endogenous sequence so that the PiggyBac-DsRed could be placed within the linker. After precise excision of the PiggyBac element, the linker is reconstituted leaving behind only the in-frame tag (Figure 4). This approach includes both the tag and the marker in a single cassette such that they are reliably inserted together into the target locus. I also incorporated a secondary marker, GMR-white, into the backbone of the donor vector. GMR-white strongly expresses the eye pigment transporter gene *white* in the eyes on adult flies making them dark red in appearance. This allows for the rapid identification of HDR events resulting in the incorporation of the donor backbone into the target locus alleviating the need for molecular screening of such events. So far, I have created three versions of this vector, pScarlessHD-sfGFP-DsRed, pScarlessHD-3xFLAG-DsRed and pScarlessHD-2xHA-DsRed, but any tag could be used.

One possible advantage of this approach is it can be used to simultaneously generate a null allele. Before the PiggyBac-DsRed cassette is removed, it disrupts the coding sequence of the tagged locus. Depending on the location of the insertion site, this could constitute a null allele. This also makes it possible to achieve tissue-specific endogenous tagging. Due to the fact that the tagged locus is interrupted by the presence of the marker, the tagged gene product is only expressed in cells in which the marker has been removed. To take advantage of this, we can use cell-specific expression of the transposase to analyze the tagged gene in specific cell types or global expression to generate stochastic labeling of cells.

N-terminal GFP tagging of *Cacophony*

I targeted the N-terminus of *Cacophony* using the pScarlessHD-sfGFP-DsRed donor vector to insert superfolder GFP. A single unique gRNA target site was selected near the translation start site of *Cacophony*. The donor vector contained ~1 kb homology arms directly flanking the target cut site. The donor vector and the DNA-based gRNA vector were co-injected into *vasa-Cas9* embryos. Based on DsRed expression, 9 independent founders were identified out of 94 fertile crosses (9.6% HDR efficiency). One of the nine founders produced progeny with red eyes indicating incorporation of the donor vector backbone while the remaining eight founders produced white-eyed flies indicating clean insertion. Candidate alleles were confirmed molecularly using PCR amplification and sequencing. As expected, all contained the full tag sequence.

The PiggyBac-DsRed cassette was removed by crossing tagged line to PiggyBac transposase-expressing flies. As before, all F1 progeny of the correct genotype displayed mosaic DsRed expression indicating mobilization of the marker cassette. Flies without

DsRed expression were recovered in the F2 generation. *sfGFP-Cacophony^{Nterm}* flies were confirmed both molecularly and visually. As seen using neuronal over expression of *UAS-Cacophony-EGFP*, the *sfGFP-Cacophony^{Nterm}* signal significantly overlaps with Bruchpilot, a central component of electron-dense cytoplasmic projections at NMJ AZs and in the neuropil of the larval ventral ganglion (Figure 5). *sfGFP-Cacophony^{Nterm}* is a homozygous viable allele indicating that the tag insertion does not significantly interfere with the function of *Cacophony*.

To test the ability of our new tagging system to generate stochastic endogenous tagging, I crossed *sfGFP-Cacophony^{Nterm}* flies with PiggyBac-DsRed still inserted to *tubulin-PiggyBac transposase* flies. Third instar larval progeny were dissected and stained with anti-GFP. The ventral ganglion showed clusters of GFP positive cells bodies indicative of mosaic endogenous tagging (Figure 6). We have now generated UAS-PiggyBac transposase for cell-specific expression and targeted generation of mosaic animals. This transposase incorporates mutations that have been shown in yeast and mammalian cells to render the enzyme defective in transposon insertion (Li *et al.*, 2013; Yusa *et al.*, 2011). However, our results with wild-type transposase suggest that reinsertion of the marker cassette is not a significant concern.

Discussion and Future Directions

With *sfGFP-Cacophony*^{Nterm}, I can now follow Ca^{2+} channels live at endogenous synapses in an intact organism for the first time. This reagent provides us with a unique opportunity to study how Ca^{2+} channels are organized at AZs and how their organization contributes to the release characteristics of individual AZs. To analyze Ca^{2+} channel localization at endogenous AZs, I will image *sfGFP-Cacophony*^{Nterm} using super-resolution microscopy. By co-labeling with other well established AZ proteins like Brp, RIM and Fife I will be able to correlate levels of AZ constituents with levels of Cacophony and build a model of endogenous Ca^{2+} channel localization in relation to the rest of the cytomatrix. Beyond light level imaging, *sfGFP-Cacophony*^{Nterm} is also a powerful tool for ultrastructure analysis of AZ architecture. Using immunoelectron microscopy, we will be able to analyze Ca^{2+} channel localization on the nanometer scale within AZs providing us with an ultrastructural measurement of coupling.

In addition to building a model of Ca^{2+} channel localization at wild-type AZs, we can analyze *sfGFP-Cacophony*^{Nterm} levels and localization at mutant AZs. *Fife*, *Rim*, *Rbp*, and *brp* mutant all exhibit defects in localizing *UAS-Cacophony-EGFP*, but endogenous Ca^{2+} channel levels have never been able to be measured before now. Using *sfGFP-Cacophony*^{Nterm} I will be able to get a more accurate measurement of Ca^{2+} channel levels at mutant AZs providing us with the most precise analysis of the role of the CAZ in VGCC clustering to date.

sfGFP-Cacophony^{Nterm} provides a unique tool for studying neurotransmitter release probability. Using sfGFP-Cacophony^{Nterm} I will be able to directly measure the Ca^{2+} channel content at each AZ and correlate these levels to activity to determine how Ca^{2+} channel clustering can regulate release probability. To accomplish this, I will generate

stocks with both sfGFP-Cacophony^{Nterm} and GCaMP so that I can simultaneously monitor Ca^{2+} channel levels and release dynamics at individual AZs. Because both sfGFP and GCaMP emits in the green spectrum I will also make a red fluorescent version of Cacophony^{Nterm} for use in these studies. As an alternative, I will also pair sfGFP-Cacophony^{Nterm} with R-CaMP2, a new a high-affinity red genetically encoded calcium indicator (Inoue *et al.*, 2015). Using these tools I can determine how different members of the CAZ work together to regulate AZ organization and Ca^{2+} channel levels to modulate release probability during synaptic plasticity.

Our improved toolset for scarless tagging and gene editing streamlines the process of using HDR, opening the door increased throughput engineering. A long-term goal of our lab is to systematically dissect the proteins that make up the CAZ to gain a more comprehensive understanding of how this complex network of proteins organize and regulate the molecular machinery underlying vesicle release. One factor contributing to the complexity of the CAZ is the highly interconnected nature of the interactions between its members and the fact that many of the CAZ constituents share common, but non-redundant, functions. A key way to disentangle the unique contributions of each member of the CAZ is to gain a better spatial and temporal understanding of their functions through a combination of live, super-resolution and ultrastructural imaging. Our endogenous tagging approaches will generate the reagents necessary to analyze the localization of CAZ members at subcellular and sub-AZ resolution. Further, these reagents will be an excellent resource for biochemical dissection of AZ function, and dynamic analysis of protein levels and localization if synaptic plasticity.

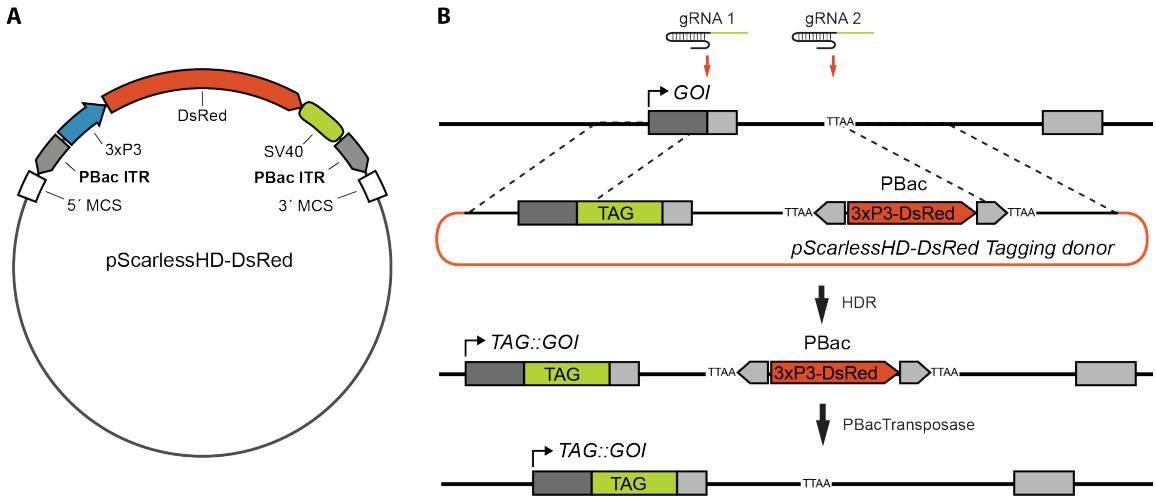

Figure 1. Scarless tagging strategy using pScarlessHD-DsRed. (A) Schematic of pScarlessHD-DsRed showing eye expressed 3xP3-DsRed marker flanked by the inverted terminal repeats of PiggyBac transposable elements (PBac ITR). The PiggyBac-DsRed cassette is bordered on each side by multiple cloning sites (MCS) for restriction enzyme cloning of homology regions. (B) Diagram of a general tagging strategy using pScarlessHD-DsRed. Two gRNA target sites are selected; one close to the insertion site of the tag and the other close to a nearby endogenous TTAA site were the PBac-DsRed will be inserted. The donor vector should contain ~1 kb homology arms directly flanking each target site. In this example the 5' homology arms contains the tag sequence and additional homology between the tag insertion site and the TTAA site. In the resulting allele, the PBac-DsRed can be scarlessly removed with PiggyBac transposase.

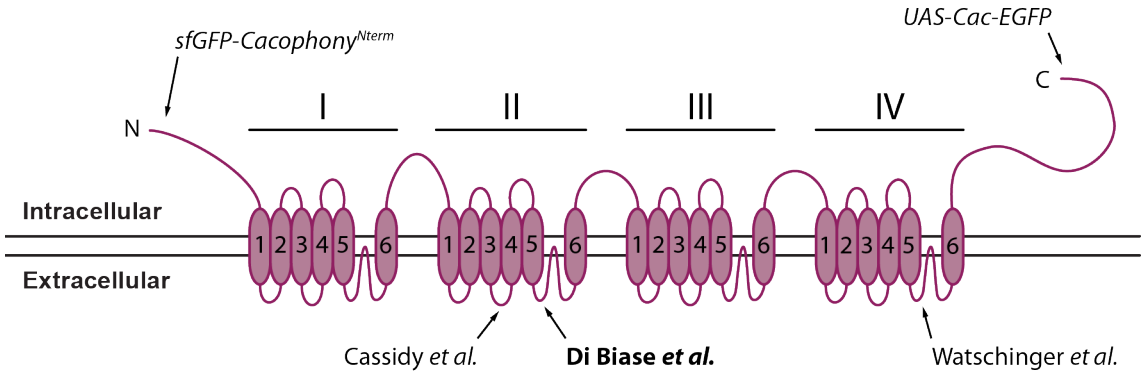

Figure 2. Tagging locations in the *Cacophony* locus. A diagram of *Cacophony* displaying the four domain structure (I-IV) and 24 transmembrane domains that make up the pore-forming subunit of the Ca^{2+} channel. Tag locations of the C-terminal *UAS-Cac-EGFP*, the extracellular loop tagging sites analogous to those used by Cassidy *et al.* (2014), Di Biase *et al.* (2011), and Watschinger *et al.* (2008), as well as the N-terminal site used to generate *sfGFP-Cacophony*^{Nterm} are shown. The extracellular loop targeted in this study is analogous to the loop targeted by Di Biase *et al.* (bold, 2011).

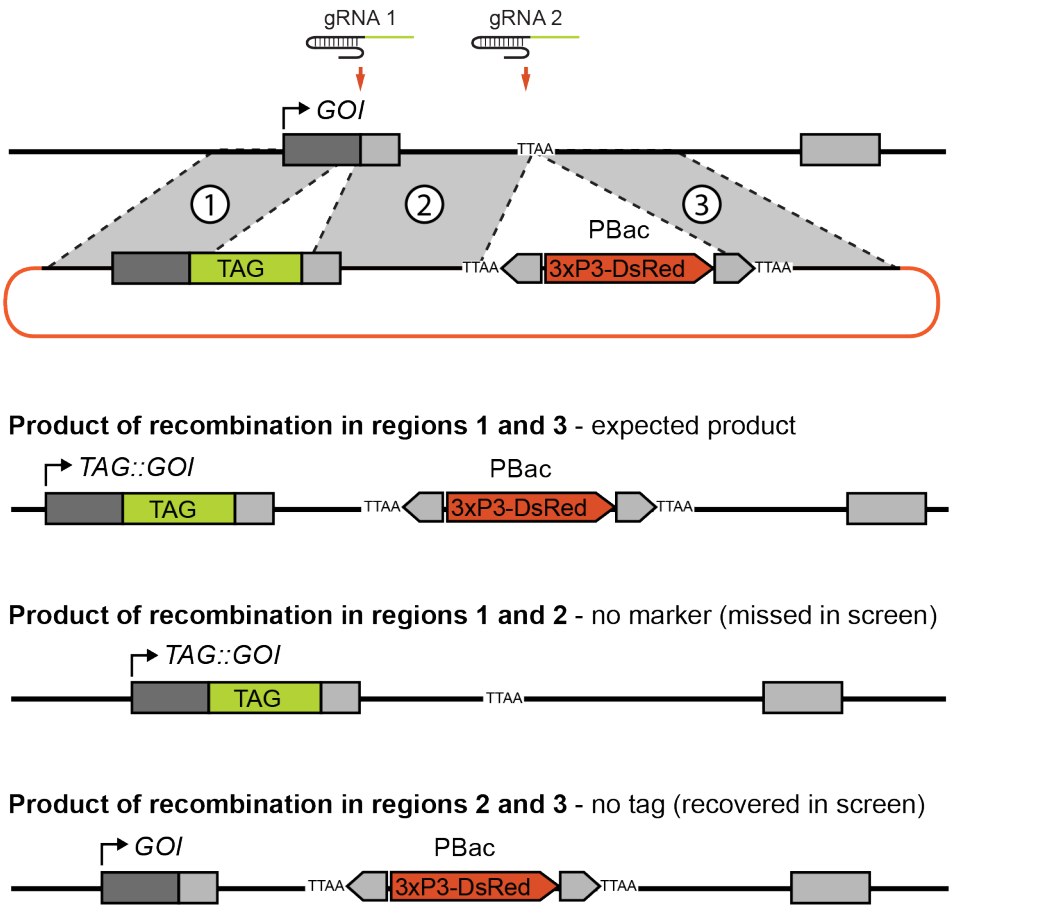

Figure 3. Results of different recombination events. When the PBac-DsRed cassette is placed in a nearby intron, the tag and the marker are separated by an intervening homology region (homology region 2) in addition to the homology arms flanking the target sites (homology regions 1 and 3). The expected allele, containing both the tag and the marker, requires homology directed repair involving regions 1 and 3. If regions 1 and 2 are involved in repair, the tag is inserted without the marker. These events are missed because candidate events are screened for using the DsRed-based marker. Alternatively, if homology regions 2 and 3 are utilized in the repair event, the marker is inserted without the corresponding insertion of the tag. These events are recovered in the screen.

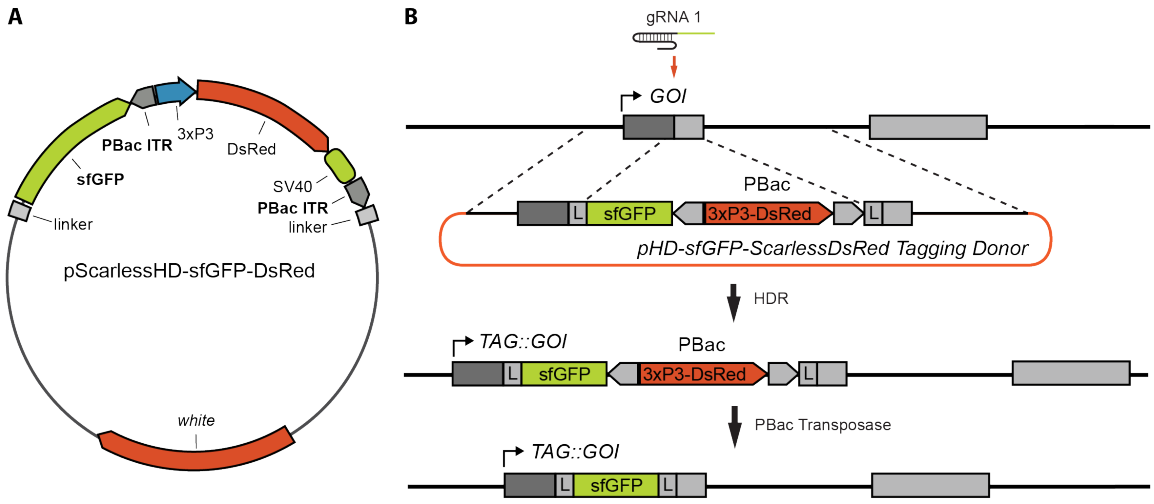

Figure 4. Scarless tagging strategy using pScarlessHD-sfGFP-DsRed. (A) A diagram of the pScarlessHD-sfGFP-DsRed donor vector. This vector contains superfolder GFP flanked by two linker sequences encoding 4 repeats of GlyGlySer to increase the flexibility between the tag and the endogenous protein. The PBac-DsRed is located in a TTAA site engineered into the 3' linker such that removal of the marker reconstitutes the complete linker sequence. A *white* gene based marker is placed in the backbone of the donor vector to serve as a screenable indicator of HDR events that result in the incorporation of backbone sequences. (B) A schematic of the general tagging strategy using pScarlessHD-sfGFP-DsRed. Because the tag and marker are contained within and single insertion cassette, a single gRNA target site is selected near the intended insertion site. Homology arms of ~1 kb flanking the DSB are cloned into the donor vector. The resulting tagged locus is disrupted by the PBac-DsRed until it is removed by PiggyBac transposase.

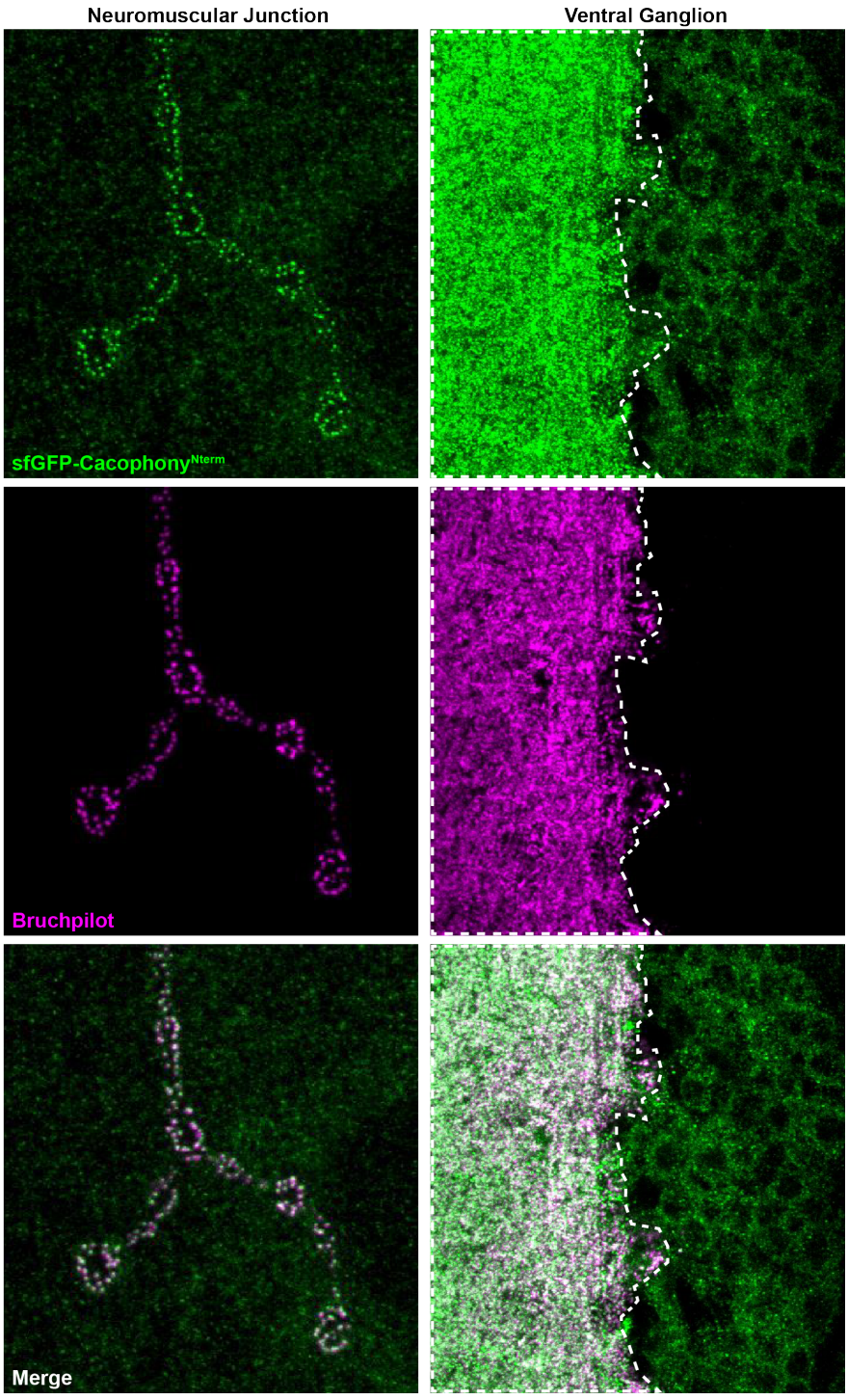

Figure 5. sfGFP-Cacophony^{Nterm} localizes to AZs. Confocal Z-projections of NMJ 4 and the larval ventral ganglion colabeled with antibodies to GFP and Bruchpilot in *sfGFP-Cacophony^{Nterm}* larvae. sfGFP-Cacophony^{Nterm} clusters at Bruchpilot labeled AZs at both the NMJ and the neuropil of the ventral ganglion (outlined).

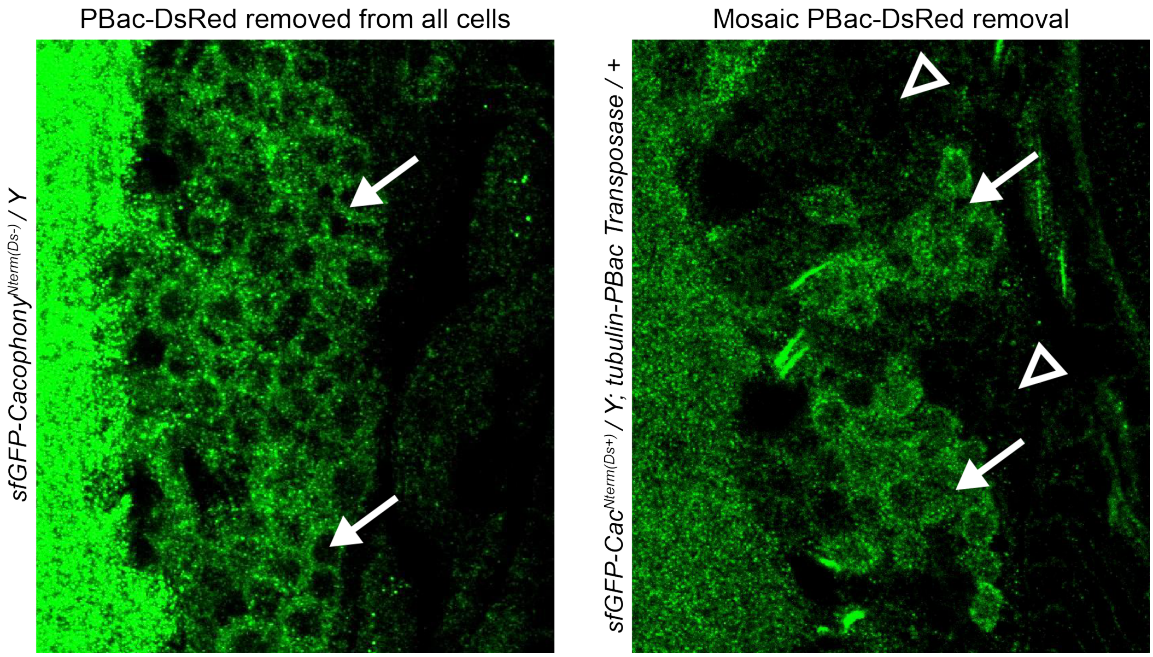

Figure 6. Mosaic expression of sfGFP-Cacophony^{Nterm}. Confocal Z-projections of the ventral ganglion of *sfGFP-Cacophony$^{Nterm(DsRed-)}$/Y* and *sfGFP-Cacophony$^{Nterm(DsRed+)}$/Y; tubulin-PiggyBac Transposase* larvae. After the PBac-DsRed cassette has been removed from the allele *sfGFP-Cacophony^{Nterm}* is expressed in the cell bodies of neurons in the ventral ganglion (arrows). *sfGFP-Cacophony^{Nterm}* expression is disrupted in cells with PBac-DsRed is still inserted (open arrow heads).

Table 1. Germline transmission rates of *Cacophony* extracellular loop targeting

donor vector	fertile injectee crosses	DsRed producing founders	% founder rate	founders producing molecularly confirmed progeny	founders producing progeny w/o tag inserted
tetracysteine donor	62	7	11.29	1	6
super-ecliptic pHluorin donor	82	9	10.98	1	8

vasa-Cas9 flies injected with gRNAs and the indicated donr vector were crossed to w^{1118} and progeny screened for DsRed positive eyes. The percentage of injected flies producing one or more DsRed positive progeny (founders) is indicated. At least two progeny per founder were molecularly characterized to determine if the tag was also inserted into the target locus.

References

Beumer, K.J. et al., 2013. Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. *G3 (Bethesda, Md.)*.

Bibikova, M. et al., 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science,* 300(5620), p.764.

Bruckner, J.J. et al., 2012. Fife, a Drosophila Piccolo-RIM Homolog, Promotes Active Zone Organization and Neurotransmitter Release. *The Journal of neuroscience*, 32(48), pp.17048–17058.

Cary, L.C. et al., 1989. Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, 172(1), pp.156–169.

Cassidy, J.S. et al., 2014. Functional exofacially tagged N-type calcium channels elucidate the interaction with auxiliary α2δ-1 subunits. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), pp.8979–8984.

Chen, Z. et al., 2015. Ca2+ Channel to Synaptic Vesicle Distance Accounts for the Readily Releasable Pool Kinetics at a Functionally Mature Auditory Synapse. *The Journal of neuroscience*, 35(5), pp.2083–2100.

Di Biase, V. et al., 2011. Surface Traffic of Dendritic CaV1.2 Calcium Channels in Hippocampal Neurons. *The Journal of neuroscience*, 31(38), pp.13682–13694.

Fraser, M.J. et al., 1995. Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology*, 211(2), pp.397–407.

Gong, W.J., 2003. Ends-out, or replacement, gene targeting in Drosophila. *Proceedings of the National Academy of Sciences*, 100(5), pp.2556–2561.

Graf, E.R. et al., 2012. RIM Promotes Calcium Channel Accumulation at Active Zones of the Drosophila Neuromuscular Junction. *The Journal of neuroscience*, 32(47), pp.16586–16596.

Gratz, S.J. et al., 2013. Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics*, 194(4), pp.1029–1035.

Gratz, S.J. et al., 2014. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. *Genetics*, 196(4), pp.961–971.

Harris, K.P. & Littleton, J.T., 2015. Transmission, Development, and Plasticity of Synapses. *Genetics*, 201(2), pp.345–375

Holderith, N. et al., 2012. Release probability of hippocampal glutamatergic terminals scales with the size of the active zone. *Nature Publishing Group*, 15(7), pp.988–997.

Huang, J. et al., 2009. From the Cover: Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 106(20), pp.8284–8289.

Inoue, M. et al., 2015. Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. *Nat. Methods*, 12(1), pp.64–70.

Kawasaki, F., Felling, R. & Ordway, R.W., 2000. A temperature-sensitive paralytic mutant defines a primary synaptic calcium channel in Drosophila. *The Journal of neuroscience*, 20(13), pp.4885–4889.

Kittel, R.J. et al., 2006. Bruchpilot promotes active zone assembly, Ca2+ channel clustering, and vesicle release. *Science,* 312(5776), pp.1051–1054.

Li, X. et al., 2013. piggyBac transposase tools for genome engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 110(25), pp.E2279–87.

Liu, J. et al., 2012. Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *Journal of Genetics and Genomics*, 39(5), pp.209–215.

Liu, K.S.Y. et al., 2011. RIM-binding protein, a central part of the active zone, is essential for neurotransmitter release. *Science,* 334(6062), pp.1565–1569.

Melom, J.E. et al., 2013. Spontaneous and evoked release are independently regulated at individual active zones. *The Journal of neuroscience*, 33(44), pp.17253–17263.

Peled, E.S. & Isacoff, E.Y., 2011. Optical quantal analysis of synaptic transmission in wild-type and rab3-mutant Drosophila motor axons. *Nature Neuroscience*, 14(4), pp.519–526.

Peled, E.S., Newman, Z.L. & Isacoff, E.Y., 2014. Evoked and spontaneous transmission favored by distinct sets of synapses. *Current biology : CB*, 24(5), pp.484–493.

Rieckhof, G.E. et al., 2003. Presynaptic N-type calcium channels regulate synaptic growth. *The Journal of biological chemistry*, 278(42), pp.41099–41108.

Rong, Y.S. & Golic, K.G., 2000. Gene targeting by homologous recombination in Drosophila. *Science,* 288(5473), pp.2013–2018.

Sheng, Z.H., Westenbroek, R.E. & Catterall, W.A., 1998. Physical link and functional coupling of presynaptic calcium channels and the synaptic vesicle docking/fusion machinery. *Journal of bioenergetics and biomembranes*, 30(4), pp.335–345.

Smith, L.A. et al., 1996. A Drosophila calcium channel alpha1 subunit gene maps to a genetic locus associated with behavioral and visual defects. *The Journal of neuroscience*, 16(24), pp.7868–7879.

Watschinger, K. et al., 2008. Functional properties and modulation of extracellular epitope - tagged Ca V2.1 voltage-gated calcium channels. *Channels*, 2(6), pp.461–473.

Yu, Z, H Chen, J Liu, H Zhang, Y Yan, N Zhu, Y Guo, et al. 2014. "Various Applications of TALEN- and CRISPR/Cas9-Mediated Homologous Recombination to Modify the Drosophila Genome." Biology Open 3 (4): 271–80.

Yusa, K. et al., 2011. A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences*, 108(4), pp.1531–1536.

Appendix A

CRISPR RNA-Cas9-mediated genome engineering and the promise of designer flies on demand.

Scott J. Gratz[1], Jill Wildonger[2], Melissa M. Harrison[3] and Kate M. O'Connor-Giles[1, 4, 5*]

[1] Genetics Training Program, University of Wisconsin-Madison, Madison, WI 53706

[2] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[3] Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706

[4] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

[5] Laboratory of Cell and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706

This appendix is adapted from the publication:

Gratz, S J, J Wildonger, M M Harrison and K M O'Connor-Giles. 2013. "CRISPR/Cas9-Mediated Genome Engineering and the Promise of Designer Flies on Demand." *Fly* 7 (4): 249-255.

Abstract

The CRISPR/Cas9 system has attracted significant attention for its potential to transform genome engineering. We and others have recently shown that the RNA-guided Cas9 nuclease can be employed to engineer the Drosophila genome, and that these modifications are efficiently transmitted through the germline. A single targeting RNA can guide Cas9 to a specific genomic sequence where it induces double-strand breaks that, when imperfectly repaired, yield mutations. We have also demonstrated that two targeting RNAs can be used to generate large defined deletions and catalyze gene replacement by homologous recombination. Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have shown similar promise in Drosophila. However, the ease of producing targeting RNAs over the generation of unique sequence-directed nucleases to guide site-specific modifications makes the CRISPR/Cas9 system an appealingly accessible method for genome editing. From the initial planning stages, engineered flies can be obtained within a month. Here we highlight the variety of genome modifications facilitated by the CRISPR/Cas9 system along with key considerations for starting your own CRISPR genome engineering project.

Introduction

An efficient, reliable means for precisely modifying the genome in living cells is a long-standing goal of biomedical science. In the clinic, such a tool will enable gene therapies to repair damaged genes, while in the lab it will be used to selectively manipulate genomic elements to study their function. In Drosophila, multiple approaches for precise genome editing have been successfully developed[1-7]. However, their significant time and labor requirements have limited the widespread adoption of genome engineering techniques in Drosophila. The CRISPR/Cas9 system is poised to change this. We and others have recently demonstrated that within one month CRISPR/Cas9-mediated genome modifications can be efficiently generated in Drosophila and transmitted through the germline[8-10].

Endogenous CRIPSR RNA/Cas9 systems comprise a single polypeptide nuclease, Cas9, that is guided to target sites by a complex of two small RNAs – the CRISPR RNA (crRNA), which contains the targeting sequence, and a common trans-activating CRISPR RNA (tracrRNA)[11,12]. For use in genome engineering, the *S. pyogenes* system was simplified to two components through the generation of a chimeric RNA (chiRNA or guide RNA (gRNA)) comprising all critical crRNA and tracrRNA sequences[12] (**Fig. 1**). The two-component system requires only a single gRNA that recognizes a 20-nt target sequence next to a trinucleotide NGG protospacer adjacent motif (PAM) to direct Cas9-dependent cleavage of both DNA strands within the target sequence[12]. This elegantly simple system has recently been shown to efficiently generate mutations in mammalian cell lines, human stem cells, yeast, bacteria, mice, zebrafish, worms and flies[8-10,13-24]. The fact that this has all been accomplished over a period of months illustrates the adaptability of the CRISPR/Cas9 system.

Engineering diverse genome modifications with CRISPR

In the short time since it was adapted for use in flies, the CRISPR/Cas9 system has already been used to successfully generate a variety of complex genome modifications, and the possibilities for expanding its application are nearly limitless[8-10]. Here, we focus on the applications likely to be of broadest interest to the Drosophila community.

Knock-outs/Deletions. The induction of site-specific double-strand breaks (DSBs) in chromosomal DNA can yield mutations when breaks are inaccurately repaired by non-homologous end joining (NHEJ). This error-prone repair process can generate small insertions and deletions (indels) at the cleavage site that disrupt gene function (**Fig 2A**). Separately, we used four different gRNAs to target Cas9 to the *yellow* and *rosy* loci, and observed germline transmission of frame-shifting indels at each targeted site[9] (**Table 1**). Bassett et al. (2013) and Yu et al. (2013) expanded the number of sites targeted for mutation by Cas9-induced NHEJ, and observed efficient germline transmission of indels at nine of twelve targeted sites in eight genes.

While the generation of random indels can disrupt function, the precise deletion of genes or other genomic sequences provides a powerful tool for unambiguously elucidating their role (**Fig. 2B**). To precisely delete the *yellow* gene, we simultaneously targeted the locus with two gRNAs, one directed to the 5′ end and the other at the 3′ end. Using this strategy we generated stable transformants harboring precise deletions of the 4.6-kb *yellow* locus as well as flies with partial deletions of the locus that abolished *yellow* function[9]. We have also successfully deleted the 6.1-kb *rosy* locus (**Table 1**). Thus, multiple gRNAs can be simultaneously employed to delete entire open-reading frames.

Knock-ins/Insertions. In addition to NHEJ, targeted DSBs in chromosomal DNA can catalyze homologous recombination (HR) using a donor template for repair. The replacement of a gene with an attP ΦC31 phage recombination site has been successfully combined with traditional ends-in and ends-out HR approaches in Drosophila to provide ongoing genetic access to a locus of interest[25,26]. Using two gRNAs targeting sequences flanking *yellow* and a single-stranded oligodeoxynucleotide (ssODN) donor template, we were able to replace the endogenous *yellow* locus with an attP recombination site[9] (**Fig. 2C**). This result demonstrates that the CRISPR/Cas9 system can be employed for homology-based genome engineering in Drosophila.

The finding that targeted insertion of exogenous sequences into the Drosophila genome can be accomplished using the readily programmable CRISPR/Cas9 system opens the door to a wide variety of genome modifications. This approach can potentially be used to introduce specific point mutations, tag endogenous loci, or flank genes with recombination sites for conditional knock-out alleles, to cite just a few of the possibilities (**Figs. 2D and E**).

Starting a CRISPR genome engineering project

The CRISPR/Cas9 system is a straightforward genome engineering technique that can be employed by any laboratory to generate targeted modifications. Here we address three key considerations for starting your own CRISPR/Cas9 genome engineering project: (1) options for identifying targeted events, (2) strategies for maximizing targeting specificity, and (3) methods for delivering system components. We also direct you to the flyCRISPR discussion board, accessible through our website (flyCRISPR.molbio.wisc.edu/news), where members of the Drosophila community are sharing their ideas and strategies.

Screening for targeted events. The initial experiments employing the CRISPR/Cas9 system in Drosophila were facilitated by the fact that mutations in *yellow*, *rosy* and *white* yield easily scored visible phenotypes[8-10]. In most cases, however, phenotypic screening will not be an option, necessitating alternative approaches for recognizing targeted events. Here we discuss three options to consider as you design your engineering strategy: (i) molecular screening, (ii) targeting a marked locus, and (iii) supplying a marked donor template.

Molecular screening: PCR-based methods are a universal option, but the associated generation and maintenance of candidate fly stocks can be labor intensive. Thus to be broadly feasible, molecular screening requires a high rate of targeted events. Using a plasmid-based injection paradigm to generate indels at four targets in two genes, we observed mutant progeny at rates ranging from 0.25 to 22%[9] (**Table 1**). Injection of CRISPR components as RNAs yielded mutant progeny at a rate of 2 to 35% in experiments targeting five genes with 8 different gRNAs[8,10]. Although two gRNAs yielded no progeny due to lethality or sterility and there were no mutants among the progeny of flies injected with one of two gRNAs targeting *white*, a gRNA targeting the *K81* gene generated mutations that were transmitted to 99% of characterized progeny[10] – indicating that, while highly variable, indels can be generated at very high rates. It is important to note that all the numbers reported here likely include clonal events. These rates allowed Yu and colleagues to screen for mutations in progeny via PCR followed by either restriction enzyme-based analysis or direct sequencing of the products. With continued optimization and greater understanding of the factors that influence targeting efficiency, molecular screening of Cas9-induced mutations is likely to become increasingly feasible.

Negative screening: An alternative to molecular screening is to generate deletions in fly lines that contain visibly marked elements in the target locus (**Fig. 3A**). Because CRISPR components can be introduced into any genetic background via injection, any line carrying a marked element in a locus of interest can be used. One need only design the deletion or gene replacement experiment to remove the marker with two flanking gRNAs and screen for loss of the visible marker. A consequence of this approach is an increase in the size of the deletion that must be generated, which might adversely affect efficiency. We previously targeted the 4.6-kb yellow locus for deletion, and found that 21% of injected flies (founders) transmitted deletions resulting in the null *yellow* phenotype to 1.4% of progeny. One of eighteen founders transmitted the full 4.6-kb deletion, while the remainder of recovered events represented partial deletions[9]. A 6.1-kb deletion of the rosy locus was fortuitously recovered in experiments designed to replace the locus with an attP site. (Table 1). To date, the CRISPR/Cas9 system has not been used to generate deletions larger than 6.1 kb[9]. Nonetheless, the simplicity of this negative screening approach combined with the presence of marked elements in most Drosophila genes makes it an attractive option for many applications.

Positive screening: A third option is to use a donor template for HR that includes a visible marker (**Figs. 3B and C**). While ssODN donors such as the one we utilized to integrate an attP docking site are useful for small modifications, larger insertions and the incorporation of visible markers for screening will require double-stranded DNA (dsDNA) donors. Injected dsDNA donor templates have been successfully utilized in Drosophila to incorporate up to 13-kb of exogenous sequence in both P element- and zinc-finger nuclease-induced HR[27,28]. Based on a comprehensive analysis of ZFN-induced HR, dsDNA donors containing flanking homology arms of at least 1-kb in length serve as effective donors[27]. dsDNA donors have been employed in mammalian cells as

templates for CRISPR/Cas9-induced HR and there is every reason to expect they will work as effectively in flies[21]. While this strategy requires generating a donor template, it will facilitate simple screening while enabling a broad range of modifications, from gene deletion and replacement (**Fig. 3B**) to the incorporation of fluorescent tags (**Fig. 3C**). Flanking the visible marker with FRT or loxP recombination sites will allow for its subsequent removal.

Strategies for maximizing targeting specificity. Once you've determined your strategy for recognizing targeted events, you will need to select the sequences to be targeted. Specificity and efficiency of cleavage are critical parameters for genome engineering methods that rely on sequence-specific DSB generation in chromosomal DNA. Although little is known about the factors affecting Cas9 efficiency, significant attention has already been devoted to identifying parameters that maximize specificity[12,15,29,30]. Based on these studies, we expect that careful target selection and optimization of injection conditions will enable high specificity when employing the CRISPR/Cas9 system.

Initial CRISPR/Cas9 genome engineering studies identified a 12-nt 'seed' region adjacent to the PAM that was necessary for gRNA-guided cleavage by Cas9[12,15]. The detrimental effects of seed-region mismatches on cleavage efficiency suggested that targets lacking perfect matches to the 12-nt seed sequence adjacent to an NGG PAM sequence elsewhere in the genome would be highly specific. More recently, two studies in mammalian cell lines indicate that somewhat more stringent criteria should be applied when selecting target sequences to minimize off-target cleavage[29,30]. Importantly, Hsu et al. (2013) also found that NAG can serve as a PAM in mammalian cells, mediating cleavage with approximately one-fifth the efficiency of NGG. Together these findings suggest that

targeting specificity can be maximized by selecting targets with the fewest potential off-target cleavage sites as defined by the rules below:

1. PAM-adjacent sites with ≥ 11/12 matches to the target seed sequence.
2. PAM-adjacent sites with ≥ 18/20 matches to the full target sequence.
3. Sites meeting the above criteria adjacent to a divergent PAM of the form NAG as well as NGG.

An online CRISPR target identification tool is available from Feng Zhang's laboratory at http://www.genome-engineering.org/crispr. This tool identifies CRISPR target sites lacking perfect matches to the seed sequence elsewhere in the genome; an updated version incorporating the more conservative rules outlined above is currently not available for the Drosophila genome.

In addition to careful selection of target sites, the concentration of CRISPR components can be titrated to maximize specificity by taking advantage of cleavage efficiency differences between on- and off-target sites[30]. This has not yet been directly investigated in Drosophila. Finally, to further reduce the potential for generating mutations at off-target cleavage sites, Cas9 can be mutated and supplied as a nickase that will generate targeted single-strand breaks capable of catalyzing HR but unlikely to be repaired by the NHEJ pathway[15,21].

Encouragingly, neither we (Gratz et al. 2013) nor Bassett et al. (2013) uncovered evidence of cleavage at potential off-target sites identified by sequence similarity. In both studies, targets were selected to avoid perfect matches to the seed sequence adjacent to a PAM elsewhere in the genome, suggesting that even using looser criteria than above,

strategic target selection can limit off-target cleavage in Drosophila. It is also possible that differences in DNA repair between germ cells and transformed cell lines result in lower apparent rates of off-target cleavage in Drosophila since only those cleavage events that are improperly repaired yield mutations. However, comprehensive analyses have not yet been performed, so further work will be required to determine the potential for off-target effects in Drosophila.

Deciding on a delivery system. The final decision to make is how you will introduce CRISPR components into Drosophila embryos. In this section, we discuss three alternatives: (i) injection of DNA plasmids encoding Cas9 and gRNA(s) for in vivo transcription/translation, (ii) injection of these components as RNAs, or (iii) injection of modification-specific components into flies expressing the common system components, Cas9 and/or tracrRNA.

DNA injection: To generate indels in the *yellow* and *rosy* loci, we injected DNA plasmids encoding Cas9 and, separately, four different gRNAs into preblastoderm embryos. 6-27% (median = 12) of injected embryos that survived to become fertile adults transmitted targeted mutations to progeny at rates ranging from 0.25 to 22% (median = 2)[9] (**Table 1**). With our expression plasmid, unique gRNA-encoding constructs can be generated in a couple days using oligonucleotides. This U6-gRNA plasmid is available from Addgene and detailed protocols are available on our website (http://flycrispr.molbio.wisc.edu).

RNA injection: Two Drosophila groups have observed efficient germline transmission when injecting CRISPR components as RNA[8,10]. Using single gRNAs to induce indels in 12 targets, mutant progeny were generated at rates ranging from 0-99% (median = 9). The percentage of founders (injected embryos that yielded mutant progeny) ranged from

0-100% (median = 59%) for the 10 targets for which this data was reported. DNA templates for in vitro transcription of gRNAs can be generated as plasmids or PCR products[8,10].

Flies expressing CRISPR components: Germline expression of CRISPR/Cas9 components may increase targeting efficiency. To that end, we have generated transgenic flies that express Cas9 under the control of the *vasa* promoter. In addition, we have generated stable transgenic flies that express tracrRNA under the control of the *snRNA:U6:96Ab* promoter, and flies that express both Cas9 and tracrRNA. To facilitate rapid and widespread use, we have deposited these flies at the Bloomington Drosophila Stock Center. All lines are homozygous viable, suggesting low toxicity of CRISPR/Cas9 components in the absence of crRNA. Using these fly lines, targeted modifications can in principle be generated by injecting only the targeting component of the system. We do not yet have data on germline transmission using these lines; however, analysis of injected embryos indicates successful generation of Cas9-induced modifications with this approach.

Choosing a delivery system: To date, results have been reported for the generation of indels using DNA and RNA injection-based approaches. Initial reports raise the possibility that RNA injections may yield fewer surviving flies than DNA injections while yielding more affected progeny[8-10].

However, no direct comparisons have been made, which will be important for determining how these approaches compare in their cleavage efficiency and toxicity given the significant variability between loci and targets that has been reported. Similarly, future experiments will be required to determine how techniques employing transgenic flies expressing CRISPR components compare to the injection-only techniques reported

so far. A drawback of the transgenic approach is decreased flexibility in choosing the genetic background in which to generate modifications. The injection-based systems, on the other hand, place constraints on target selection due to the requirements for one or two Gs at the start of the gRNA sequence for efficient transcription from the U6 and T7 promoters, respectively. These additional requirements decrease the frequency of target sites in the genome from 1/8 to 1/32 for the plasmid-based system and 1/128 for an RNA injection approach. However, a new report suggests that these constraints may be easily overcome. In zebrafish, the G nucleotides can simply be added to the end of the gRNA construct without undue effects on efficiency[31]. Finally, higher or lower cleavage rates may be desirable depending on the particulars of a given genome modification experiment. For instance, when targeting an essential gene, high cleavage rates might frequently yield biallelic breaks that would decrease the likelihood of recovering a targeted event. Suggesting this is a significant concern, Yu et al. (2013) observed 100% male infertility in the targeting of two out of three sex-linked genes and 100% larval lethality when targeting an essential gene. These factors will likely combine to make different approaches ideal for distinct circumstances. The options already available suggest the CRISPR/Cas9 system will be a versatile technique adaptable for a broad range of applications.

Outlook

The CRISPR/Cas9 system has sparked extraordinary interest as a tool for genome engineering due to its simplicity and adaptability. In the short time since the development of the two-component system, work in Drosophila has already demonstrated that RNA-guided Cas9 can be used to generate indels, gene deletions and gene replacements – all of which are efficiently transmitted through the germline[8-10,12]. In the coming months, these successes are sure to be expanded upon as targeting

efficiency and specificity are optimally balanced and more complex modifications are attempted. With Drosophila researchers worldwide eagerly embracing CRISPR technology, we can look forward to realizing the promise of designer flies on demand.

Note

While this manuscript was in review, germline transmission of indels and defined deletions generated in transgenic flies expressing both Cas9 and a custom gRNA was reported[32]. Fillip Port and Simon Bullock also report the efficient generation and transmission of mutations in *yellow* and *ebony* in transgenic flies expressing Cas9 ubiquitously under the control of the *act5* promoter, while Hui-Min Chen and Tzumin Lee report their successful targeting of *yellow* in transgenic flies expressing UAS-Cas9, a germline-specific Gal4 and a gRNA. These data are available at http://www.crisprflydesign.org.

Acknowledgements

We are grateful to Alex Cummings, Laura Donohue, Danielle Hamm, Jenny Nguyen and Anna Zeidman for their assistance with experiments, and to Dustin Rubinstein and David Loehlin for helpful comments on the manuscript. We thank the Genetics Society of America for permission to use the schematic in Figure 1, which was originally published as Figure S1A of Gratz et al., 2013. This work was funded by startup funds from the University of Wisconsin to MMH, JW and KOCG and grants from the National Institutes of Health to JW (R00 NS072252) and KOCG (R00 NS060985 and R01 NS078179). Plasmids and transgenic fly lines described here are available through the non-profit distributor Addgene and the Bloomington Drosophila Stock Center, respectively. Detailed protocols and reagent information are available at flyCRISPR.wisc.edu.

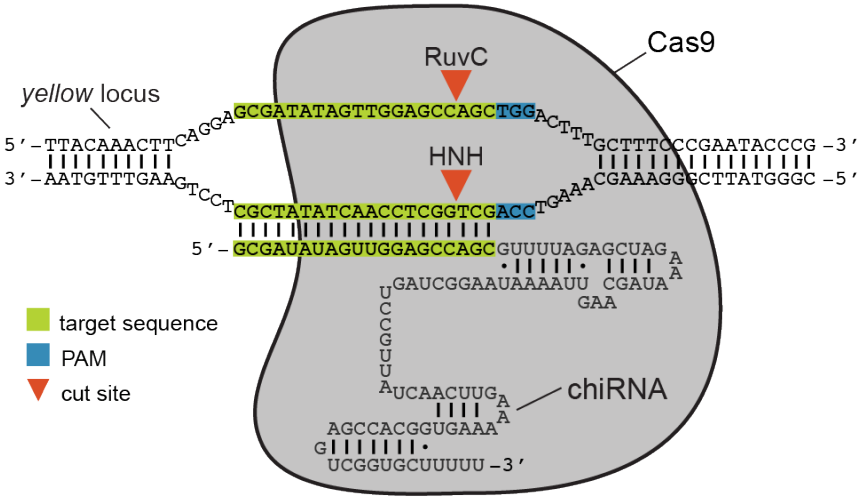

Figure 1. Schematic of the two-component CRISPR/Cas9 system. A target site in the *yellow* locus is shown as an example. Cas9 is guided to a cleavage site by a chimeric RNA containing critical crRNA and tracrRNA sequences, including 20-nt of homology to a target site. This RNA has alternately been referred to as a gRNA, a single-guide RNA (sgRNA) or a guide RNA (gRNA). Cas9 (grey) contains two distinct endonuclease domains, a HNH domain and a RuvC-like domain, that independently cleave both stands at the target site to generate a DSB (red arrowheads). Cleavage of target sites requires a high degree of homology to the gRNA and a 3-bp PAM (NGG) immediately 3′ of the target sequence.

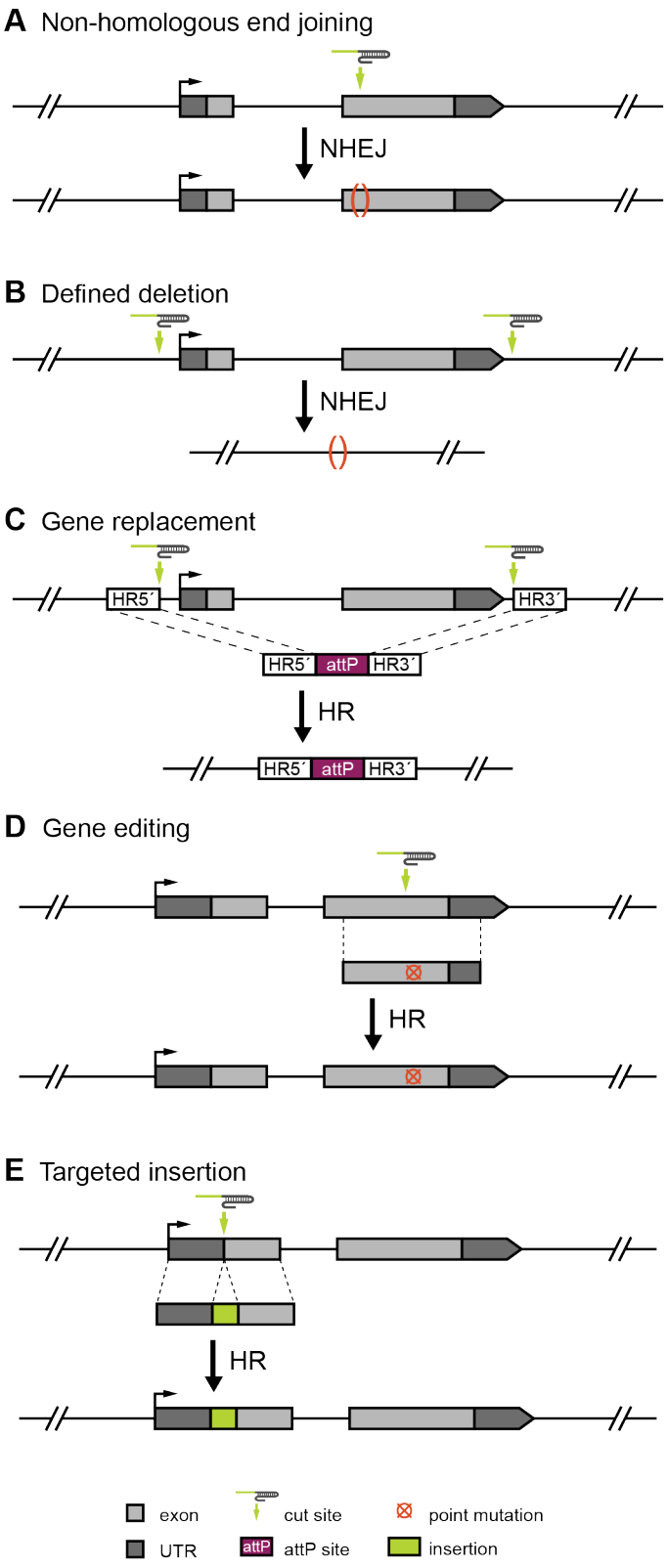

Figure 2. Engineering diverse genome modifications with the CRISPR system. (A) Using a single gRNA, targeted DSBs can be generated and repaired

imperfectly by NHEJ resulting in disruptive mutations[8-10]. (B) Through a process that is likely mediated by NHEJ, targeted deletions can be generated using two gRNAs targeting the limits of the region to be removed. This approach has been used to delete *yellow* and *rosy*[9]. (C) Employing two gRNAs targeting the limits of the region to be replaced and a donor template for HR, targeted genomic regions can be replaced with exogenous sequences. In this example, a gene is replaced with an attP PhiC31 phage recombination site to allow for subsequent manipulation of the locus. In addition to the attP sequence, the donor template contains homology arms corresponding to the sequences immediately adjacent to the predicted cleavage sites. This approach has been used to replace *yellow* with attP[9]. (D) Cas9-mediated HR can be used to engineer point mutations into a target locus. In this example, a target site is chosen near the sequence to be mutated, and novel sequence introduced via a donor template containing the modification flanked by sequences homologous to the target region. This approach has been used to introduce point mutations in mice and zebrafish[23,31]. (E) Insertions of exogenous sequence including fluorescent molecules and epitope tags can be incorporated into a target locus via HR. A target site at or near the intended insertion site is chosen and used in conjunction with a donor template comprising two homology regions flanking the sequence to be inserted.

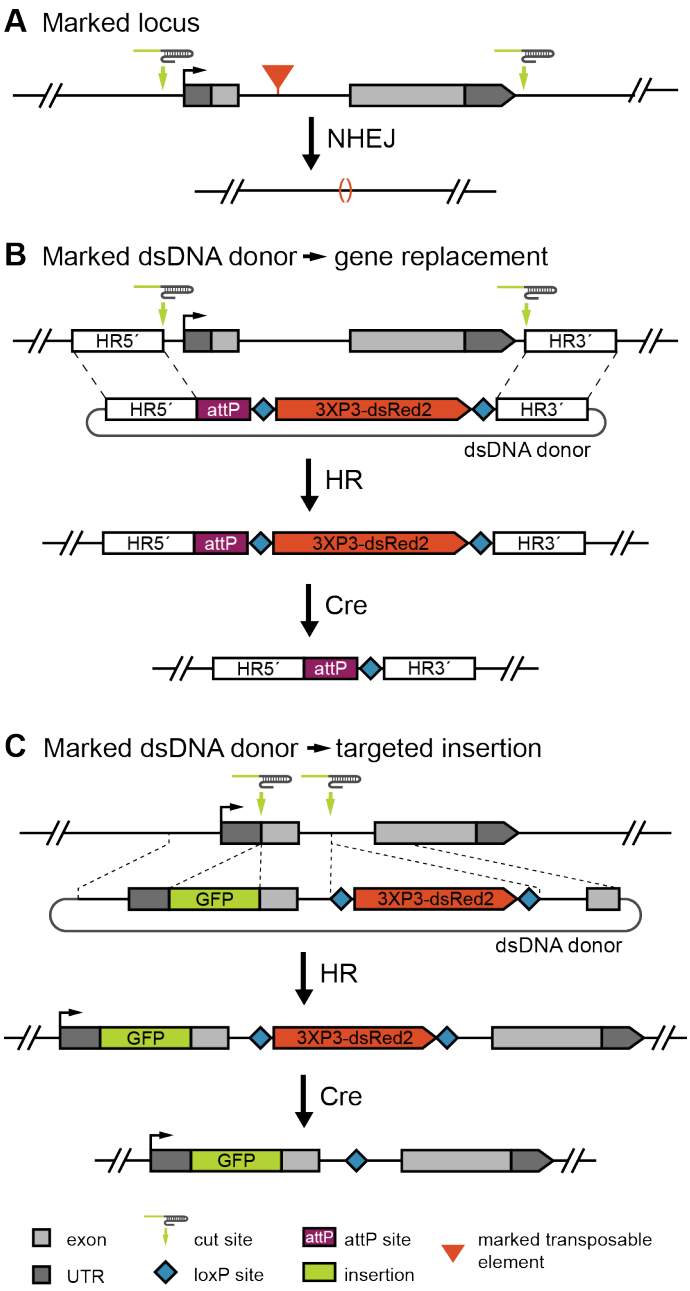

Figure 3. Strategies for detecting targeted events. (A) Targeting a locus with a marked element will allow screening for targeted deletions or gene replacements via loss of the marker. In this example, a locus containing a visibly marked transposable element marked is targeted for deletion. (B) A dsDNA donor with a removable positive marker can be used to facilitate screening for targeted deletions and gene replacements. In this example, a loxP-flanked fluorescent marker driven in the eye allows for identification of transformants based on eye color. The visible marker can later be removed using Cre recombinase. (C) In this example, a GFP tag is inserted at the translation start site of a gene using a donor template containing a removable positive marker to aid in screening. The donor was designed so the visible marker will be inserted in a non-conserved region of an adjacent intron as its subsequent removal will leave a single loxP site behind. This strategy might be facilitated by the use of two target sites – one at the site of the intended

insertion and another at the insertion site for the positive marker. With this approach, the donor vector would contain homology arms corresponding to sequences adjacent to each cleavage site, sequences homologous to the region between the insertion sites, GFP coding sequence, and sequences encoding a visible marker.

Table 1. Germline transmission rates of targeted mutations in *rosy*

gRNA(s)	ssODN donor	Male crosses		Female crosses		% (#) founders yielding targeted event	% (#) overall germline transmission	% (#) overall progeny
		% (#) founders	% (#) progeny	% (#) founders	% (#) progeny			
R1	-	10 (1/10)	1.2 (6/508)	20 (2/10)	2.5 (5/201)	100 (3/3)	15 (3/20)	1.6 (11/709)
R2	-	27 (3/11)	22 (88/404)	NA	NA	100 (3/3)	27 (3/11)	22 (88/404)
R3	-	9.7 (3/31)	1.7 (27/1621)	5.3 (1/19)	0.49 (2/406)	100 (4/4)	8 (4/50)	1.4 (29/2027)
R5', R3'	+	33 (4/12)	5.8 (35/603)	0 (0/6)	0 (0/549)	0 (0/4)	0 (0/18)*	3.0 (35/1152)

Flies injected with plasmids for expression of Cas9 and the indicated gRNAs with or without an ssODN donor template for the HR-mediated replacement of rosy with an attP ΦC31 phage recombination site were outcrossed and progeny screened for rosy eye color. The percentage of injected flies producing one or more rosy progeny (founders) is indicated along with the percentage of total progeny exhibiting rosy eyes. At least one progeny per founder was sequenced to determine if the targeted event had occurred. The percentage of founders in which the expected event occurred in one or more progeny is reported, as is the overall germline transmission rate (% injected flies yielding expected event).

*We recovered the precise deletion of the 6.1-kb *rosy* locus without attP incorporation from 1/18 crosses (5.6%) in 7 progeny (0.6% of total progeny screened).

References

1. Banga, S.S. & Boyd, J.B. Oligonucleotide-directed site-specific mutagenesis in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **89**, 1735-9 (1992).

2. Beumer, K.J. *et al.* Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**, 19821-6 (2008).

3. Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169-75 (2002).

4. Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R. & Engels, W.R. Targeted gene replacement in Drosophila via P element-induced gap repair. *Science* **253**, 1110-7 (1991).

5. Gong, W.J. & Golic, K.G. Ends-out, or replacement, gene targeting in Drosophila. *Proc Natl Acad Sci U S A* **100**, 2556-61 (2003).

6. Huang, J., Zhou, W., Watson, A.M., Jan, Y.N. & Hong, Y. Efficient ends-out gene targeting in Drosophila. *Genetics* **180**, 703-7 (2008).

7. Rong, Y.S. & Golic, K.G. Gene targeting by homologous recombination in Drosophila. *Science* **288**, 2013-8 (2000).

8. Bassett, A.R., Tibbit, C., Ponting, C.P. & Liu, J.L. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Rep* **4**, 220-8 (2013).

9. Gratz, S.J. *et al.* Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* (2013).

10. Yu, Z. *et al.* Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in Drosophila. *Genetics* (2013).

11. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-86 (2012).

12. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-21 (2012).

13. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res* **23**, 465-72 (2013).

14. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-2 (2013).

15. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23 (2013).

16. DiCarlo, J.E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* **41**, 4336-43 (2013).

17. Friedland, A.E. *et al.* Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods* **10**, 741-3 (2013).

18. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-9 (2013).

19. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233-9 (2013).

20. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).

21. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).

22. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* **23**, 720-3 (2013).

23. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910-8 (2013).

24. Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res* (2013).

25. Gao, G., Wesolowska, N. & Rong, Y.S. SIRT combines homologous recombination, site-specific integration, and bacterial recombineering for targeted mutagenesis in Drosophila. *Cold Spring Harb Protoc* **2009**, pdb prot5236 (2009).

26. Huang, J., Zhou, W., Dong, W., Watson, A.M. & Hong, Y. From the Cover: Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. *Proc Natl Acad Sci U S A* **106**, 8284-9 (2009).

27. Beumer, K.J., Trautman, J.K., Mukherjee, K. & Carroll, D. Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. *G3 (Bethesda)* (2013).

28. Keeler, K.J., Dray, T., Penney, J.E. & Gloor, G.B. Gene targeting of a plasmid-borne sequence to a double-strand DNA break in Drosophila melanogaster. *Mol Cell Biol* **16**, 522-8 (1996).

29. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* (2013).

30. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* (2013).

31. Hwang, W.Y. *et al.* Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS One* **8**, e68708 (2013).

32. Kondo, S. & Ueda, R. Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in Drosophila. *Genetics* (2013).

Appendix B

CRISPR-Cas9 Genome Editing in Drosophila

Scott J. Gratz[1], C. Dustin Rubinstein[2], Melissa M. Harrison[3], Jill Wildonger[4] and Kate M. O'Connor-Giles[1,2, 5*]

[1] Genetics Training Program, University of Wisconsin-Madison, Madison, WI 53706

[2] Laboratory of Cell and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706

[3] Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706

[4] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[5] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

This appendix is adapted from the publication:

Gratz, S J, C D Rubinstein, M M Harrison, J Wildonger, and K M O'Connor-Giles. 2015. "CRISPR-Cas9 Genome Editing in Drosophila." *Current Protocols* 31 (2): 1-0.

ABSTRACT

The CRISPR-Cas9 system has transformed genome engineering of model organisms from possible to practical. CRISPR-Cas9 can be readily programmed to generate sequence-specific double-strand breaks that disrupt targeted loci when repaired by error-prone non-homologous end joining or to catalyze precise genome modification through homology-directed repair (HDR). Here we describe a streamlined approach for rapid and highly efficient engineering of the Drosophila genome via CRISPR-Cas9-mediated HDR. In this approach, transgenic flies expressing Cas9 are injected with plasmids to express guide RNAs (gRNAs) and positively marked donor templates. We detail target site selection; gRNA plasmid generation; donor template design and construction; and the generation, identification and molecular confirmation of engineered lines. We also present alternative approaches and highlight key considerations for experimental design. The approach outlined here can be used to rapidly and reliably generate a variety of engineered modifications, including genomic deletions and replacements, precise sequence edits, and incorporation of protein tags.

INTRODUCTION

The CRISPR-Cas9 system is significantly advancing the ability of researchers to engineer targeted genome modifications for functional studies of genes and genetic elements. In Drosophila, the CRISPR-Cas9 system has been used to disrupt, delete, replace, tag and edit multiple genes and genetic elements (Bassett et al., 2013; Gratz et al., 2013a; Gratz et al., 2014; Kondo and Ueda, 2013; Lee et al., 2014; Port et al., 2014; Ren et al., 2013; Sebo et al., 2014; Xue et al., 2014; Yu et al., 2014; Yu et al., 2013). The rapid and widespread adoption of CRISPR-Cas9 illustrates the utility of this novel genome engineering platform for generating a wide variety of modifications, and its power for addressing fundamental biological questions, understanding and treating disease, and engineering agriculturally relevant species and their pests.

Endogenous CRISPR-Cas9 systems have been adapted as simple and highly robust genome engineering tools that are being widely adopted by the research community. The most widely used *Streptococcus pyogenes* system was simplified to two components to facilitate genome engineering: a common endonuclease called Cas9 and a single chimeric RNA referred to as a guide RNA (gRNA) (Jinek et al., 2012). gRNAs interact with Cas9 and guide the nuclease to specific DNA sequences through an easily programmed 20-nt target sequence that directly base pairs with complementary DNA. Upon binding its target, Cas9 utilizes its two nuclease domains to generate a double-strand break (DSB). The only known requirement for a potential cleavage site is the presence of a 3-bp protospacer adjacent motif (PAM) of the form NGG immediately 3′ of the 20-nt target sequence. Thus, *S. pyogenes* CRISPR-Cas9 target sites occur an average of once in every eight basepairs of genomic sequence.

Induction of a DSB in genomic DNA triggers repair by one of two general cellular repair pathways, both of which can be co-opted for genome engineering. Non-homologous end-joining (NHEJ) is an error-prone process in which broken ends are simply ligated together. This repair pathway can yield small insertions and deletions (indels) that disrupt function at cleavage sites. In contrast, homology-directed repair (HDR) employs homologous DNA sequences as templates for precise repair. By supplying donor templates comprising exogenous sequence flanked by homology-containing stretches (commonly referred to as homology arms), the HDR pathway can be appropriated to make precise modifications including defined deletions, sequence substitutions, or insertions. Beyond the genome engineering applications of the CRISPR-Cas9 system, nuclease-dead Cas9 is being used as a sequence-specific repressor or activator of gene expression and is being developed as a tool for probing genome structure and function without causing mutations (Anton et al., 2014; Bikard et al., 2013; Chen et al., 2013; Cheng et al., 2013; Fujita and Fujii, 2013; Gilbert et al., 2013; Kearns et al., 2014; Maeder et al., 2013; Perez-Pinera et al., 2013; Qi et al., 2013)

Here we detail a rapid and efficient CRISPR-Cas9 method for HDR-mediated engineering of the Drosophila genome (Gratz et al., 2013a; Gratz et al., 2014). We have used this approach to generate numerous genome modifications, including gene replacements, in-frame protein tag insertions, and conditional alleles. The Basic Protocol covers target site selection; gRNA generation; donor design and construction; and the generation, identification and molecular confirmation of engineered lines. We begin with key considerations for experimental design and discuss alternative approaches.

STRATEGIC PLANNING

Figure 1 shows a decision tree that can be used as a guide in designing the appropriate strategy for different types of CRISPR-Cas9 genome engineering experiments. Here we discuss the key considerations for each decision point.

1. What do you want to achieve?

   Loss-of-function (lof) allele: If your goal is to generate a lof allele, an approach that relies on either NHEJ or HDR can be employed to achieve your aim. The key difference in practice is whether a donor repair template is included. If you choose to go with NHEJ (no donor), you will have two options: (1) disrupt the locus by targeting a single cleavage event in a critical sequence and recovering disruptive indels or (2) delete the locus with two flanking gRNAs. You will need to screen candidate mutants using an appropriate molecular approach unless you can screen phenotypically for the desired mutant. The simplest way to identify relatively small indels is by using high-resolution melt analysis (HRMA), while PCR can be used to detect larger deletions (Bassett et al., 2013; Gratz et al., 2013a). Following sequence verification, you will have an unmarked lof allele.

   Complex modifications: To engineer defined modifications, such as specific changes to a nucleotide sequence or insertion of a tag or other exogenous DNA (e.g. FRT sites for a conditional allele), you will need to employ the HDR pathway by supplying a donor repair template and screening for the desired repair event. This pathway can also be employed to generate defined deletions or to insert screenable marker genes that facilitate recovery of engineered flies. Subsequently, markers that have been flanked by LoxP or FRT sites can be readily removed to

minimize alterations to the engineered locus or retained to provide a marked allele.

2. Irrespective of whether your goal is to employ the NHEJ or HDR pathway, you will need to decide how you will introduce Cas9.

   Transgenic Cas9 source: Unless your experimental goals limit you to a specific genetic strain, we recommend using a transgenic Cas9 source for the highest efficiency and reliability. We have generated transgenic fly lines expressing Cas9 under the control of the germline *vasa* promoter. These and related fly lines are available at the Bloomington Drosophila Stock Center (Gratz et al., 2014; Port et al., 2014; Sebo et al., 2014).

   Injection of Cas9 as DNA or RNA: If you wish to work in a particular genetic background (or non-*melanogaster* species), you can introduce Cas9 as either plasmid DNA or mRNA through injection. RNA-based approaches can generate targeted mutations at higher efficiency than the DNA constructs attempted to date, but yield somewhat more experiment-to-experiment variability, possibly due to complications encountered with the injection of RNA (Bassett et al., 2013; Yu et al., 2013). A disadvantage in using RNA is the relatively high cost compared to DNA, which is inexpensive to generate and easy to handle. The primary drawback of injected Cas9 DNA is that, while it appears to work quite reliably, it does so with lower efficiency than Cas9 mRNA (Gratz et al., 2013a; Ren et al., 2013). Optimized promoters are likely to mitigate this difference. In some cases, such as attempts to generate recessive lethal alleles, the lower efficiency observed

with DNA constructs may provide an advantage by reducing the levels of biallelic targeting.

3. All CRISPR experiments require a gRNA to guide Cas9 to the targeted locus. As with Cas9, gRNAs can be supplied transgenically or injected as DNA or RNA.

   Transgenic gRNA source: gRNAs can be integrated into the genome using phiC31-mediated transgenesis (Kondo and Ueda, 2013; Port et al., 2014; Xue et al., 2014). This approach has the advantage of catalyzing extremely efficient NHEJ – and likely HDR, though to date only one such experiment has been published (Port et al., 2014). The disadvantage of an integrated gRNA approach is the time and cost of generating a unique gRNA-expressing transgenic fly line for each targeting experiment. Integrating a gRNA more than doubles the timeline for obtaining an engineered fly line and costs several hundred dollars if injections are outsourced, as is commonly the case. However, given the advantage of increased efficacy, this approach is particularly useful if a gene or genomic sequence will be manipulated frequently.

   Injection of gRNA as DNA or RNA: Injection of gRNAs, as either DNA or RNA, is the most rapid method for engineering flies and both are highly efficient, particularly when combined with transgenic sources of Cas9 (Gratz et al., 2014; Ren et al., 2013; Xue et al., 2014; Yu et al., 2014).

4. For HDR, but not NHEJ, you will need to supply a donor repair template. Both dsDNA and single-stranded DNA (ssDNA) donors have been used successfully in Drosophila.

dsDNA donor: dsDNA donors are highly versatile as they can incorporate large DNA sequences (Gratz et al., 2014; Port et al., 2014; Xue et al., 2014; Yu et al., 2014). We have generated vectors for rapidly cloning locus-specific homology arms into donors with visible markers. The pHD-DsRed-attP and pHD-DsRed vectors, described in detail below, are available from Addgene.

ssDNA donor: ssDNA donor templates can be used to incorporate small modifications (Gratz et al., 2013a; Port et al., 2014; Xue et al., 2014). The primary advantage of using ssDNA donors is that they can be rapidly synthesized, obviating the need for cloning. However, most synthesis companies have a size limit of ~200 nt for single-stranded DNA synthesis, so they cannot be used for larger modifications such as the incorporation of fluorescent tags or for the inclusion of a selection marker for identifying engineered flies. Thus, molecular screening is required, increasing the labor necessary to identify and recover the intended allele.

In the Basic Protocol below, we detail our preferred method for efficient generation of engineered flies via HDR: injection of *vasa-Cas9* flies with gRNA plasmids and a positively marked dsDNA donor template. This choice represents a favorable balance of time, cost, efficiency, and reliability. With this approach, we obtain engineered alleles within one month at a total reagent cost of approximately $150. Injections generally cost an additional $200 if outsourced. In our experience, an average of 25% (range = 7-42%) of fertile injected flies transmit the targeted event to their progeny. In Alternate Protocol 1, we detail HDR with ssDNA donor templates. Alternate Protocol 2 covers HDR in other genetic backgrounds using Cas9 supplied as DNA. In Alternate Protocol 3, we outline our

approach for NHEJ using a transgenic source of Cas9 and gRNA supplied as DNA. This approach has also been used successfully by Ren et al. (2013). Together these protocols offer a versatile toolset amenable for generating a variety of genome modifications in Drosophila.

BASIC PROTOCOL

Target Site Selection

Selection of high-quality target sites is essential for the success of any CRISPR-based genome engineering experiment. It is important to identify target sites that will generate DSBs close to the location of the intended modification. In choosing a target site, location must be balanced with target-site specificity and, thus, the potential for off-target DSBs. While originally raised as a significant concern in the editing of transformed cell lines (Fu et al., 2013), with careful target site selection, off-target cleavage does not seem to be a significant problem for genome editing of organisms or human stem cells (Bassett et al., 2013; Chiu et al., 2013; Duan et al., 2014; Gratz et al., 2013a; Gratz et al., 2014; Kiskinis et al., 2014; Smith et al., 2014; Suzuki et al., 2014; Veres et al., 2014; Yang et al., 2013). Nonetheless, our current understanding of Cas9-induced cleavage is far from complete, so it is important to select the most specific sites possible to minimize the potential for off-target mutagenesis. To facilitate the rapid identification of high-quality target sites, we have developed a web-based tool, CRISPR Optimal Target Finder, that identifies gRNA cleavage sites and evaluates their specificity (Gratz et al., 2014).

It is essential that target sites be identified in sequence obtained from the fly strain that will be edited, not the reference genome. Polymorphisms between a given fly strain and the reference genome are frequent, especially in intergenic regions, and could eliminate or significantly decrease cleavage if they occur within your target sequence. Thus,

CRISPR Optimal Target Finder identifies gRNA target sites in user-supplied DNA sequence rather than reference genome sequences. In the Basic Protocol, we use *vasa-Cas9* flies. However, as described in Alternate Protocol 2 below, our approach can be readily adapted to engineer any fly strain.

Materials

vasa-Cas9 fly stocks (Bloomington Drosophila Stock Center)

y^1 M{vas-Cas9.RFP-}ZH-2A w^{1118}/FM7a, P{Tb^1}FM7-A (stock number 55821)

w^{1118}; PBac{vas-Cas9}VK00037/CyO, P{Tb^1}Cpr^{CyO-A} (stock number 56552)

w^{1118}; PBac{vas-Cas9}VK00027 (stock number 51324)

Total DNA purification kit

PCR and sequencing primers

Phusion High-Fidelity DNA Polymerase

Gel extraction kit

Protocol steps

1. Isolate genomic DNA from the Cas9 (or other) fly strain in which the genome modifications will be made. Purify total DNA from about 50 adult flies. This large genomic DNA preparation can be used for many subsequent CRISPR experiments.

2. Design PCR primers to amplify a region of about 500-1,000-bp centered around the target region.

Only one gRNA is necessary to catalyze HDR. We often design our experiments to include two gRNAs targeting either end of the region to be modified (See Figure 3). However, the effect of one vs. two gRNAs, if any, on efficiency or choice of cellular repair pathway is not known. Two gRNAs increase the likelihood that at least one gRNA will efficiently induce cleavage. On the other hand, two gRNAs increase the potential of off-target cleavage.

3. Prepare a 50-µl reaction to amplify the target region.

4. Use gel electrophoresis to purify the PCR product.

5. Sequence the PCR product using Sanger sequencing.

6. Identify CRISPR target sites in the sequenced region. Submit the sequence returned from step 5 to the CRISPR Optimal Target Finder (tools.flycrispr.molbio.wisc.edu/targetFinder).

7. Select the genome release you wish to search. Target Finder is currently configured to search *D. melanogaster* release 5 by default.

Release 6 is also available, as are genomes for D. simulans, D. yakuba, D. mauritiana, D. sechellia, D. ananassae, D. erecta, D. persimlis, D. pseudoobscura, D. virilis, D. mojavensis, D. willistoni, D. grimshawi, Anopheles gambiae *(strains M and S),* Aedes aegypti, Apis mellifera, Tribolium castaneum *(release 2 and draft release 4), and* C. elegans.

8. Select the length of the target sites you would like to identify. Target Finder will identify all target sequences of the selected length in the 500-1000 bp sequenced region from step 2. Full-length target sites of 20-nt offer the highest cleavage efficiency and are the most commonly used. Target sequences of a shorter length have been shown to increase the specificity of Cas9 cleavage while decreasing cleavage efficiency (Fu et al., 2014).

9. Target Finder will identify all CRISPR target sequences in the input sequence. If you would like to limit identified target sites to those that start with a G (for efficient expression from the U6 promoter when supplying gRNA as plasmid DNA) or those that start with GG (for efficient expression from the T7 promoter using in vitro transcription of gRNAs), select the appropriate option.

   Alternatively, a G can simply be added to the 5′ end of any target sequence when cloning the gRNA plasmid (see below) for efficient U6 transcription.

10. Once CRISPR targets are identified, their specificity will be evaluated based on user-selected criteria. The final 12 nt of the CRISPR target sequence, often referred to as the 'seed' sequence, are more critical for specificity than the distal eight nucleotides. The CRISPR Optimal Target Finder algorithms consider both the number and location of mismatches in the evaluation of potential off-target cleavage sites.

    High-stringency (default setting) defines potential off-target sites as those with (i) perfect matches (zero mismatches) to the seed sequence or (ii) one mismatch in the seed sequence and one or zero mismatches in the distal sequence.

Maximum stringency defines potential off-target as sites with (i) perfect matches (zero mismatches) to the seed sequence, (ii) one mismatch in the seed sequence and four or fewer mismatches in the distal sequence, or (iii) two mismatches in the seed sequence and a maximum of one mismatch in the distal sequence.

PAM: By default, the program will only consider sequences adjacent to a canonical NGG PAM in the evaluation of potential off-target cleavage sites. Putative off-target sites adjacent to a non-canonical PAM sequence of the form NAG can be considered by selecting the 'NGG and NAG' option. In transformed cell lines, target sites adjacent to an NAG PAM were cleaved at one-fifth the efficiency of those adjacent to a canonical NGG PAM sequence (Hsu et al., 2013).

We recommend use of the default settings for most applications in Drosophila where little off-target cleavage has been observed to date (Bassett et al., 2013; Gratz et al., 2013a; Gratz et al., 2014).

11. Select specific target site(s) for your genome engineering project by balancing proximity to the site you are editing and the potential for off-target cleavage. Target Finder returns all identified target sites in order of specificity. For each target site, the program also provides oligonucleotide sequences designed for gRNA plasmid cloning into a pU6-BbsI-gRNA vector.

gRNA Plasmid Preparation

To supply gRNAs containing the target-specific sequences from a plasmid DNA source, we have generated vectors for rapid cloning of target-specific sequences using short

complementary oligonucleotides and a simple annealing and ligation process. The pU6-BbsI-gRNA vectors utilize the small RNA promoter of a Drosophila U6 gene to express the gRNA.

Materials

pU6-BbsI-gRNA (Addgene; Plasmid 45946)

BbsI endonuclease

Sense and antisense gRNA oligonucleotides

T4 Polynucleotide Kinase

T4 DNA Ligase

Gel extraction kit

E. coli DH5α or other suitable cloning strain

Plasmid miniprep kit

Sequencing Primers

T7 (5′-TAATACGACTCACTATAGGG-3′)

T3 (5′- AATTAACCCTCACTAAAGGG-3′)

Endotoxin-Free Plasmid Midi Kit

12. Order a pair of short complementary oligonucleotides for each target site. You can design them yourself using the following guidelines. The oligonucleotide design incorporates the target sequence and cohesive ends for cloning into the pU6-BbsI-gRNA backbone. The top strand should be designed in the format of 5′-CTTCG$(N)_{19}$-3′, where G$(N)_{19}$ corresponds to your unique target site sequence beginning with a G for efficient transcription from the Drosophila U6 promoter (Figure 2). The bottom strand is designed in the format of 5′-AAAC$(N)_{19}$C-3′, with $(N)_{19}$C representing the reverse compliment of the targeting sequence.

Alternatively, to aid in design process, the CRISPR Optimal Target Finder has a feature that will output the oligonucleotide sequences needed for cloning selected target sites.

You can either use T4 Polynucleotide Kinase (PNK) to add the 5′ phosphates to standard oligonucleotides, as described below, or order 5′ phosphorylated oligonucleotides.

As noted above, target sites without an endogenous 5′ G can be used by simply adding a G to the 5′ end of a 20-nt target site in the format of G(N)$_{20}$ to achieve efficient transcription. In this case the top strand should be designed in the format of 5′-CTTCG(N)$_{20}$-3′ and the bottom strand designed in the format of 5′-AAAC(N)$_{20}$C-3′.

13. Resuspend oligonucleotides at a concentration of 100 μM in nuclease-free H_2O.

14. Combine 1 μL of the top strand oligonucleotide (100 μM stock), 1 μL of the bottom strand oligonucleotide (100 μM stock), 1 μL of T4 DNA Ligase buffer (10X), 6 μL of H_2O, and 1 μL of T4 Polynucleotide Kinase (10U/μL). Incubate at 37°C for 30 minutes, 95˚C for 5 min, then ramp to 25˚C at a rate of -5˚C/min.

15. Digest 1 μg of pU6-BbsI-gRNA or pU6-BbsI-U63-gRNA with 10 Units *BbsI* for two hours at 37°C.

The pU6-BbsI-gRNA plasmid expresses the gRNA under the control of the U6-2 promoter. The pU6-BbsI-U63-gRNA plasmid expresses the gRNA

under the control of the U6-3 promoter, but is otherwise identical. Port et al. (2014) reported that U6-3 was more efficient than either U6-1 or U6-2. However, we have not observed consistent differences between the two promoters in our HDR experiments. This might be explained by low numbers in all cases (only 3 flies per promoter were analyzed in the published work and we have not made exhaustive comparisons); differences between integrated and injected gRNA or between NHEJ- and HDR-based experiments; and/or locus or gRNA-specific differences. We use the two promoters interchangeably.

16. Use gel electrophoresis to purify the digested vector. Determine the DNA concentration using a spectrophotometer.

17. Ligate the annealed insert into pU6-BbsI-gRNA. Combine 1 μL of annealed insert (Step 14), 50 ng of *BbsI* digested pU6-BbsI-gRNA, 1 μL of T4 DNA Ligase buffer, 1 μL of T4 DNA Ligase, and enough H_2O to bring the reaction to 10 μL. Incubate at 25°C for 1 hour.

18. Transform the ligation reaction into DH5α cells and select colonies on plates containing 100 μg/mL ampicillin. A control transformation of digested pU6-BbsI-gRNA vector alone can be performed to ensure no contaminating undigested plasmid was collected in Step 16.

19. Isolate plasmids from 2-4 individual colonies using a plasmid miniprep kit. Screen for plasmids with incorporated oligonucleotides by Sanger sequencing using T7 and/or T3 primers.

Alternatively, positive colonies can be identified via colony PCR using a target site oligonucleotide in combination with a T3 or T7 primer. Use a pipette tip to scratch a visible amount of an individual colony into a PCR tube and use the tip to inoculate 4 mL of LB with ampicillin. Add 10 ul of PCR master mix to the PCR tube and use the appropriate cycling parameters to amplify the diagnostic product.

20. Prepare DNA for injection from a positive clone using an Endotoxin-Free Plasmid Midi Kit.

As with all Drosophila injection-based techniques, preparing high-quality DNA is critical to successful CRISPR-mediated genome engineering.

Donor Vector Design

dsDNA donor vectors can be made in many configurations to facilitate the generation of an endless variety of genome modifications (Figure 3). The Basic Protocol focuses on the design of donor constructs using the pHD-DsRed-attP or pHD-DsRed vectors available through Addgene. The design of ssDNA donors is described in Alternate Protocol 1 below. The pHD-DsRed-attP vector is used for generating marked knock-out alleles that harbor an attP phage recombination site for serial manipulations of the target locus catalyzed by phiC31. pHD-DsRed is used for generating positively marked targeted insertions or sequence edits. Both vectors include a removable DsRed marker expressed strongly in the eye for visual identification of lines with targeted events and contain multiple cloning sites for inserting locus-specific homology arms.

21. Identify the two ~1-Kb sequences that flank the cleavage site and the genomic region that will be modified.

    Extensive analysis of zinc-finger nuclease-catalyzed HDR demonstrated that homology arms of 1 Kb mediate efficient HDR (Beumer et al., 2013), and we and others have found this to be the case with CRISPR-based HDR (Gratz et al., 2014; Port et al., 2014; Xue et al., 2014; Yu et al., 2014).

22. Ensure that neither AarI nor SapI sequences occur within your homology arms, as these restriction enzymes are used for cloning the donor vector.

    If your homology arms contain AarI or SapI sites, alter your design and/or experimental protocol accordingly. For example, if the 5′ homology arm contains an AarI cut site, it may be necessary to 'invert' the locations of the 5′ and 3′ homology arms in the donor such that the donor cassette is placed on the antisense strand. If the 3′ arm contains an AarI cut site, it will simply be necessary to clone the 5′ homology arm first. Alternatively, the multiple cloning sites can be used instead of AarI or SapI.

23. Design and order primers to amplify both homology arms. Homology arms should contain sequence immediately adjacent to your cleavage sites (Cas9-mediated DSBs are generated 3-bp upstream of the PAM) for efficient HDR. When deleting or replacing a locus, the homology arms will simply comprise the sequence flanking the deleted region (Figure 3A). For inserting exogenous

sequences or editing genomic sequences, design overlapping extension PCR oligonucleotides to construct a fragment that contains (i) a 1-Kb homology region, (ii) the insertion or edit you wish to make, and (iii) the genomic sequence between the edit and the marker insertion site, usually in an adjacent intron (Figure 3B). For detailed guidance on the design of overlapping extension PCR primers, we refer the reader to a previous Current Protocols publication (Miklos et al., 2012).

Left Homology Arm (AarI): The forward primer should follow the format 5′-NNNN*CACCTGC*NNNNTCGC$(N)_{20}$-3′ where the *AarI* site is italicized, the cohesive end generated by *AarI* digestion in underlined, Ns represent spacer sequences required for efficient cleavage and accurate cohesive end generation, and $(N)_{20}$ represents the hybridization sequence. The reverse primers vary slightly for the two vectors because only one includes the attP sequence in the AarI overhang. For the pHD-DsRed-attP vector, the reverse primer should follow the format 5′-NNNN*CACCTGC*NNNNCTAC$(N)_{20}$-3′ (Figure 4A and B). For pHD-DsRed cloning, the reverse primer should follow the format 5′-NNNN*CACCTGC*NNNNTTAT$(N)_{20}$-3′ (Figure 4A and C).

Right Homology Arm (SapI): The right homology arm primer design is the same for both pHD-DsRed-attP and pHD-DsRed. The forward primer should follow the format 5′-NNNN*GCTCTTC*NTAT$(N)_{20}$-3′ where the *SapI* site is italicized, the cohesive end generated by *SapI* digestion is underlined, Ns represent spacer sequences required for efficient cleavage and accurate cohesive end generation, and $(N)_{20}$ represents the hybridization sequence (Figure 4D). The reverse primer should follow the format 5′-NNNN*GCTCTTC*NGAC$(N)_{20}$-3′ (Figure 4D).

Alternatives to overlapping extension PCR abound, including Gibson and Golden Gate cloning or DNA synthesis, which is becoming an increasingly cost-effective option.

24. If your target sequences remain intact in your homology arms, incorporate silent mutations to disrupt the PAM or two seed sequence nucleotides into your design. This will ensure that neither your donor construct nor the successfully engineered locus are targets for cleavage.

Donor Vector Construction

Below we outline the steps for rapidly constructing dsDNA donor constructs using the type IIS restriction sites *AarI* and *SapI* in the pHD-DsRed-attP or pHD-DsRed vectors for seamless integration of homology arms. These vectors also contain multiple cloning sites for an alternative cloning method.

Materials

pHD-DsRed-attP (Addgene; Plasmid 51019, also called pDSRed-attP)

pHD-DsRed (Addgene; Plasmid 51434)

Phusion High-Fidelity DNA Polymerase

dNTPs

Gel extraction kit

AarI endonuclease

SapI endonuclease

T4 DNA Ligase

Gel extraction kit

E. coli DH5α or other suitable cloning strain

LB with ampicillin

Plasmid Mini Kit

Primers

pHD-BB-1 (5′-ACGAAAGGCTCAGTCGAAAG-3′)

pHD-BB-2 (5′-TGATATCAAAATTATACATGTCAACG-3′)

pHD-HSP70-R (5′-CGGTCGAGGGTTCGAAATCGATAAG-3′)

pHD-SV40-F (5′-GGCCGCGACTCTAGATCATAATC-3′)

Endotoxin-Free Plasmid Midi Kit (Macherey-Nagel NucleoBond Xtra Midi EF or similar)

25. Prepare a 50-µl reaction to amplify each homology arm. Homology arms should be amplified from the targeted fly line as maximal homology increases the efficiency of HDR (Deng and Capecchi, 1992; Nassif and Engels, 1993).

0.5 µl of template DNA

2.5 µl of 10mM forward primer

2.5 µl of 10mM reverse primer

1.0 µl of 10mM each dNTPs

10 µl of 5X HF Phusion Buffer

0.2 µl of Phusion polymerase

Water to 50 µl

26. Perform PCR with the following parameters.

1 cycle:	2 min	94°C	(initial denaturation)

30 cycles:	10 sec	98°C	(denaturation)
	30 sec	60°C	(annealing)
	30 sec	72°C	(extension)
1 cycle:	10 min	72°C	(final extension)

27. Purify homology arms via gel electrophoresis. Determine the DNA concentration using a spectrophotometer.

28. Simultaneously digest 1 μg of the pHD-DsRed-attP or pHD-DsRed vector and 1 μg of the left homology arm PCR product with 4 Units of *AarI* for four hours at 37°C.

29. Use gel electrophoresis to purify both the *AarI* digested pHD-DsRed-attP or pHD-DsRed vector and left homology arm PCR product. Determine the DNA concentration using a spectrophotometer.

30. Ligate the left homology arm into pHD-DsRed-attP or pHD-DsRed vector. Combine 50 ng of *AarI* digested pHD-DsRed-attP, 3:1 molar ratio of digested left homology arm PCR product, 1 μL of T4 DNA Ligase buffer, 1 μL of T4 DNA Ligase, and enough H_2O to bring the reaction to 10 μL. Incubate at 25°C for 1 hour.

31. Transform the ligation reaction into DH5α cells and select colonies on plates containing 100 μg/mL ampicillin. A control transformation of digested plasmid alone can be included to ensure the sample is not contaminated with undigested plasmid.

32. Select 2-4 individual colonies and culture in 4 mL of LB overnight. Isolate plasmid DNA using a plasmid miniprep kit. Screen for plasmids with incorporated homology arms by Sanger sequencing using pHD-BB-1 and pHD-HSP70-R.

    Alternatively, positive colonies can be identified via colony PCR as described above.

33. Simultaneously digest 1 μg of positive constructs from Step 32 and 1 μg of the right homology arm PCR product with 5 Units of *SapI* for two to four hours at 37°C.

34. Use gel electrophoresis to purify both the *SapI* digested construct and right homology arm PCR product. Determine the DNA concentration using a spectrophotometer.

35. Ligate the right homology arm into pHD-DsRed-attP. Combine 50 ng of *SapI* digested pHD-DsRed-attP, 36.6 ng of digested left homology arm PCR product, 1 μL of T4 DNA Ligase buffer, 1 μL of T4 DNA Ligase, and enough H_2O to bring the reaction to 10 μL. Incubate at 25°C for 1 hour.

36. Transform the ligation reaction into DH5α cells and select colonies on plates containing 100 μg/mL ampicillin. A control transformation of digested plasmid alone can be included to ensure the sample is not contaminated with undigested plasmid.

37. Isolate plasmids from 2-4 individual colonies using a plasmid miniprep kit. Screen for plasmids with incorporated oligonucleotides by Sanger sequencing using pHD-BB-2 and pHD-SV40-F.

    Alternatively, positive colonies can be identified via colony PCR as described above.

38. Prepare high-quality DNA for injection from a positive clone using an Endotoxin-Free Plasmid Midi Kit.

Injection of CRISPR Components

CRISPR components are injected using standard Drosophila injection techniques. Here we provide the injection mixture for HDR in *vasa-Cas9* flies.

Materials

pU6-BbsI-gRNA(s)

pHD-DsRed-attP or pHD-DsRed donor vector

Sterile PCR-grade water

vasa-Cas9 fly stocks (Bloomington Drosophila Stock Center)

39. Prepare the injection mixture with 500 ng/µl plasmid donor vector and 100 ng/µl of each gRNA vector in sterile PCR-grade water.

    This parameter can be altered based on your experience. We have tested a range of concentrations and find that gRNA plasmid concentrations

between 50 and 250 ng/µl and dsDNA donor concentrations between 250 and 500 ng/µl yield successful HDR.

40. Inject 150-300 embryos using standard Drosophila techniques.

Identification and Molecular Confirmation of CRISPR Alleles

Following injection and an appropriate outcross of injected flies, candidate CRISPR alleles are easily identified by screening for flies with red fluorescent eyes in F1 progeny. Once these candidates have been crossed to a balancer line, they can be sacrificed for molecular characterization to verify recovery of the intended genome modification. Below we describe our strategy of performing three PCRs that, in combination with Sanger sequencing, confirm targeted and precise editing (See Figure 3).

Materials

Adult fly homogenization buffer (see Reagents and Solutions for recipe)

PCR and sequencing primers

pHD-HSP70-R (5′-CGGTCGAGGGTTCGAAATCGATAAG-3′)

pHD-SV40-F (5′-GGCCGCGACTCTAGATCATAATC-3′)

Left-Genomic-F (target locus specific)

Right-Genomic-R (target locus specific)

Phusion High-Fidelity DNA Polymerase

dNTPs

Gel extraction kit

41. Design primers (see Figure 3). First, we perform two flanking PCRs to amplify regions extending from within the DsRed marker cassette to the right and left

flanking genomic regions. To confirm precise incorporation at the targeted locus, it is critical that the genomic primer-binding sites are outside the homology arms of the donor vector as depicted in Figure 3. Candidate alleles are then confirmed via a spanning PCR using the two genomic primers. This reaction will amplify the entire modified region and rule out undesirable crossover ("ends-in") repair events that result in the incorporation of the entire donor vector including backbone (Yu et al., 2014).

42. Isolate genomic DNA from the F1 candidate flies identified by DsRed expression in the eyes. Anesthetize a single fly and place it in a 0.2 mL PCR tube. Using a P200 pipette tip draw up 50 µL of freshly prepared adult fly homogenization buffer. Keeping the buffer in the pipette, use the tip to disrupt and homogenize the fly. Once the fly is homogenized, dispense the remaining buffer.

43. Incubate at 37°C for 30-45 minutes followed by a 5 minute incubation at 95°C.

44. Use 1 µL of genomic DNA per 50 µL PCR.

45. Prepare 3 separate 50 µL reactions for each candidate.

    Primers sets:

    1 - Left-Genomic-F to pHD-HSP70-R

    2 - pHD-SV40-F to Right-Genomic-R

    3 - Left-Genomic-F to Right-Genomic-R

    Recipe:

1 µl of the genomic DNA from step 42

2.5 µl of 10mM forward primer

2.5 µl of 10mM reverse primer

1.0 µl of 10mM each dNTPs

10 µl of 5X Phusion Buffer

0.2 µl of Phusion polymerase

Water to 50 µl

46. Use gel electrophoresis to purify positive PCR products.

47. Sequence the PCR product using Sanger sequencing to confirm expected sequence.

ALTERNATE PROTOCOL 1

HDR with Single-Stranded DNA Donors

When engineering small modifications, it may be desirable to use ssDNA donors, which can be rapidly synthesized. However, ssDNA donors are generally limited to 200 nt and, thus, cannot be used for engineering large modifications (such as the integration of a fluorescent tag). They also cannot be designed to include a visible marker for screening, so molecular screening is required, which increases the time and labor required to recover engineered flies.

Protocol steps

1. Design an ssDNA donor with homology arms corresponding to sequences immediately adjacent to the targeted cleavage sites and the intended modification

or insertion. Homology arms of ~40-60 nt have been shown to mediate efficient ssDNA-based HDR in Drosophila (Banga and Boyd, 1992; Beumer et al., 2013; Gratz et al., 2013a; Port et al., 2014; Xue et al., 2014). The orientation of the ssDNA is critical to the success of the experiment. During DNA repair, free 3′ ends created by rescission at the DSB invade homologous DNA. Therefore, to serve as a template for repair, it is essential for the ssDNA be complementary to the free 3′ end. Note that the PAM is not required for cleavage of single-stranded DNA so it is important that the ssDNA does not include an intact target site or it may be a target of cleavage (Jinek et al., 2012).

2. The injection mixture should contain 100 ng/μl of ssDNA.

3. Engineered flies will need to be identified by modification-specific PCR amplification or a modification-spanning PCR followed by sequencing. DNA can be obtained from F1 flies after outcrossing. Alternatively, a non-lethal method, such as isolating wing or leg DNA can be used prior to outcrossing of F1 candidates.

ALTERNATE PROTOCOL 2

HDR in Any Genetic Background

For many applications, it is necessary or desirable to engineer a specific fly strain. This is easily accomplished using an injectable source of Cas9 such as pBS-Hsp70-Cas9. Targeting efficiency is lower than with a transgenic Cas9 source, so it is advisable to inject a larger number of embryos.

Additional Materials

pBS-Hsp70-Cas9 plasmid (Addgene; plasmid 46294)

1. Prepare the injection mixture with 250-500 ng/μl of pBS-Hsp70-Cas9, 500 ng/μl plasmid donor vector and 100 ng/μl of each gRNA vector in sterile PCR-grade water.

2. Inject 250-500 embryos using standard Drosophila techniques.

ALTERNATE PROTOCOL 3

Generation of Disruptive Indels and Deletions via NHEJ

If your goal is to generate a disruptive allele, you can target the NHEJ repair pathway by introducing Cas9 and one or two gRNAs in the absence of a donor repair template. Using one gRNA to target a single cleavage event in critical sequence, you can recover disruptive indels. With two gRNAs, you can delete the intervening sequence.

Protocol steps

1. Design and generate gRNAs. For introducing a disruptive indel, design one gRNA in a location where the insertion or deletion of a small number of nucleotides is expected to interfere with function. Indels are often less than 10 bp, but can be much longer (Koike-Yusa et al., 2014). For deleting a genomic region, design two gRNAs flanking the region you wish to delete. We have used this approach to delete regions as large as 14 Kb in Drosophila (Gratz et al., 2014). Note that the deletion will not be precise and you will likely observe the loss or gain of a small number of nucleotides at the repaired junction.

The generation of deletions by NHEJ with two gRNAs can be conducted in lines with a marked element in the targeted locus (Gratz et al., 2014). This allows for the detection of deletions by loss of the marker in the element.

2. The injection mixture should contain 100-250 ng/µl of each gRNA.

   When using a single gRNA, we have increased the concentration to 500 ng/µl.

3. Engineered flies will need to be identified molecularly. Indels can be identified by HRMA, while PCR can be used to detect larger deletions.

REAGENTS AND SOLUTIONS

Adult fly homogenization buffer:

10 mM Tris-HCl (pH 8.2)

25 mM NaCl

1 mM EDTA

Store at room temperature for up to 6 months

1 µl of 20 mg/mL proteinase K added to 99 µl of homogenization buffer just prior to use

COMMENTARY

Background Information

The CRISPR-Cas9 system is a highly accessible and effective tool for genome engineering in *Drosophila* (Bassett and Liu, 2014; Gratz et al., 2013b; Harrison et al., 2014; Kondo, 2014). The Basic Protocol outlined above details an optimized CRISPR-Cas9 approach that has several advantages. We use a transgenic source of Cas9, expressed in the germline under the control of the *vasa* promoter, to achieve highly efficient and reliable genome engineering. The introduction of gRNA using rapidly constructed plasmids is quick and inexpensive. Our donor vectors facilitate streamlined cloning of locus-specific donor templates, and the incorporation of a removable DsRed marker makes identification of candidate alleles markedly easier than identification through molecular characterization.

While not covered in our protocol, other groups have successfully applied the CRISPR-Cas9 system in Drosophila using a variety of methods for introducing gRNAs and Cas9. NHEJ has been successfully accomplished using transgenic Cas9 + gRNA supplied as RNA (Xue et al., 2014), transgenic Cas9 + transgenic gRNA (Kondo and Ueda, 2013; Port et al., 2014; Xue et al., 2014), Cas9 DNA + gRNA plasmid (Gratz et al., 2013a; Ren et al., 2013), and Cas9 mRNA + gRNA supplied as RNA (Bassett et al., 2013; Yu et al., 2013). Successful HDR has been reported using Cas9 DNA, gRNA plasmid and either a dsDNA or ssDNA donor (Gratz et al., 2013a; Gratz et al., 2014). However, efficiency is higher with a transgenic Cas9 source, and all other HDR experiments reported in Drosophila to date have been conducted in Cas9-expressing flies using either gRNA plasmid (Gratz et al., 2014), gRNA supplied as RNA (Xue et al., 2014; Yu et al., 2014), or a transgenic gRNA source (Port et al., 2014).

Critical Parameters

Sequencing of target sites: Due to the prevalence of polymorphisms between distinct genetic backgrounds in Drosophila, it is critical to sequence the intended target locus in the genetic background in which the genome engineering experiment will be performed. Even a single basepair change in a target site can be detrimental to the success of the experiment.

Donor template construction: To protect both the donor template and the modified locus from unintended cleavage, it is critical that your donor template not contain an intact gRNA target site

Molecular confirmation of engineered lines: Because unexpected events can always occur during DNA repair, it is important to thoroughly confirm all candidate alleles. To do this, we suggest three PCRs (See Figure 3B) that together will confirm that engineered DNA has been incorporated at the target locus and that the locus is free of additional modifications, including the integration of donor vector backbone sequences (Yu et al., 2014).

Finally, it is important to note that, while CRISPR-Cas9 is working quite well in Drosophila, the system is not yet fully understood. For example, locus- and sequence-specific effects on cleavage efficiency are poorly understood. The Perrimon group (Harvard Medical School) has developed a tool that uses data from high throughput experiments in S2 cells to predict cleavage efficiency based on gRNA target sequence (www.flyrnai.org/evaluateCrispr/). An understanding of how donor composition and other experimental design features may influence the efficiency of HDR or the DNA repair pathway utilized by the cell awaits further study.

Troubleshooting

Poor viability: Reduce the concentration of gRNAs and donor vector in the injection mixture. While this may reduce efficiency, we have found that reducing the overall concentration of the injection mixture can increase viability. If the poor viability is due to highly efficient generation of a lethal allele, use an injected DNA source of Cas9 (pBS-hsp70-Cas9) instead of *vasa-Cas9*. This will decrease cleavage efficiency, and thus the occurrence of biallelic events, facilitating the recovery of recessive lethal lesions.

Poor efficiency: In the event that a given targeting experiment fails to yield engineered alleles, the gRNAs should be tested for cleavage efficiency. For genome engineering strategies using a pair of gRNAs, a simple PCR spanning the two gRNA target sites can be performed on embryos 24 hours after injection of both gRNAs into *vasa-Cas9*. Amplicons indicating a deletion between the two targeted cleavage sites demonstrate that both gRNAs are capable of generating DSBs. For strategies using one gRNA, cleavage efficiency can be tested using HRMA or a mismatch-specific nuclease assay on embryos 24 hours post injection. This approach can also be used to assess gRNA efficiency prior to embarking on a CRISPR-based experiment.

Anticipated Results

Using the approach described above to replace genes with attP docking sites in multiple experiments, an average of 24% of injected flies produced correctly engineered progeny. The deleted genes ranged in size from 2 to 27 Kb. Interestingly, we have not observed a strong correlation between deletion size and efficiency, suggesting that locus- or gRNA-specific effects may play a larger role in determining differences in efficiency between targeting experiments. We have also used this approach to insert in-frame tags in a number of loci at a similar average efficiency of 26%.

The highest probability off-target cleavage sites can be identified by sequence similarity to the targeted site. In a subset of our experiments, we have assayed these sites and found no evidence of off-target cleavage (Gratz et al., 2014). Based on this and similar findings by others, we expect that with careful target site selection, engineered fly lines can be generated with few or no off-target mutations in most cases (Bassett et al., 2013; Gratz et al., 2013a; Gratz et al., 2014).

Time Considerations

Target site selection, preparation of gRNA plasmids, and construction of donor templates can be accomplished in 1 week with 1-4 hours hands-on time per day. Embryo injection can be accomplished in 1 day with 4 hours hands-on time. After 10 days, injected flies can be outcrossed in approximately 1 hour. After another 10 days, F1 progeny are screened and outcrossed (2 hours) before molecular confirmation, which can be completed in 2 days (5 hours hands-on time).

ACKNOWLEDGEMENTS

We thank the members of the Harrison, Wildonger and O'Connor-Giles labs for invaluable input throughout this work. Plasmids and transgenic fly lines described here are available through the non-profit distributor Addgene and the Bloomington Drosophila Stock Center, respectively. Additional information and resources are available at flyCRISPR.molbio.wisc.edu and tools.flycrispr.molbio.wisc.edu/targetFinder. All software code is available upon request. This work was funded by startup funds from the University of Wisconsin, a grant from the McKnight Foundation to KOCG, grants from the National Institute of Neurological Disorders and Stroke, National Institutes of

Health to KOCG (R01 NS078179 and R21 NS088830) and JW (R00 NS072252), and grants from the Wisconsin Partnership Program and March of Dimes to M.M.H.

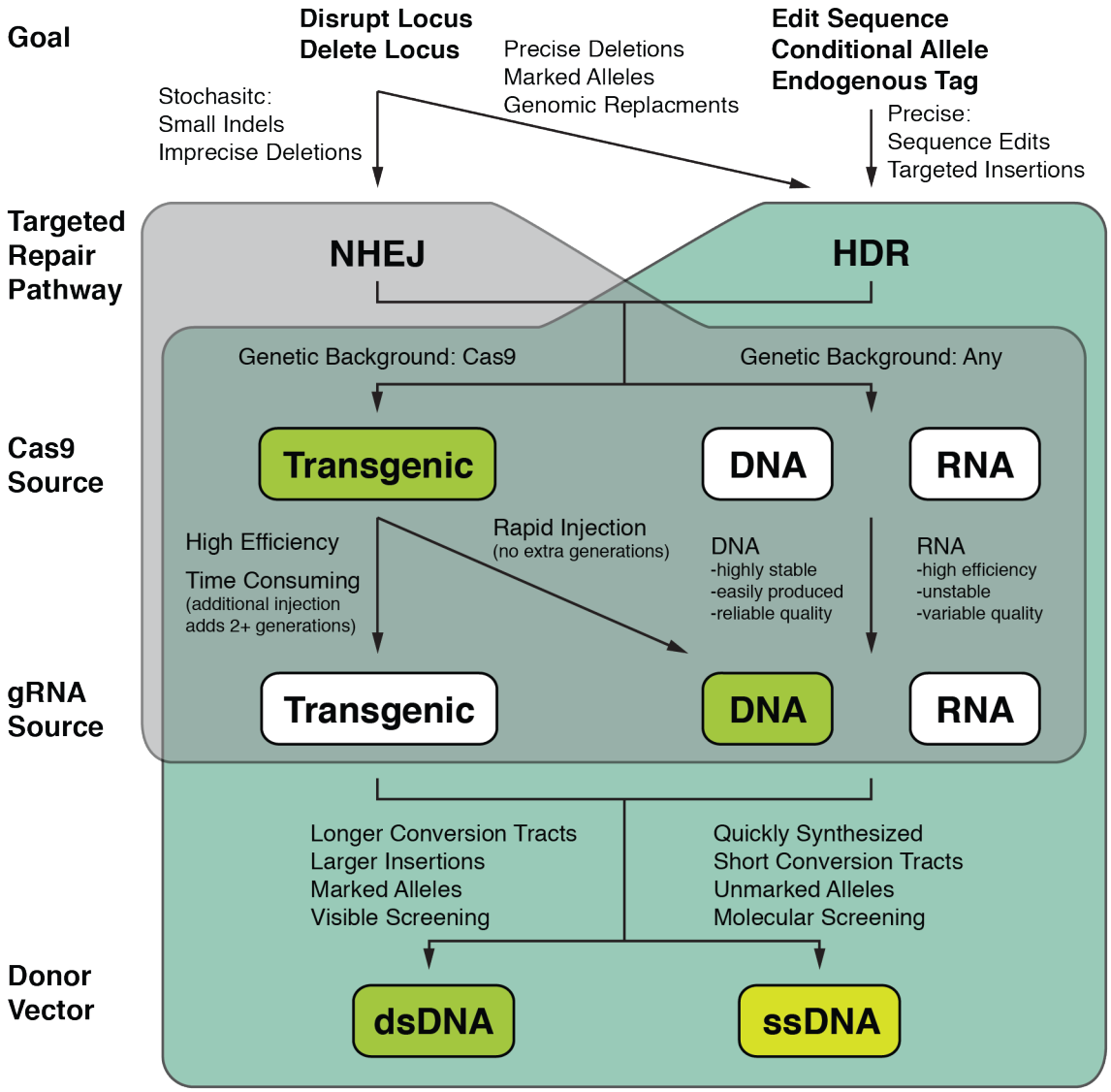

Figure 1. Strategic planning flowchart. The options outlined in this protocol are indicated by green boxes. See text (Strategic Planning) for a detailed discussion of each choice point, including the advantages and disadvantages of each strategy.

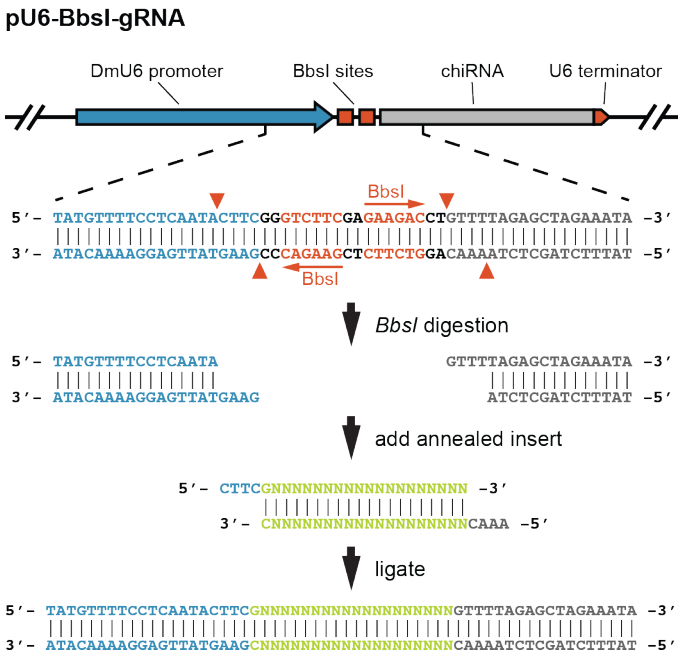

Figure 2. gRNA plasmid cloning. The pU6-BbsI-gRNA vector contains two BbsI cut sites between the Drosophila U6-2 (snRNA:U6:96Ab) promoter and the common portion of the gRNA. Specific target site sequences are synthesized as complementary oligonucleotides designed to generate appropriate cohesive 5′ overhangs once annealed. The annealed oligonucleotides, once phosphorylated, are then ligated into the BbsI digested pU6-BbsI-gRNA vector. U6-2 sequence (blue), BbsI recognition sequences (red), target site specific sequence (green), and the common portion of the gRNA (grey) are indicated. Red arrowheads denote the breakpoints generated by BbsI cleavage.

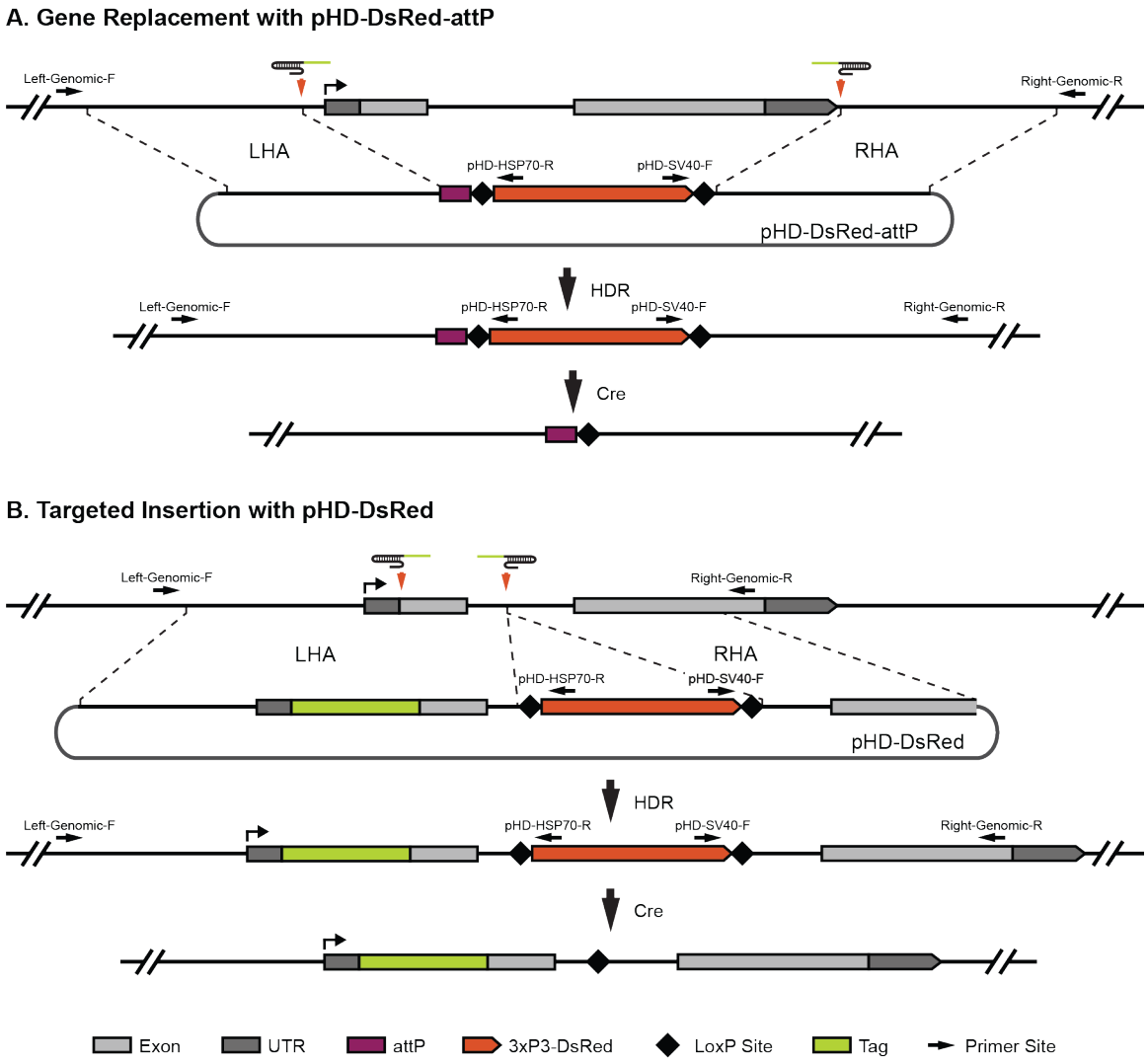

Figure 3. Donor construct design. The (A) pHD-DsRed-attP vector and (B) pHD-DsRed donor vectors and their typical uses are depicted. Both vectors contain a removable 3xP3-dsRed marker flanked by LoxP sites and two multiple-cloning sites for insertion of the left (LHA) and right (RHA) homology arms. pHD-DsRed-attP also contains the recombination-based docking site attP. (A) In the case of replacing a locus using the pHD-DsRed-attP vector, two target sites flanking the region to be replaced are chosen. Homologous sequences immediately flanking the cleavage sites should be cloned into the MCSs. Upon Cas9-mediated cleavage and HDR, the region between the two gRNA cut sites is replaced with the attP site and removable DsRed marker. Using Cre recombinase, the DsRed marker can be removed leaving only the attP docking site and a single LoxP site. (B) In the case of tagging a gene using the pHD-DsRed vector, select a target site close to the tag insertion site and another target site in a nearby intron where the DsRed marker will be placed. Homology arms will include sequences immediately flanking the cleavage sites. In addition, one of the homology arms will contain the in-frame tag and sequences between the tag and the DsRed marker. Upon Cas9 cleavage and HDR, the untagged region is replaced with a tagged region and a visible 3xP3-DsRed marker. Using Cre recombinase, the DsRed marker can be removed leaving only the tagged coding sequence and a single LoxP site. Black arrows indicate the primer binding sites

used for molecular characterization for candidate alleles. Note that the two locus specific primers are in the genomic region outside of the homology regions used in the donor vector.

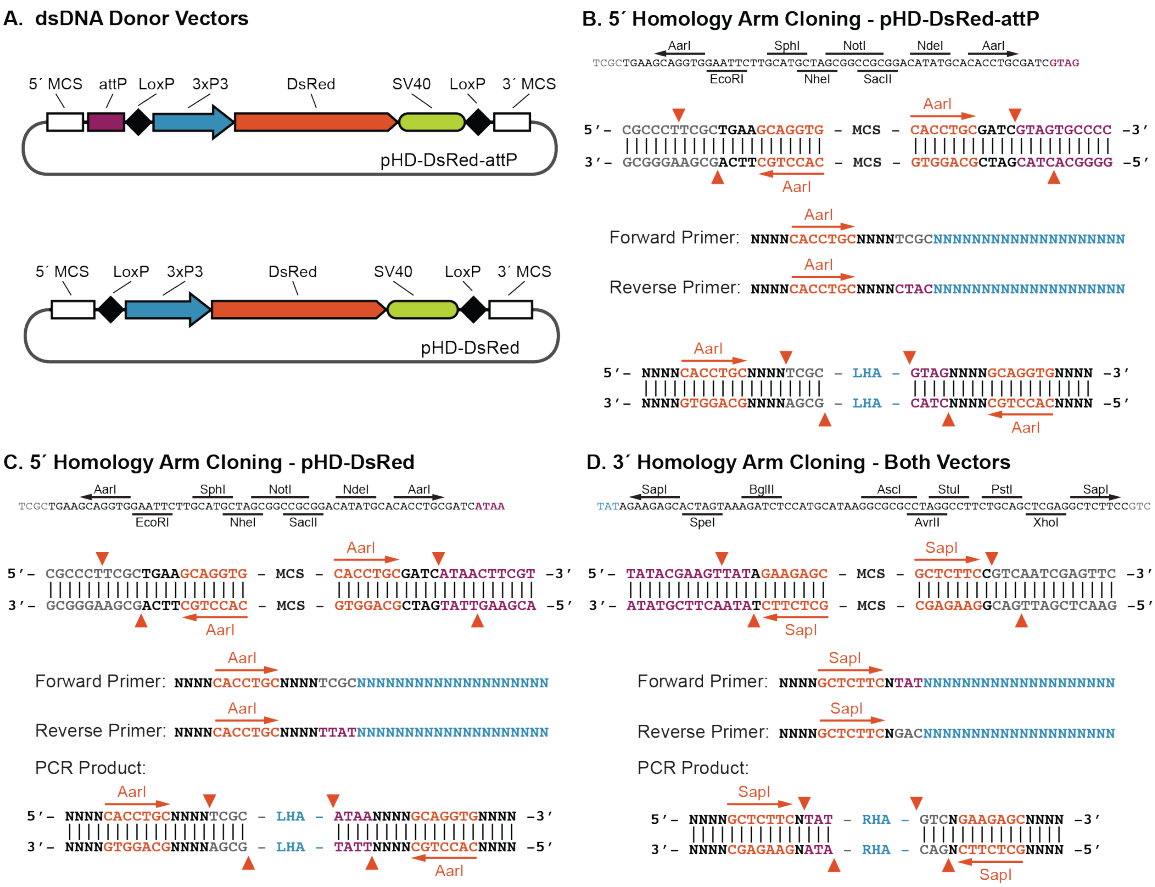

Figure 4. Donor plasmid cloning. (A) Schematic of the pHD-DsRed-attP and pHD-DsRed donor vector including the MCSs. (B-D) Multiple cloning site sequences and primer design for type IIS restriction site (AarI or SapI) based cloning of homology arms for the LHA of pHD-DsRed-attP (B), the LHA of pHD-DsRed (C), and the RHA of both pHD-DsRed-attP and pHD-DsRed (D). Note that the LHA of pHD-DsRed-attP and pHD-DsRed requires slightly different primers due to the presence or absence of the attP site. Vector backbone sequences (grey), attP/LoxP sequence (purple), AarI/SapI recognition sites (red), and locus specific hybridization sequences (blue) are indicated. Red arrows indicate the breakpoints generated by AarI or SapI digestion.

References

Anton, T., Bultmann, S., Leonhardt, H., and Markaki, Y. 2014. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5.

Banga, S.S. and Boyd, J.B. 1992. Oligonucleotide-directed site-specific mutagenesis in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* 89:1735-1739.

Bassett, A.R. and Liu, J.L. 2014. CRISPR/Cas9 and genome editing in Drosophila. *Journal of genetics and genomics = Yi chuan xue bao* 41:7-19.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell reports* 4:220-228.

Beumer, K.J., Trautman, J.K., Mukherjee, K., and Carroll, D. 2013. Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. *G3*.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* 41:7429-7437.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155:1479-1491.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research* 23:1163-1171.

Chiu, H., Schwartz, H.T., Antoshechkin, I., and Sternberg, P.W. 2013. Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. *Genetics* 195:1167-1171.

Deng, C. and Capecchi, M.R. 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Molecular and cellular biology* 12:3365-3371.

Duan, J., Lu, G., Xie, Z., Lou, M., Luo, J., Guo, L., and Zhang, Y. 2014. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell research*.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* 31:822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* 32:279-284.

Fujita, T. and Fujii, H. 2013. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and biophysical research communications* 439:132-136.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., and Qi, L.S. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154:442-451.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. 2013a. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194:1029-1035.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. 2014. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. *Genetics* 196:961-971.

Gratz, S.J., Wildonger, J., Harrison, M.M., and O'Connor-Giles, K.M. 2013b. CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly* 7:249-255.

Harrison, M.M., Jenkins, B.V., O'Connor-Giles, K.M., and Wildonger, J. 2014. A CRISPR view of development. *Genes & development* 28:1859-1872.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., and Zhang, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* 31:827-832.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816-821.

Kearns, N.A., Genga, R.M., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. 2014. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* 141:219-223.

Kiskinis, E., Sandoe, J., Williams, L.A., Boulting, G.L., Moccia, R., Wainger, B.J., Han, S., Peng, T., Thams, S., Mikkilineni, S., Mellin, C., Merkle, F.T., Davis-Dusenbery, B.N., Ziller, M., Oakley, D., Ichida, J., Di Costanzo, S., Atwater, N., Maeder, M.L., Goodwin, M.J., Nemesh, J., Handsaker, R.E., Paull, D., Noggle, S., McCarroll, S.A., Joung, J.K., Woolf, C.J., Brown, R.H., and Eggan, K. 2014. Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell* 14:781-795.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C., and Yusa, K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature biotechnology* 32:267-273.

Kondo, S. 2014. New horizons in genome engineering of Drosophila melanogaster. *Genes & genetic systems* 89:3-8.

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

Lee, J.S., Kwak, S.J., Kim, J., Kim, A.K., Noh, H.M., Kim, J.S., and Yu, K. 2014. RNA-guided genome editing in Drosophila with the purified Cas9 protein. *G3* 4:1291-1295.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature methods* 10:977-979.

Miklos, A.E., Hughes, R.A., and Ellington, A.D. 2012. Design and assembly of large synthetic DNA constructs. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 3:Unit3 23.

Nassif, N. and Engels, W. 1993. DNA homology requirements for mitotic gap repair in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 90:1262-1266.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* 10:973-976.

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. *Proceedings of the National Academy of Sciences of the United States of America* 111:E2967-2976.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173-1183.

Ren, X., Sun, J., Housden, B.E., Hu, Y., Roesel, C., Lin, S., Liu, L.P., Yang, Z., Mao, D., Sun, L., Wu, Q., Ji, J.Y., Xi, J., Mohr, S.E., Xu, J., Perrimon, N., and Ni, J.Q. 2013. Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proceedings of the National Academy of Sciences of the United States of America* 110:19012-19017.

Sebo, Z.L., Lee, H.B., Peng, Y., and Guo, Y. 2014. A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. *Fly* 8:52-57.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., and Ye, Z. 2014. Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. *Cell Stem Cell* 15:12-13.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebl, A., Tang, S., Ren, R., Aizawa, E., Zhang, F., Xu, X., Soligalla, R.D., Chen, F., Kim, J., Kim, N.Y., Liao, H.K., Benner,

C., Esteban, C.R., Jin, Y., Liu, G.H., Li, Y., and Izpisua Belmonte, J.C. 2014. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* 15:31-36.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Talkowski, M.E., and Musunuru, K. 2014. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell* 15:27-30.

Xue, Z., Ren, M., Wu, M., Dai, J., Rong, Y.S., and Gao, G. 2014. Efficient gene knock-out and knock-in with transgenic Cas9 in Drosophila. *G3* 4:925-929.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154:1370-1379.

Yu, Z., Chen, H., Liu, J., Zhang, H., Yan, Y., Zhu, N., Guo, Y., Yang, B., Chang, Y., Dai, F., Liang, X., Chen, Y., Shen, Y., Deng, W.M., Chen, J., Zhang, B., Li, C., and Jiao, R. 2014. Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the Drosophila genome. *Biology open* 3:271-280.

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